Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040308

A Size-Biased Probability Distribution for the Number of Male Migrants

Priti and B. P. Singh[*]

Department of Statistics, Banaras Hindu University, Varanasi, 221005, India.

Received: 18 Jun. 2015, Revised: 18 Aug. 2015, Accepted: 20 Aug. 2015
Published online: 1 Nov. 2015

Abstract: In this paper, an attempt is made to study the distribution of the number of male migrants from household through a size-biased probability model based on certain assumptions. The parameters of the proposed model have been estimated by method of moments and its suitability is tested by applying it to a number of the observed set of population data collected under some sample surveys. We found that the proposed model describes the observed data satisfactorily well.

Keywords: Migration, Size-biased probability model, Negative Binomial distribution and Estimation.

1 Introduction

Process of migration includes various factors affecting the movements of an individual of a household. These factors are characterised by age, sex, marital status, education etc. Various studies have been conducted to study migration at macro-level based on net or gross migration flows. These studies explain the aggregate migration flow or the rate of migration by identifying factors which make certain areas attractive to migrants and those which cause others to experience out-migration (Bannerji.B., 1986). But to study the factors affecting the movement process is carried out with models that incorporate factors at the micro, or individual/ household level (Bills Borrow, R. E. et al 1987) Micro-level analysis of migration is important for several regional planning, housing policies and sociological models (Pryor. R.J., 1975).

Adult males are more prone to migrate than other people of the community. The process of migration can take place as an individual or complete household or children or dependents with individual migrant. When an investigator records an observation by nature according to a certain stochastic model the recorded observation will not have the original distribution unless every observation is given an equal chance of being recorded (G. P. Patil and C. R. Rao, 1978).

The idea of weighted distributions was first given by Fisher (1934) to make up for determination bias in real life situation more precisely while the size-biased distribution which is a special case of weighted distributions was introduced by Cox 1962. These two concepts have been proved to be very useful for various applications in the field of bio-statistics such as family history of disease, survival studies etc. but their application for studying the human and wildlife populations was first traced by an article given by Patill and Rao (1978). Size biased or length biased distribution is a special case of various forms of weight functions given by Patil which are useful in scientific and statistical literature. The present study is an attempt to propose an alternative probability model to describe the pattern of male migrants aged 15 years and above. It has been obtained by compounding the size biased Poisson and the Gamma distribution.

2 Size Biased Distribution

Let X be a non negative random variable having probability function $f(x,\theta)$, with unknown parameter θ , where $\theta \in \Theta$, the parameter space and the $E(X)$ expected value. Then $f^*(x,\theta)$ is known as Size-biased distribution with probability

[*] Corresponding author e-mail: praj.preeti@gmail.com

© 2015 NSP
Natural Sciences Publishing Cor.

function as

$$f^*(x,\theta) = \frac{xf(x,\theta)}{E(x)} \tag{1}$$

3 The Models

3.1 Models 1

Yadava et al. (2004) proposed a probability model for the number of male out-migrants aged 15 years and above from a household. Let X denotes the number of migrants from a household, then it follows the size-biased geometric with parameter whose probability density function is given by

$$P(X=x) = x\theta^2(1-\theta)^{x-1} \tag{2}$$

Where θ is the parameter.
The above model contains only one unknown parameter which is estimated by the method moments as follows:

$$\mu = \sum_{x=1}^{n} xP[X=x] = \sum_{x=1}^{\infty} x * x\theta^2(1-\theta)^{x-1} \tag{3}$$

Solving the above expression we get

$$\mu_1 = \frac{2(1-\theta)}{\theta} + 1 \tag{4}$$

Solution of $\mu_1 = \bar{x}$ gives $\hat{\theta} = \frac{2}{\bar{x}+1}$.

3.2 Proposed Model: Model 2

A probability distribution for the number of male migrants (aged 15 years and above) from a household has been derived under the following assumptions:

– Let X be a random variable denoting the number of male migrants whose conditional distribution is a size-biased Poisson distribution given by

$$P(X=x|\lambda) = \frac{exp(-\lambda)(\lambda^{x-1})}{(x-1)!} \quad ;x=1,2,3...,\lambda>0 \tag{5}$$

Where λ is the parameter denoting risk of migration.
– The risk of migration varies from migrant to migrant and follows a gamma distribution i.e.

$$g(\lambda) = \frac{a^r exp(-a\lambda)(\lambda^{r-1})}{\Gamma r} \quad ;a>0,\lambda>0,0<x<\infty \tag{6}$$

Under these two assumptions (5) and (6), the probability model for the number of male migrants at household level becomes:

$$\begin{aligned} P_x(\lambda) &= \int_{\lambda=0}^{\infty} P(X=x|\lambda)g(\lambda)d\lambda \\ &= \frac{a^r}{\Gamma r(x-1)!}\int_0^{\infty} exp(-(1+a)\lambda)(\lambda^{x+r-2})d\lambda \\ &= \binom{x+r-2}{x-1}\left(\frac{a}{a+1}\right)^r\left(\frac{1}{a+1}\right)^{x-1} \end{aligned} \tag{7}$$

© 2015 NSP
Natural Sciences Publishing Cor.

Putting $p = \frac{a}{a+1}$ and $q = \frac{1}{a+1}$, the above equation can be written as

$$P_x(\lambda) = \binom{x+r-2}{x-1} p^r q^{x-1} \qquad ;x = 2,1,3... \tag{8}$$

which is the form of size biased negative binomial distribution i.e. the marginal distribution of X is a negative binomial distribution with parameter (r, p). The other form of the above equation is

$$P_x(\lambda) = \binom{-r}{x-1} p^r (-q)^{x-1} \qquad ;x = 1,2,3... \tag{9}$$

If we put$p = \frac{1}{Q}$ and $q = \frac{P}{Q}$, then the equation take the following form

$$P_x(\lambda) = \binom{-r}{x-1} Q^{-r} \left(-\frac{Q}{P}\right)^{x-1} \qquad ;x = 1,2,3... \tag{10}$$

4 Estimation of Parameters

The proposed model contains two parameters r and p which can be estimated using method of moments of estimation.
The mean and variance of the proposed distribution can be obtained as follows:
If we put $X-1 = Y$ in (6) then it takes the form of standard Negative binomial distribution i.e.

$$P_y(\lambda) = \binom{-r}{y} Q^{-r} \left(-\frac{P}{Q}\right)^{y} \qquad ;y = 0,1,2,3... \tag{11}$$

Then moments i.e. mean and variance of the above probability distribution of Y are given by

$$E(Y) = rP, \quad V(Y) = rPQ$$

Then this is implies that;

$$E(X) = rP+1, \quad V(X) = rPQ$$

If we use, $P = \frac{1}{Q}$ and $q = \dfrac{P}{Q}$, then we have;

$$E(X) = \frac{rq}{p} + 1, \quad V(X) = \frac{rq}{p^2}$$

Thus, from the above expressions of mean and variance, the moment estimates for the parameters p and r are given by following expressions.

$$\hat{p} = \frac{\bar{x}-1}{\sigma^2} \tag{12}$$

$$\hat{r} = \frac{(\bar{x}-1)^2}{\sigma^2 - \bar{x} - 1} \tag{13}$$

where; $\bar{x} = E(X)$ and $\sigma^2 = V(X)$

5 Application of the Models

The proposed probability model is fitted using some real data sets collected in two different sample surveys entitled Demographic Survey of Chandauli District (Rural Area) 2001-2002 and Rural Development and Population Growth Survey -1978, both sponsored by the centre of Population Studies , Banaras Hindu University, Varanasi, India.

The model application is done for the distribution of the households in which at least one male aged fifteen years and above has been migrated. First the parameters have been estimated from the observed data and then the expected values have been calculated using them.

© 2015 NSP
Natural Sciences Publishing Cor.

Table 1: Observed and expected number of the households with at least one male migrant according to the number of male migrants aged 15 years and above (survey 2001 data)

Number of Migrants	Number of households		
	Observed	Expected	
		Model 1	Model 2
1	97	91.57	93.56
2	35	44.59	42.29
3	19	16.29	15.85
4	6	7.55	8.3
5	3		
Total	160	160	160
χ^2		3.12	2.065
d.f.		2	1
Estimates of parameters		$\hat{\theta} = 0.7565$	$(\hat{p},\hat{r})$=(0.7021, 1.5175)
p-value		0.2101	0.1507

Table 2: Observed and expected number of the households with at least one male migrant according to the number of male migrants aged 15 years and above (survey 1978 data)

Number of Migrants	Number of households		
	Observed	Expected	
		Model 1	Model 2
1	375	367.07	371.07
2	143	155.07	149.98
3	49	49.13	48.71
4	17	18.72	14.53
5+	6		5.72
Total	**590**	**590**	**590**
χ^2		2.09	0.802
d.f.		2	2
Estimate of parameters		$\hat{\theta} = 0.7888$	$(\hat{p},\hat{r})$=(0.7546,1.6471)
p-value		0.3517	0.6697

In the tables 1 to 3 the observed and expected number of the households in relation to the number of male migrants aged 15 years and above is shown for both the models and the calculated χ^2 values are given for fitting of each data set. Also the estimated values of the parameters are given for both the models. For all sets of data the χ^2 values have been found to be insignificant at 5% level of significance.

Tables 1 and 2 show that the expected values provided by proposed model are closer to the observed set of data than model 1 which show that the proposed model presents a better approximation for the data even at different times. Table 3 gives the observed and expected number of the households with at least one male migrant according to the number of male migrants (survey 1978 data) in the three types of households: Semi-Urban, Remote and Growth centre. The details of data are provided in Yadava (1995). In each type of household, proposed model gives better fit than that by Yadava (2004).

It is also clear from the χ^2 values and p-values that the proposed model is more appropriate for describing the pattern of male migration than the previous one.

© 2015 NSP
Natural Sciences Publishing Cor.

Table 3: Observed and expected number of the households with at least one male migrant according to the number of male migrants aged 15 years and above (survey 1978 data) in the three types of households.

Number of migrants	Type of households								
	Semi-Urban			Remote			Growth centre		
	Observed	Expected		Observed	Expected		Observed	Expected	
		Model 1	Model 2		Model 1	Model 2		Model 1	Model 2
1	95	86.54	92.78	176	169.61	175.15	154	145.68	150.65
2	19	31.32	23.41	59	66.81	59.67	47	59.48	53.19
3	10	11.14	7.97	18	19.74	19.28	18	18.21	17.68
4	2		5.01	6	6.85	8.91	9	6.63	8.48
5	3			4			2		
Total	129	129	129	263	263	263	230	230	230
χ^2		7.01	1.26		2.76	0.23		5.98	1.55
d.f.		1	1		2	1		2	1
Estimate of parameters		$\hat{\theta}$=0.819	$\hat{p}$=0.5711 $\hat{r}$=0.5883		$\hat{\theta}$=0.8031	$\hat{p}$=0.6946 $\hat{r}$=1.1154		$\hat{\theta}$= 0.7958	$\hat{p}$=0.6882 $\hat{r}$= 1.1324
p-value		0.0082	0.2618		0.2516	0.6312		0.0503	0.2133

6 Conclusion

The distribution of the rural migration at household level by the number of male migrants has been studied by a size-biased geometric model as proposed by Yadava et al (2004) and a size biased negative binomial distribution with the application of some real data sets. For each data set the proposed model is found to give better fit than the previous model. Therefore we can conclude that the size-biased negative binomial model can be utilized to study the risk of male migration at the household level.

References

[1] C.R. Rao, On discrete distributions arising out of methods of ascertainment in Classic. Contag Discrete Distributions. G.P. Patil, Ed. Statistical Publishing Society, Calcutta, pp. 320- 332,1965.
[2] Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequellcies. Annals of Eugenics 6, 13-25.
[3] G. P. Patil and C. R. Rao, (1978) , Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations, and Human Families, Biometrics, Vol. 34, No. 2, pp. 179-189.
[4] Pryor R.J. (1975), Conceptualization migration behaviour: a problem at micro-demographic analysis in L.A. Kosinski, J.W. Webb, population at micro-scale, New Zeland, Geographical society, Auckland, pp. 105-120.
[5] Yadava, K. N. S. and Yadava,G.S., (1980), On some Probability models and their applications to the distribution of the number of migrants from a household. Janasamkhya, Vol.6(2).
[6] Yadava, K. N. S. and Singh.,S.R.J., (1983), A model for the number of Rural out migrants at households level. Rural Demography, Vol.10, 23-33
[7] Yadava K.N.S., Tripathi, S. and Singh,V.S., (1994), A Probability model for the total number of migrants from the Household, An alternative Journal of Scientific Research, B. H. U., India Vol.49, 125-134.
[8] Yadava K.N.S., Tripathi, S. and K.K. Shukla, (2004), Size-biased geometric distribution and its application to migration data, International Journal of Mathematical Sciences, Vol. 3, pp. 233-239.

© 2015 NSP
Natural Sciences Publishing Cor.