CHAPTER VI

SOME PROBABILITY MODELS FOR OUT-MIGRATION
AT HOUSEHOLD LEVEL

6.1 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. 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.

The first approach of model building in this direction was initiated by Singh and Yadava (1981) who found that the observed distribution of migrants (> 15 years of age) from a household can be described well by the Negative Binomial Distribution (NBD). They proposed this probability distribution under the assumptions that the occurrence of migration from a household is a rare event, and the number of migrants from the household is a random variable and follows a Poisson distribution
and that the risk of migration varies from household to household and follows a Type III distribution. It must, however, be mentioned that the negative binomial distribution is obtained by many sets of assumptions. In one way it can be obtained by the compound Poisson process and in another by the Polya process (Chiang, 1968). Thus there are two different mechanisms at the micro-level which cannot be distinguished by the observed distribution of the random variable (number of migrants).

Yadava and Singh (1983) introduced an idea of cluster for the occurrence of the number of migrants from a household and proposed a model assuming that migrants from a household occur in clusters, and the risk of migration in a cluster varies from household to household and follows a Poisson distribution. Also that the number of migrants to a cluster follows a Poisson distribution increased by one (since a cluster has at least one migrant). The probability distribution under these assumptions is known as the Thomson distribution and is discussed in Johnson and Kotz (1969). Sharma (1984) assumed that the number of migrants per cluster follows the displaced geometric distribution he got the probability distribution for the number of migrants(>15 years of age) as Polya-Aeppli distribution (Jonhson and Kotz, 1969).

Assuming that the number of migrants from a cluster cannot be infinite and at most it may be equal to the size of the household, Yadava and Yadava (1988) took a truncated displaced geometric distribution in place of taking Poisson distribution for the occurrence of number of migrants and truncated a Truncated Polya-Aeppli distribution

However, it was observed that all the households are not equally exposed to the risk of migration at a particular point of time. Keeping this in mind, the present study is an attempt to propose some alternative probability models to describe the pattern of male migrants aged 15 years and above.
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 (Patil, P. G. and Rao, R. C., 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 Patil 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.

In this chapter, two probability models based on the concept of weighted distribution have been proposed to describe the phenomenon. The parameters involved have been estimated using method of moments and their applicability has been checked by some real data of adult members in the household. The models have also been compared with some previous ones.

**6.1.1 SIZE BIASED DISTRIBUTION**

Let $X$ be a non negative random variable having probability function $f(x, \theta)$, with unknown parameter $\theta$, where $\theta \in \Theta$, the parameter space. (pmf when $X$ is discrete). Also suppose an investigator records $x$ of $X$ under $f(x, \theta)$ with a probability proportional to $w(x, \beta)$. Here, $w(x, \beta)$ is a non-negative weight function with parameter $\beta$. Then the recording $x$ will be considered as an observation of rv $X^w$ say with pdf
Some Probability Models for Out-Migration at Household Level

\[ f^w = \frac{w(x, \beta) f(x, \theta)}{\omega} \]

Where \( \omega \) is the normalizing factor which makes the total probability equal to unity.

\( X^w \) is a weighted version of \( X \) by Choosing \( \omega = E[w(x, \beta)] \) and the distribution becomes the weighted distribution with weight function \( w \).

Taking \( w(x, \beta) = X \), \( X^w = X \) is known as Size-biased distribution with probability function as

\[ f^*(x, \theta) = \frac{xf(x, \theta)}{\mu} \]

Where \( \mu = E(x) \) and \( f^* \) gives the size biased probability function of \( f \).

6.2 Probability models for male migrants aged 15 years and above

6.2.1 MODEL-I

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

\[ P(X = x) = x\theta^2.(1 - \theta)^{x-1}; \quad x = 1, 2, 3... \quad (6.1) \]

\[ \theta > 0 \]

Where \( \theta \) is the risk of migration.
6.2.2 MODEL-II

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

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

$$P(X = x \mid \lambda) = \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!}; \quad x = 1, 2, 3 \ldots \ldots \quad (6.2)$$

Where $\lambda$ denotes average number of migrants.

(ii) The average number of migrants varies from migrant to migrant and follows a gamma distribution i.e.

$$g(\lambda) = \frac{a^r}{\Gamma(r)} e^{-a\lambda} \lambda^{r-1}; \quad a > 0, \lambda > 0, \ 0 < x < \infty \quad (6.3)$$

Under these two assumptions (6.2) and (6.3), the probability model for the number of male migrants at household level becomes:

$$P(X = x) = P_x(\lambda) = \int_0^\infty P(X = x \mid \lambda) g(\lambda) d\lambda$$

$$= \frac{a^r}{\Gamma(r)(x-1)!} \int_0^\infty e^{-(1+a)\lambda} \lambda^{x+r-2} d\lambda$$

$$= \frac{a^r}{\Gamma(r)(x-1)!} (1+a) e^{-x/1+a} \left( \frac{1}{1+a} \right)^{x-1} \quad (6.4)$$

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

$$P(X = x) = x^{-1+r-1} C_{x-1} p^r q^{x-1}; \quad x = 1, 2, 3 \ldots \quad (6.5)$$
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

\[
P(X = x) = \binom{x-1}{r-1} p^r (q^{x-1}) ; \quad x = 1,2,3\ldots
\]  

(6.6)

Put \( p = \frac{1}{Q} \) & \( q = \left( \frac{P}{Q} \right) \) the equation (6.6) takes the form

\[
P(X = x) = \binom{x-1}{r-1} Q^{-r} \left( -\frac{P}{Q} \right)^{x-1} ; \quad x = 1,2,3\ldots
\]  

(6.7)

### 6.2.3 MODEL-III

Yadava and Hossain (1998) proposed a logarithmic series distribution in place of geometric distribution for the purpose assuming that probability of migrating one male from a household is greater than the probability of migrating two males and probability of migrating two males is greater than the probability of migrating three males and so on.

The logarithmic series distribution (LSD) characterized by a parameter \( \alpha \) is given by

\[
P(X = x) = \frac{1}{\log(1-\alpha)} \frac{\alpha^x}{x} ; \quad x=1,2,3,\ldots, \alpha>0
\]  

(6.8)

The probability model given in equation (6.8) is a limiting form of zero-truncated negative binomial distribution.

The mean of this distribution is given as

\[
\mu_i = \frac{\alpha}{(1-\alpha) \log(1-\alpha)}
\]  

(6.9)

A size-biased logarithmic distribution (SBLSD) is obtained by taking the weight of the LSD given in equation (6.8) as \( x \).

Thus we have from equations (6.8) and (6.9)
Some Probability Models for Out-Migration at Household Level

\[ P(X = x) = (1 - \alpha)\alpha^{x-1} \]  

(6.10)

6.3 ESTIMATION OF PARAMETERS

In this section, some estimation procedure for the model under consideration will be discussed for the number of migrants in the household. For this purpose the method of moment estimation procedure has been considered to estimate the parameters involved in the models.

6.3.1 MODEL-I

The first model contains only one unknown parameter \( p \) which is estimated by the method moments as follows:

\[ E(x) = \sum_{x=1}^{n} xP(X = x) = \sum_{x=1}^{\infty} x \cdot x\theta^2 (1 - \theta)^{x-1} \]  

(6.11)

Solving the above equation (6.11), we get

\[ E(x) = \frac{2\theta}{(1 - \theta)} + 1 \]  

(6.12)

Solution of \( E(x) = \bar{x} \) gives the estimate of \( \theta \) as

\[ \hat{\theta} = \frac{2}{\bar{x} + 1} . \]  

(6.13)

6.3.2 MODEL-II

The proposed SBNB 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 equation (6.7), then it takes the form of standard Negative binomial distribution i.e.

\[ P(Y = y) = \gamma^{-\gamma}C_y Q^{-\gamma} \left( -\frac{P}{Q} \right)^y ; \quad y = 0, 1, 2, 3 .... \]  

(6.14)

This has first and second moments as
Some Probability Models for Out-Migration at Household Level

\[ E(Y) = rP \quad \text{or} \quad E(X - 1) = rP \]

\[ \Rightarrow E(X) = rP + 1 \quad \text{(6.15)} \]

and

\[ \text{Var}(Y) = rPQ \quad \text{or} \quad \text{Var}(X - 1) = rPQ \]

\[ \Rightarrow \text{Var}(X) = rPQ \quad \text{(6.16)} \]

Putting the values \( p = \frac{1}{Q} \) & \( q = \frac{p}{Q} \) in the above two moments viz. mean and variance of \( Y=X-1 \), given in equations (6.15) and (6.16), we get

\[ E(X) = \frac{rq}{p} + 1 \quad \text{and} \quad \text{Var}(X) = \frac{rq}{p^2} \]

Solution of \( E(X) = \bar{x} \) and \( \text{Var}(X) = \sigma^2 \) gives

\[ \hat{p} = \frac{\bar{x} - 1}{\sigma^2} \quad \text{and} \quad \hat{r} = \frac{(\bar{x} - 1)^2}{(\sigma^2 - \bar{x} - 1)} \]

6.3.3 MODEL-III

The SBLS model contains only one parameter \( \alpha \) which can easily be estimated using method of moments of estimation which is given below

\[ E(x) = \sum_{x=1}^{\bar{x}} xP(X = x) = \sum_{x=1}^{\bar{x}} x(1-\alpha)^x \alpha^{x-1} \quad \text{(6.17)} \]

Solution of \( E(X) = \bar{x} \) gives

\[ \bar{x} = \frac{\alpha}{1-\alpha} \]

Thus we get the estimate of \( \alpha \) as

\[ \hat{\alpha} = 1 - \frac{1}{\bar{x}} \]
6.4 APPLICATION OF THE MODELS

The proposed probability models have been applied on 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 models are also applied to the data used by Hossain (2000) and Aryal (2002) from the surveys conducted in some rural areas of Comilla district of Bangladesh (1997) and a sample survey of the Rupandehi and Palpa districts in Nepal.

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.

In the tables 6.1 to 6.5 the observed and expected number of the households in relation to the number of male migrants aged 15 years and above is shown for each model and the calculated $\chi^2$ values are given for fitting of each data set. Also the estimated values of the parameters are given for the models.

For all sets of data the $\chi^2$ values have been found to be insignificant at 5% level of significance which shows that the proposed probability model presents a good approximation for the data used. It is also clear from the $\chi^2$ values and p-values that the proposed models are more appropriate for describing the pattern of male migration than the previous one. Plots of the observed and expected frequencies of the number of migrants have also been given for a better and quick analysis of applicability of various models used in this chapter.
It is observed that the given models (SBNB and SBLS) give better fitting for each dataset used as compared to SBG model. although by comparing, it cannot be said that which of these models (SBNB and SBLS) can be more useful to study the phenomenon under consideration as for some data sets, SBNB gives better estimates than SBLS and in rest of the cases, SBLS overcome SBNB. But we can observe that both the models are appropriate for the study of migration trend for males from household as for all sets of data the $\chi^2$ values have been found to be insignificant at 5% and 1% level of significance for these two models.

6.5 CONCLUSION

In this chapter, two size biased probability distributions have been proposed to portray the pattern of male migrants fifteen years and old at household level. Their suitability has been tested with numerous real data sets taken from various sample surveys. These probability models are found to fit the data satisfactory well. Therefore we can conclude that these models can be utilized to study the pattern of male migration at the household level.
Table 6.1: Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi

| Number of migrants | Number of households | Observed | Expected | | | | | Model 1 | Model 2 | Model 3 |
|-------------------|---------------------|---------|---------| | | | | | | | |
| 1                 | 97                  | 91.57   | 93.56   | 97.34 |
| 2                 | 35                  | 44.59   | 42.29   | 38.12 |
| 3                 | 19                  | 16.29   | 15.85   | 14.93 |
| 4                 | 6                   | 7.55    | 8.30    | 9.61  |
| 5                 | 3                   |         |         |       |
| Total             | 160                 | 160     | 160     | 160   |
| $\chi^2$          |                     | 3.12    | 2.065   | 1.405 |
| d.f.              |                     | 2       | 1       | 2     |
| Estimate of parameters | $\hat{\theta}=0.7565$ | $\hat{p}=0.7021$ | $\hat{\alpha}=0.3916$ |
| p-value           |                     | 0.2101  | 0.1507  | 0.4953 |

Data Source: Survey 2001

Fig. 6.1 Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi
Table 6.2. Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi

<table>
<thead>
<tr>
<th>Number of migrants</th>
<th>Observed</th>
<th>Number of households</th>
<th>Expected</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>375</td>
<td>367.07</td>
<td>371.07</td>
<td>384.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>143</td>
<td>155.07</td>
<td>149.98</td>
<td>134.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>49.13</td>
<td>48.71</td>
<td>46.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>18.72</td>
<td>14.53</td>
<td>16.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+</td>
<td>6</td>
<td>5.72</td>
<td></td>
<td>8.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>590</td>
<td>590</td>
<td>590</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \chi^2 \] 2.09 0.802 1.818

\[ \text{d.f.} \] 2 2 3

Estimate of parameters
\[ \hat{\theta} = 0.7888 \]
\[ \hat{\rho} = 0.7546 \]
\[ \hat{\delta} = 0.3488 \]

\[ p\text{-value} \] 0.3517 0.6697 0.6111

Data Source: Survey 1978

Fig. 6.2 Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi
### Table 6.3. Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi in the three types of households

<table>
<thead>
<tr>
<th>Number of migrants</th>
<th>Semi-Urban</th>
<th></th>
<th></th>
<th>Remote</th>
<th></th>
<th></th>
<th>Growth centre</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
<td>Expected</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
</tr>
<tr>
<td>1</td>
<td>95</td>
<td>86.54</td>
<td>92.78</td>
<td>89.47</td>
<td>176</td>
<td>169.61</td>
<td>175.15</td>
<td>176.45</td>
<td>154</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>31.32</td>
<td>23.41</td>
<td>27.42</td>
<td>59</td>
<td>66.81</td>
<td>59.67</td>
<td>58.07</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>7.97</td>
<td>11.14</td>
<td>6</td>
<td>18</td>
<td>19.74</td>
<td>19.28</td>
<td>19.11</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5.01</td>
<td>12.11</td>
<td>6</td>
<td>6.85</td>
<td>7.81</td>
<td>6.29</td>
<td>9</td>
<td>6.63</td>
</tr>
<tr>
<td>5+</td>
<td>3</td>
<td>4</td>
<td></td>
<td>4</td>
<td>3.08</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>263</td>
<td>263</td>
<td>263</td>
<td>230</td>
<td>230</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>7.01</td>
<td>1.26</td>
<td>3.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.76</td>
<td>0.23</td>
</tr>
<tr>
<td>d.f.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Estimate of parameters</td>
<td>$\hat{\theta} = 0.819$</td>
<td>$\hat{p} = 0.5711$</td>
<td>$\hat{\alpha} = 0.3065$</td>
<td>$\hat{\theta} = 0.8031$</td>
<td>$\hat{p} = 0.6946$</td>
<td>$\hat{\alpha} = 0.3291$</td>
<td>$\hat{\theta} = 0.7958$</td>
<td>$\hat{p} = 0.6882$</td>
<td>$\hat{\alpha} = 0.3391$</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0082</td>
<td>0.2618</td>
<td>0.0573</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2516</td>
<td>0.6312</td>
</tr>
</tbody>
</table>

Data Source: Survey 1978
**Some Probability Models for Out-Migration at Household Level**

**Fig. 6.3** Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi in Semi-Unban Households

**Fig. 6.4** Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi in Remote areas

**Fig. 6.5** Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Varanasi in Growth Centre households
Table 6.4. Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Nepal

<table>
<thead>
<tr>
<th>Number of migrants</th>
<th>Number of households</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Model 1</td>
</tr>
<tr>
<td>1</td>
<td>126</td>
<td>120.83</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>47.92</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>14.25</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3.77</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.93</td>
</tr>
<tr>
<td>6+</td>
<td>1</td>
<td>0.29</td>
</tr>
<tr>
<td>Total</td>
<td>188</td>
<td>188</td>
</tr>
</tbody>
</table>

\( \chi^2 \)  \hspace{1cm} 1.8711  \hspace{1cm} 0.5657  \hspace{1cm} 0.2794

\( \) d.f. \hspace{1cm} 2  \hspace{1cm} 1  \hspace{1cm} 2

Estimate of parameters

\( \hat{\theta} = 0.8017 \hspace{0.5cm} \hat{p} = 0.6596 \hspace{0.5cm} \hat{\alpha} = 0.3164 \)

p-value \hspace{1cm} 0.3924  \hspace{0.5cm} 0.4520  \hspace{0.5cm} 0.8696

Data Source: Aryal (2002)

Fig. 6.6. Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Nepal
**Table 6.5.** Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Bangladesh

<table>
<thead>
<tr>
<th>Number of migrants</th>
<th>Observed</th>
<th>Expected</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>296</td>
<td>283.41</td>
<td>292.57</td>
<td>292.32</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>79</td>
<td>98.97</td>
<td>86.59</td>
<td>86.91</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>25.92</td>
<td>25.82</td>
<td>25.84</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>6.03</td>
<td>7.72</td>
<td>7.68</td>
<td></td>
</tr>
<tr>
<td>5+</td>
<td>3</td>
<td>1.66</td>
<td>3.30</td>
<td>3.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>416</td>
<td>416</td>
<td>416</td>
<td>416</td>
<td></td>
</tr>
</tbody>
</table>

\[ \chi^2 \]

| d.f. | 2 | 1 | 2 |

<table>
<thead>
<tr>
<th>Estimate of parameters</th>
<th>$\theta = 0.8254$</th>
<th>$\hat{\beta} = 0.6995$</th>
<th>$\hat{\alpha} = 0.2973$</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.0252</td>
<td>0.2767</td>
<td>0.5333</td>
</tr>
</tbody>
</table>

Data Source: Hossain (2000)

**Fig. 6.7.** Observed and expected frequency of the number of households according to the number of male migrants aged 15 years and above in Bangladesh