ISSN: 2469-9062 (print), 2467-9240(e)

*Journal of Institute of Science and Technology, 2015, **20(1):** 1-5, © IOST, Tribhuvan University*

Stochastic Analysis of Infant Deaths by Age and Estimation of Parameters

Brijesh P. Singh[1], Shweta Dixit[2], Tapan Kumar Roy[3]

[1]*Faculty of Commerce & DST-Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University, Varanasi, India, E-mail:brijesh@bhu.ac.in*

[2]*Department of Statistics, Banaras Hindu University, Varanasi, India. Email:sdshwetadixit@gmail.com*

[3]*Department of Population Sciences & HRD, Rajshahi University, Rajshahi, Bangladesh*

ABSTRACT

Infant Mortality Rate (IMR) is a sensitive and powerful index of development. Birth and death, registered through vital registration system in the developing countries suffer from age misreporting, omissions and under count. To overcome this defectiveness in data and to obtain reliable estimates of birth and death rates, India introduced Sample Registration System (SRS) in 1960 but still they suffer from considerable degree of errors. It is observed in retrospective surveys that events are misreported due to ignorance and digit preference of the respondents. Thus, the data on deaths collected, suffer from one defect or other as mentioned above. To resolve this problem attempts have been made to develop and fit suitable models to data on age distribution of deaths. In this paper an attempt has been made to develop a model with an idea of the majority of infant deaths occurs within the first month of their life. The model is used to give a functional shape to the phenomenon of infant deaths distribution and apply on real data taken from NFHS.

Keywords: Infant death, probability models, maximum likelihood, Bayesian estimation.

INTRODUCTION

It is well known that among the many available indicators of socio-economic development, Infant Mortality Rate (IMR) is used as a sensitive and powerful index of development. Reduction in IMR is also known to reduce the fertility as probability of survivorship of children increases. It is observed in retrospective surveys that events are misreported due to ignorance and digit preference of the respondents. Thus, the data on deaths collected, suffer from one defect or other as mentioned above. To resolve this problem attempts have been made to develop and fit suitable models to data on age distribution of deaths. In this paper an attempt has been made to develop a model with an idea of the majority of infant deaths occurs within the first month of their life. The model is used to give a functional shape to the phenomenon of infant deaths distribution and apply on real data. Decline in mortality has been quite rapid because of better medical facilities, but in developing countries, the rate of child and infant mortality is still very high in comparison to developed countries. Thus, the primary concern for these countries is to reduce the level of infant as well as child mortality. The levels and trends of infants and child mortality are often taken as indicators of the development of a country. Infant and child mortality is an excellent measure of the level and quality of health care and other social activities prevailing in a population. A low infant and child mortality reflects availability of good health services and its proper utilization by community. Child mortality is measured by the probability of dying between the first and fifth birthday, where as infant mortality is measured by the probability of death before the first birthday. Indian government has launched several child survival programs over the past decades.

Due to these programs, perhaps, initially a rapid decline in infant and child mortality was noted but during the last decade of twentieth century the rate of decline was very slow and at present, it is more or less stagnant. During last fifty years, studies on early age death were mostly related to infant mortality. But it has been increasingly realized that child loss i.e. mortality under age five needs to be examined in addition to infant mortality. The most common problem in such studies is associated with the data of deaths during infancy and childhood which suffer from substantial degree of errors. Usually errors occur due to recall lapses which result in omission of events, misplacements of dates and distortion of reports on duration of vital events. In such situations development of probability models is perhaps the most appropriate way to minimize the effects of these errors. Actually, a probability model smoothes the data and provide a reasonable explanation of phenomenon under study.

The representation of mortality data via a parametric model has attracted the attention of demographers and statisticians for over a century. Perhaps, the most popular model was given by Gompertz (1825) to study the mortality, which is still used by demographers today. Renshaw (1991) presented generalized linear and nonlinear model approach to mortality graduation and provided arguments in their favour. An attempt to represent mortality across the entire age range was eight parameter non-linear models of Heligman and Pollard (1980). These models have been used in the past for a wide range of mortality data. Keyfitz (1977) might be the first person to attempt a study on infant mortality by using hyperbolic function. Later, for study of infant and child mortality, Hartman (1982) proposed a logarithmic approximation and weibull function was recommended by Cheo (1981). Krishnan and Jin (1993) fitted the Pareto distribution on national data of Canada for studying distribution of infant deaths. Chauhan (1997) suggested the use of finite range model for describing distribution of infant mortality of various state of India.

In this study our aim is to develop a suitable discrete probability model for analyzing the risk of infant mortality in the population and to use it for comparing the risk of infant mortality between different populations. The application of the proposed models is illustrated through real data of four states of India such as Uttar Pradesh, Madhya Pradesh, Bihar and Rajasthan taken from National Family Health Surveys (NFHS-III) (http://www.nfhsindia.org)

MODEL

Probability models provide concise and clear representations of extensive data sets in a better way thus an attempt has been made about the proposition and derivation of probability models for the distribution of infant mortality, which is the combination of neonatal death and post neonatal death. Neonatal death occurs within the month after birth and post neonatal death is occurs after one month of survival and before first birthday. There is more chance of neonatal death of a child than post neonatal death. We assume geometric distribution for the number of months required for infant deaths, but this distribution does not fit to the data because of high number of deaths occurred/reported with in the first month after birth. Thus we need to inflate the above distribution at zero in order to get a better explanation.

- At any point in time, let α be the proportion of the death before reaching first birthday and (1-α) be the proportion of extra deaths within the first month. These deaths may be post neonatal death reported as neonatal deaths, still birth, induced or spontaneous abortion which recorded as a death within the month.
- If p represents the probability of a death and the pattern of infant death follows the geometric distribution.

If x represents the number of months required for infant death, thus under the above assumption the distribution of infant death follows with probability density function as

$$\left.\begin{aligned} P(\mathrm{x}=0) &= 1-\alpha+\alpha\,\mathrm{p} \\ P(\mathrm{x}=\mathrm{k}) &= \alpha\, q^{k} p \quad \text{for k} = 1,\ 2,\ 3,\ \ldots \end{aligned}\right\} \qquad (1)$$

where p+q=1. This distribution is known as inflated geometric distribution.

ESTIMATION OF THE PARAMETERS

Method of moment

Inflated geometric distribution has two parameters α and p to be estimated. Let we have

$$E(x) = \frac{\alpha(1-p)}{p} = \bar{x} \qquad (2)$$

And, $$E(x^2) = \frac{\alpha(1-p)(2-p)}{p^2} \qquad (3)$$

Thus variance =

$$E(x^2) - [E(x)]^2 = \frac{\alpha(1-p)(2-p)}{p^2} - [\frac{\alpha(1-p)}{p}]^2 \qquad (4)$$

From (2) and (3) we get

$$E(x^2) = \bar{x}.[\frac{2-p}{p}] \Rightarrow \frac{2-p}{p} = \frac{E(x^2)}{\bar{x}}$$

$$\Rightarrow \frac{2}{p} - 1 = \frac{E(x^2)}{\bar{x}} \Rightarrow \frac{2}{p} = \frac{E(x^2)+\bar{x}}{\bar{x}}$$

$$\hat{p} = \frac{2\bar{x}}{(E(x^2)+\bar{x}} \qquad (5)$$

$$or,\ \hat{p} = \frac{2\bar{x}}{\text{variance} + \bar{x}^2 + \bar{x}}$$

Putting the estimate of p in equation (2) we get the estimate of α, i.e.

$$\hat{\alpha} = \frac{p\bar{x}}{1-p} = \frac{2\bar{x}^2}{\text{variance} + \bar{x}^2 - \bar{x}}$$

Maximum likelihood estimate

Let ($x_1, x_2, \ldots\ldots x_n$) denote a random sample of size n. Each x_i count the number of month required for first conception. Assuming that $n_k\ (k = 1, 2, 3 ....m)$ denotes the number of observations with value *k*. The likelihood function of estimating the parameters α and *p* can be expressed as bellow:

$$L = (1-\alpha+\alpha p)^{n_0} \prod_k (\alpha p q^k)^{n_k} \tag{6}$$

$$= (1-\alpha+\alpha p)^{n_0} \alpha^{n-n_0} p^{n-n_0} q^S \qquad S = \sum_{k=1}^{m} k n_k$$

Taking logarithms and differentiating with respect to α and *p* respectively and equating to zero, we have

$$\frac{\partial \log L}{\partial \alpha} = \frac{n_o(p-1)}{(1-\alpha+\alpha p)} + \frac{n-n_o}{\alpha} = 0 \ and \ \frac{\partial \log L}{\partial p} = \frac{n-n_o}{p} - \frac{s}{1-p} + \frac{n_o \alpha}{(1-\alpha+\alpha p)} = 0$$

On solving we easily estimate

$$\hat{\alpha} = \frac{n-n_0}{n(1-p)} \quad and \quad \hat{p} = \frac{n-n_0}{\sum_{k=1}^{m} n_k k} = \frac{n-n_0}{n\overline{x}} \tag{7}$$

Method of zero cell frequency

We know that from equation (2) and also we have

$$\frac{n_0}{n} = 1-\alpha+\alpha p = 1-\alpha(1-p) = 1-p\overline{x} \tag{8}$$

$$p\overline{x} = \frac{n-n_0}{n} \Rightarrow \hat{p} = \frac{n-n_0}{n\overline{x}} \quad and \quad \hat{\alpha} = \frac{n-n_0}{n(1-p)} \tag{9}$$

This method is easier and quicker to get the estimate. In the above method we need *E(x)*, $E(x^2)$ or *Var(x)* but in this method we need only $\overline{x}$ and the estimates are equivalent to the maximum likelihood estimate of *p* and α.

Bayesian Estimation

We know that the distribution of infant death is

$$P(X=k) = \begin{cases} (1-\alpha)+\alpha p & if\ k=0 \\ \alpha p q^k & if\ k = 1, 2, ... \end{cases}$$

$$0 < \alpha < 1 \ and \ 0 < p < 1$$

Likelihood is as follows

$$L = (1-\alpha+\alpha p)^{n_0} \prod_k (\alpha p q^k)^{n_k} \qquad S = \sum_{k=1}^{m} k n_k$$

$$= (1-\alpha+\alpha p)^{n_0} \alpha^{n-n_0} p^{n-n_0} q^S$$

Using Binomial expression, L becomes

$$L = \sum_{r=0}^{n_0} {}^{n_0}_{r}c\,(1-\alpha)^r (\alpha - p)^{n-n_0} p^{n-n_0} q^S \quad = \sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\alpha^{n_0-r+n-n_0} (1-\alpha)^r p^{n_0-r+n-n_0} q^S$$

$$= \sum_{r=0}^{n_0} {}^{n_0}_{r}c\,(1-\alpha)^r \alpha^{n-r} p^{n-n_0} q^S \tag{10}$$

Since α and *p* lies between 0 and 1 so that we may take conjugate prior i.e., Beta distribution *α~β(a,b) and p~β(c,d)*. Also α and *p* are independent then the Joint prior is

$$M(\alpha, p) = \frac{\alpha^{a-1}(1-\alpha)^{b-1} p^{c-1}(1-p)^{d-1}}{\beta(a,b)\beta(c,d)} \tag{11}$$

Now the joint posterior is

$$M(\alpha, p) = \frac{L\,M(\alpha, p)}{\int_0^1\int_0^1 L\,M(\alpha,p)\,d\alpha\ dp}$$

$$= \frac{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\alpha^{n-r}(1-\alpha)^r p^{n-r} q^S \alpha^{a-1}(1-\alpha)^{b-1} p^{c-1} q^{d-1}}{\sum_{r=0}^{n_0} {}^{n_0}_{r}c \int_0^1\int_0^1 \alpha^{n-r}(1-\alpha)^r p^{n-r} q^S \alpha^{a-1}(1-\alpha)^{b-1} p^{c-1} q^{d-1} dp\, d\alpha} \tag{12}$$

$$= \frac{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\alpha^{n-r+a-1} p^{n-r+c-1}(1-p)^{S+d-1}(1-\alpha)^{b-1+r}}{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\beta(n-r+a, b+r), \beta(n-r+c, s+d)}; \quad q = 1-p \tag{13}$$

and the marginal posterior is

$$M_1(\alpha) = \int_0^1 M(\alpha,p)dp \ = \frac{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\alpha^{n-r+a-1}(1-\alpha)^{b+r-1}\beta(n-r+c, s+d)}{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\ \beta(n-r+c, s+d)\ \ \beta(n-r+a, b+r)} \tag{14}$$

$$M_2(p) = \int_0^1 M(\alpha,p)d\alpha \tag{15}$$

$$= \frac{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\ p^{n-r+c-1}(1-p)^{d+S-1}\beta(n-r+a, b+r)}{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\ \ \beta(n-r+a, b+r)\ \ \beta(n-r+c, d+s)}$$

Thus the Bayes estimate under squared error loss function

$$\hat{\alpha} = \int_0^1 \alpha M_1(\alpha)dp \ = \frac{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\beta(n-r+a-1, b+r)\beta(n-r+c, s+d)}{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\ \ \beta(n-r+c, s+d)\ \ \beta(n-r+a, b+r)} \tag{16}$$

$$\hat{p} = \int_0^1 pM_2(p)dp \ = \frac{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\,\beta(n-r+a, b+r)\beta(n-r+c+1, s+d)}{\sum_{r=0}^{n_0} {}^{n_0}_{r}c\ \ \beta(n-r+a, b+r)\ \ \beta(n-r+c, s+d)}$$

APPLICATION AND INTERPRETATION

We apply the proposed distribution to data taken from NFHS-III, which was collected in 2005-06. The analysis is based on the data taken for three year prior to the survey. The expected frequencies for each state have been obtained only through maximum likelihood method and Bayesian method in the present study. The expected number of infant deaths by age is very close to the observed number of infant death. Also the value of chi-square and *p*-value indicates that the proposed model explain remarkably the data on infant deaths by age in all the four states of India but estimate obtained by Bayesian method is more close to the observed value. Maximum likelihood estimate of the risk of infant mortality (*p*) is 283 per thousand per three year in Uttar Pradesh, 215 per thousand per three year in Madhya Pradesh, 235 per thousand per three year in Bihar and 194 per thousand per three year in Rajasthan. Thus infant mortality is about 94, 72, 78 and 65 per thousand per year in Uttar Pradesh, Madhya Pradesh, Bihar and Rajasthan respectively. As far as the Bayesian estimate of the risk of infant mortality (*p*) is concerned, 243 per thousand per three year in Uttar Pradesh, 212 per thousand per three year in Madhya Pradesh, 221 per thousand per three year in Bihar and 142 per thousand per three year in Rajasthan so that infant mortality per year is about 81, 71, 74 and 47 in Uttar Pradesh, Madhya Pradesh, Bihar and Rajasthan respectively. In Uttar Pradesh and Rajasthan estimate of risk of infant mortality is remarkably different for both the estimation procedure.

The another estimate α be the proportion of death before reaching first birth is about 45, 39, 44 and 39 percent in Uttar Pradesh, Madhya Pradesh, Bihar and Rajasthan respectively and $(1-\alpha)$ be the proportion of extra infant death within a month after birth is 55, 61, 56 and 61 percent for Uttar Pradesh, Madhya Pradesh, Bihar and Rajasthan respectively. In fact as discussed above the infant mortality is the combination of neonatal and post neonatal mortality. This indicates that the death during first month i.e. proportion of neonatal death is major part of the infant death that is not captured by the simple distribution but this can be possible with modified distribution i.e. inflated distribution. Thus proposed model is a suitable choice for understanding this phenomenon. Also the present model could be extended by considering the risk of infant deaths varies according to the beta distribution.

REFERENCES

Chauhan, R.K. 1997. Graduation of infant death by age. *Demography India* **26**: 261-274.

Cheo, M.K. 1981. Fitting of age pattern of infant and child mortality with the Weibull Survival Function. *Asian and Pacific Census Forum* **7**: 10-13.

Gompertz, B. 1825. On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies. *Philosophical Transactions* **11527**: 513-585.

Hartman, M. 1982. Infant and Child Mortality. Department of Statistics, University of Lund. Sweden.

Heligman, L. and Pollard, J.H. 1980. Age pattern of mortality. *Journal of Institute of Actuaries* **107**: 49-80.

Keyfitz, N. 1977. Introduction to the mathematics of the population with revisions. Addition-Wesley Publishing Company, Reading, Massachusetts.

Krishnan, P. and Jin, Y. 1993. A statistical model for infant mortality. Paper presented at the IUSSP Meeting, Montreal, Canada, August 25-September 1.

National Family Health Survey-III 2005-06. International Institute of Population Sciences, Mumbai.

Renshaw, A.E. 1991. Actuarial graduation practice and generalized linear and nonlinear models. *Journal of Institute of Actuaries* **118**: 295-312.

Table 1. Observed and expected number of infant death by age in Uttar Pradesh

Age at deaths (in month)	Observed number of infant death	Expected number of infant death	
		ML method p =0.2830 α =0.4499	Bayesian Method p =0.2429 α =0.4237
0	315	315.00	315.85
1	51	42.45	36.23
2	21	30.44	27.43
3	18	21.82	20.77
4	15	15.65	15.72
5	10	11.22	11.90
6	12	8.04	9.01
7	5	5.77	6.82
8	6	4.13	5.17
9	5	2.96	3.91
10	4	2.13	2.96
11	3	5.38	9.23
Total	465	465.00	465.00
Chi-square		8.311	10.896
p-value		0.216	0.143

Table 2. Observed and expected number of infant death by age in Madhya Pradesh

Age at deaths (in month)	Observed number of infant death	Expected number of infant death	
		ML method p =0.2154 α =0.3900	Bayesian Method p =0.2124 α =0.3878
0	127	127.00	127.11
1	11	12.06	11.87
2	6	9.46	9.35
3	9	7.43	7.36
4	7	5.83	5.80
5	4	4.57	4.57
6	5	3.59	3.60
7	1	2.81	2.83
8	4	2.21	2.23
9	1	1.73	1.76
10	4	1.36	1.38
11	4	4.95	5.14
Total	183	183.00	183.00
Chi-square		2.081	1.979
p-value		0.556	0.577

Table 3. Observed and expected number of infant death by age in Bihar

Age at deaths (in month)	Observed number of infant death	Expected number of infant death	
		ML method p =0.2347 α =0.4420	Bayesian Method p =0.2214 α =0.4335
0	90	90.00	90.09
1	13	10.80	10.16
2	6	8.26	7.91
3	5	6.32	6.16
4	3	4.84	4.80
5	3	3.70	3.74
6	6	2.83	2.91
7	1	2.17	2.26
8	1	1.66	1.76
9	4	1.27	1.37
10	2	0.97	1.07
11	2	3.17	3.76
Total	136	136.00	136.00
Chi-square		1.439	1.479
p-value		0.487	0.477

Table 4. Observed and expected number of infant death by age in Rajasthan

Age at death (in month)	Observed number of infant death	Expected number of infant death	
		ML method p =0.1942 α =0.3878	Bayesian Method p =0.1419 α =0.3696
0	88	88.00	87.40
1	5	7.77	5.76
2	3	6.26	4.94
3	9	5.04	4.24
4	2	4.06	3.64
5	1	3.28	3.12
6	5	2.64	2.68
7	6	2.13	2.30
8	3	1.89	1.97
9	2	1.52	1.69
10	3	1.23	1.45
11	1	4.62	8.79
Total	128	128.00	128.00
Chi-square		5.991	-
p-value		0.050	-

*degree of freedom becomes zero after pooling expected frequency less than 5, thus chi-square is not calculated.