The Nepali Math. Sc. Report
Vol. 26, No. 1 & 2, 2006.

Probability distributions to describe the pattern of child loss from a family

TIKA RAM ARYAL*

Abstract: This paper attempts to investigate tne distribution of child loss from a family through probability models. The suitability of the models has been tested with the real sets of sample survey data from Nepal, Libya and Brazil. Poisson, geometric and displaced geometric distributions have been fitted to describe the distribution of families according to the number of child deaths. The parameters have been estimated by maximum likelihood method. It was found that the proposed geometric and displaced geometric distributions more or less provided a suitable description of child loss pattern at micro level (family level). These distributions fitted satisfactorily well to all the sets of sample data, which may be utilised to predict the risk of child loss in a family in any society. Findings may help planners and policy makers for designing proper policies and programs in a country.

1. Introduction

The force of mortality is still high at the younger ages particularly during the infancy [1] Level and trend of early childhood mortality indicate the standard of development of a country. A high rate of child mortality in a society indicates the reproductive wastages of physical, economical and psychological potential of a woman, and consequently shows a low level of success of a country's health program [2]. The child mortality has been of interest to researchers because of its apparent relationship with fertility and indirect relationship with the acceptance of modern contraceptive means [3]. The distribution of deaths with respect to age during the infancy is usually not governed by any single universal law because there exist a number of distinct patterns, which might be changed over time. It has been increasingly realised for several reasons that child loss from a family needs to be examined besides infant deaths [4]. The reported data of deaths during infancy and childhood suffer from substantial degree of errors where vital registration system of such information is not available [5]. Usually errors occur due to recall lapse, which result in omission of events, misplacement of dates and the distortion of reports on the duration of vital events [2, 5]. To overcome such limitations, models served the purpose and are needed to be developed to remove such variations due to these biases from one age to another. In fact, a model may smooth the data and provides a reasonable distribution of deaths according to age.

A number of attempts have been made to study the age pattern of mortality by using models [2, 6, 7, 8, 9, 10, 11]. Initially, Keyfitz [12] used a hyperbolic function to study the

* Dr. Aryal is associated as Postdoçtoral Research Fellow at the Institute of Economic Research, Hitotsubashi University, Tokyo, Japan and Cenetral Department of Statistics, Tribhuvan University, Kirtipur, Kathmandu Nepal.

E-mail: traryal@rediffmail.com & traryal@ier.hit-u.ac.jp

infant and child mortality. Choe [13] used Weibull function and Hartman [14] applied a logarithmic function for the same purpose. Krishnan and Jin [15] utilised Pareto distribution to describe the distribution of infant deaths and found a reasonable fitting. Chauhan [16] suggested finite range model to study the deaths under age one by months. Finite range model was proposed by Mukherjee and Islam [17] in reliability analysis. Krishnamoorthy and Rajna [18] checked the suitability and modified it for graduating the survivorship function. Bhuyan and Deogratias [19] used a modified Polya Aeppli model to study the pattern of child loss in Northeast Libya. These reflect an increasing application of probability models for describing the pattern of mortality.

An adequate research work on the pattern of child loss from a family has not been done yet in Nepal, which may be due to lack of interest among researchers or lack of reliable data. Child loss from a family has an important indicator for the well-being of the country, in general, and, in particular, well-being of the women health. Thus, this paper attempts to study the distributional pattern of families according to the number of child loss (under 5 years) through some probability models. Various sets of data from Nepal and other countries have also been used to discuss the applicability of the procedures proposed in this paper.

2. Models

As mentioned earlier, some probability distributions have been discussed this section to study the distribution of families according to the number of child deaths (deaths within the first five years of life). The assumption of using probability distributions is that only those families have been considered in which at least one birth prior to the survey date (study point) has had occurred.

2.1. Poisson Distribution

Let X denote the number of child deaths in a family at the survey point which follow a Poisson distribution. The probability mass function of X is

(1) $$p(x=k)=\frac{e^{-\lambda k}\lambda^{k}}{k!} \quad \text{for } x=0,1,2,\ldots$$

when λ is the risk of death of a child in a family. The maximum likelihood equation of (1) is

(2) $$L=\prod_{k=0}^{n}\left[p(x=k)\right]=\prod_{k=0}^{n}\left[\frac{e^{-\lambda k}\lambda^{k}}{k!}\right]$$

Taking logarithms of (2) and differentiating with respect to λ and equating it to zero, we get

(3) $$\frac{\delta \log L}{\delta\lambda}=-n+\frac{n\bar{k}}{\lambda}=0$$

On solving (3), the estimates of λ can easily be obtained as

(4) $$\hat{\lambda}=\bar{k}$$

2.2. Geometric Distribution

Let X denote the number of child deaths in a family at the survey point which follows a geometric distribution. The probability mass function of X is

(5) $$p(x=k)=q^{k}p \quad \text{for } k=0,1,2,3,\ldots$$

It involves a single parameter p to be estimated from the observed distribution of families according to the number of child deaths. The likelihood function can be expressed as

$$L = \prod_{k=0}^{n} [p(x=k)] = \prod_{k=0}^{n} [pq^k] \tag{6}$$

$$L = p^n q^{\sum_{k=0}^{m} k} \tag{7}$$

Taking logarithms of (7) and differentiating with respect to p and equating it to zero, we get

$$\frac{\delta \log L}{\delta p} = \frac{n}{p} - \frac{n\bar{k}}{1-p} = 0 \tag{8}$$

On solving (8), the estimate of p can easily be obtained

$$\hat{p} = \frac{1}{1+\bar{k}} \tag{9}$$

2.3. A Mixture of Two Displaced Geometric Distribution

Let X denote the number of child deaths in a family at the survey point. The distribution of X is derived under some assumptions as (i) only those families are considered in which at least one birth prior to the survey has had occurred, (ii) at the survey point, a family either has experienced a child loss or not and let β and $(1-\beta)$ be the respective proportions, (iii) out of β proportion of families, let ξ be the proportion of families in which only one child death has occurred, and (iv) remaining $(1-\xi)\beta$ proportion of families, experiencing multiple child deaths, which follows a displaced geometric distribution with parameter p according to the number of child deaths.

Hence with these assumptions, the probability distribution of X is written as

$$\begin{aligned} p(x=k) &= 1-\beta && \text{for } k=0 \\ &= \xi\beta && \text{for } k=1 \\ &= (1-\xi)\beta p q^{k-2} && \text{for } k=2,3,\dots \end{aligned} \tag{10}$$

Model (10) involves three parameters ξ, β and p to be estimated from the observed distribution of families according to the number of child deaths. Let $x_1, x_2, \dots, x_N$ denote a random sample of size N from the population (10). Further, suppose that n_k ($k = 0, 1, 2, \dots, m$) be the number of observations corresponding to the value of k and $\sum_{k=0}^{m} n_k = N$. The likelihood function for the given sample can be expressed as

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

$$L = (1-\beta)^{n_0} \xi^{n_1} \beta^{n-n_0} (1-\xi)^{n-n_0-n_1} p^{n-n_0-n_1} q^{\sum_{k=3}^{m}(k-2)n_k} \tag{11}$$

Taking logarithm of (11) and differentiating with respect to β, ξ and p respectively and equating it to zero, we get

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

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

(14) $$\frac{\delta \log L}{\delta p} = \frac{n-n_0-n_1}{p} - \frac{\sum_{k=3}^{m}(k-2)n_k}{1-p} = 0$$

From (12), (13) and (14), the estimates of β, ξ and p can easily be estimated as

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

Variances and covariances of the estimates of the parameters are obtained by taking second partial derivatives of log L as

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

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

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

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

Taking the ract that $E(n_0) = E[\sum_{i=1}^{n} 1_{(X_i=0)}] = \sum_{i=1}^{n} 1\, p(X_i=0) = \sum_{i=1}^{n}(1-\beta) = n(1-\beta)$, in similar way we can write $E(n_1) = n\xi\beta$, $E(n_k) = n(1-\xi)\beta p q^{k-2}$ for $k = 2, 3, \ldots, m.$, $E(n-n_0) = np$, $E(n-n_0-n_1) = n\beta(1-\xi)$ and

$$E[\sum_{i=3}^{m}(k-2)n_k] = E[n_3 + 2n_4 + 3n_5 + \ldots + (m-2)n_m]$$
$$= n(1-\xi)\beta p q\,[1 + 2q + 3q^2 + \ldots + (m-2)q^m]$$
$$= n(1-\xi)\beta q\,[\frac{1-q^{m-2}}{p} - (m-2)q^{m-2}], \text{ for small } m$$
$$= \frac{n(1-\theta)\beta q}{p}, \text{ for large } m,$$

Using above facts, the expected values of the second partial derivatives can be obtained as,

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

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

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

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

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

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

Since $E\left(\frac{\delta^2 \log L}{\delta\beta\delta\xi}\right) = E\left(\frac{\delta^2 \log L}{\delta\xi\delta p}\right) = E\left(\frac{\delta^2 \log L}{\delta\beta\delta p}\right) = 0$, so the covariances between the estimators becomes zero. Hence asymptotic variances of the estimators can be obtained as,

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

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

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

3. Applications

The proposed distributions have been applied to the real sets of sample survey data of Demographic Survey of Fertility and Mobility in Rural Nepal (DSFM): A Study of Palpa and Rupandehi Districts, which was carried out in 2000 [2]. Besides, one set of data has been taken from a household sample survey of Brazil in 1996 [19]. Another set of data from Northeast Libya conducted under a sample survey in 1987 [20].

The parameters of the proposed models have been estimated by the method of maximum likelihood. The observed and expected number of families (along with the estimates of the parameters) according to the number of child deaths is presented in Tables 1 to 3. It is seen that the Poisson distribution does not give a good fit to the data sets whereas the geometric distribution and displaced geometric distribution provided a good fit to all the data sets. The estimated value of β that shows the proportion of families experienced a child loss, was found slightly higher for Libya (0.36) as compared to Brazil (0.27) and Nepal (0.21). However, the proportion of families having a single child death was found higher for Nepal as compared to the data of other countries.

The average number of child deaths per family $[\hat{\xi}\hat{\beta} + (1-\hat{\xi})\hat{\beta}(1+\frac{1}{\hat{p}})]$ for Libya, Brazil and Nepal were found to be 0.53, 0.41 and 0.30 respectively. One of the applications of analyzing such data through model is to estimate the number of child deaths if family size is known. For example, if on an average total fertility rate (TFR) in Nepal is 4.1 (according Nepal Demographic Health Survey, 2001), then expected child mortality would be around 73 per

1000 live births (against the reported mortality rate 65 per 1000 live births) during the preceding (0.4) years from the survey date of NDHS. This is a close estimate of IMR (73) to observed IMR (65). The lower observed value of IMR of NDHS data may be due to under reporting in infant deaths in the survey or two different sources of data, which are not comparable [2]. Thus through such models if the size (TRR) of the family be known, child mortality may be predicted for a given society.

4. Conclusions

It was found that the proposed geometric distribution and displaced geometric distribution provided a suitable description of child deaths at micro level (family level). The distribution fitted satisfactorily well to several sets of sample data of Nepal, Libya and Brazil. So these distributions may be utilized to predict the risk of child deaths from a family in any society of the country. The findings may help planners and policy makers for designing policies and programs of a country.

Acknowledgements

My deepest appreciation and profound gratitude go to my esteemed Postdoctoral Research Supervisor, Professor Satoru Kanoh, Institute of Economic Research, Hitotsubashi University, Tokyo, Japan, for his inspiring guidance, constant guidance, constant encouragement and help in preparing the paper. I would like to thank to the Japan Society for the Promotion of Science (JSPJS), Tokyo, Japan, for offering me Postdoctoral Research Fellowship to carry out research at Japan.

Table 1 Observed And Expected Number Of Families According To The Number Of Child Deaths In Nepal (2000)

No. of dead children	Observed No. of families	Expected No. of families		
		Poisson distribution	Geometric	Displaced geometric
0	669	623.76	649.65	669.24
1	137	195.24	154.87	137.06
2	32		36.92	28.54
3	6			10.45
4	3			
5	2	34.00	11.55	
6	2			5.74
7.	0			
Total	851	851.00	851.00	851.00
χ^2		24.21	4.327	2.591
d.f		1	2	1
		λ=0.2973	p=0.7616	β=0.2139 ξ=0.7527 p=0.6338
				$v(\hat{\beta})$=0.00020 $v(\hat{\xi})$=0.00102 $v(\hat{p})$=0.00327

Table 2 Table 1 Observed And Expected Number Of Families According To The Number Of Child Deaths In North East Libya (1996)

No. of dead children	Observed No. of families	Expected No. of families		
		Poisson distribution	Geometric	Displaced geometric
0	805	740.19	820.67	805.79
1	306	389.04	282.73	306.17
2	93	102.24	97.41	94.26
3	36	20.53	33.56	31.27
4	7		17.39	9.38
5	2			5.41
6	1			
7	2			
Total	1252	1252.00	1252.00	1252.00
χ^2		60.99	4.261	1.367
d.f		2	3	2
		λ=0.5256	p=0.6555	β=0.3570 ξ=0.6846 p=0.6682
				$v(\hat{\beta})$=0.00018 $v(\hat{\xi})$=0.00048 $v(\hat{p})$=0.00106

Table 3 Table 1 Observed And Expected Number Of Families According To The Number Of Child Deaths In North East Brazil (1987)

No. of dead children	Observed No. of families	Expected No. of families		
		Poisson distribution	Geometric	Displaced geometric
0	769	698.12	745.85	769.40
1	185	285.60	216.55	185.07
2	60	58.42	62.87	63.61
3	26	7.97	18.26	21.88
4	9	8.85	7.42	11.05
5	1			
6	1			
7	0			
Total	1051	1051.00	1051.00	1051.00
χ^2		132.16	10.465	0.981
d.f		4	4	1
		λ=0.4091	p=0.7096	β=0.2683 ξ=0.6560 p=0.6554
				$v(\hat{\beta})$=0.00019 $v(\hat{\xi})$=0.00033 $v(\hat{p})$=0.00158

REFERENCES

[1] Aryal, T. R. Some techniques to estimate child mortality in Nepal, *Journal of Institute of Science and Technology*, vol. 13: 48–61, 2004.

[2] Aryal, T. R. *Some demographic models and their applications with reference to Nepal*, Unpublished Ph. D, thesis Department of Statistics, BHU, India, 2003.

[3] Karib, M., Amin, R., and Chowdhury, J., Determinants of infant and child mortality in Bangladesh, Journal of Statistical Studies, vol. 13: 54–64, 1993.

[4] U. N., *Mortality of children under age* 5; World estimates projections, 1950–2025, Population Studies vol. 105, United Nations, New York, U.S.A. 1988.

[5] Aryal, T. R. Some indirect techniques to estimate maternal mortality in Nepal, Journal of Population and Social Studies, vol. 13(1): 61–76, 2004.

[6] Gavrilov, L.A., and Gavrilov, N. S. *The biology of lifespan: A quantitative approach*, Chur, Switzerland: Harwood Academic Publishers, 1991.

[7] Gompertz., B., 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 R.S.*, vol. **115**(27): 513–585, 1825.

[8] Heligman, L., and Pollard, J.H., Age pattern of mortality, *Journal of Institute of Actuaries*, vol. 107: 49–80, 1980.

[9] Mitra, S., and Denny, C., On the application of a model of mortality, *Canadian Studies in Population*, vol. **21**(2): 117–132, 1994.

[10] Perks, W., On some experiments in the graduation of mortality statistics, *Journal of Institute of Actuaries*, vol. 63:12, 1932.

[11] Theile, P. N., On a mathematical formula to express the rate of mortality throughout the whole life, *Journal of the Institute of Actuaries*, vol. 16 : 313, 1972.

[12] Keyfitz, N., *Applied Mathematical Demography*, New York, John Wiley and Sons, 1977.

[13] Choe, M. K., Fitting the age patterns of infant and child mortality with the Weibull survival function: *Asian and Pacific Census Forum*, vol. 7(4): 10–13, 1981.

[14] Hartman, M., *Infant and child mortality*, Dept. of Statistics, University of Lund. Sweden, 1982.

[15] Krishnan, P., and Jin, Y., A statistical model for infant mortality, paper presented at the IUSSP Meetings, Montreal, Canada, August 25–Sept. 1, 1993.

[16] Chauhan, R. K., Graduation of infant deaths by age, *Demography India*, vol. 26(2):261–274, 1997.

[17] Mukherjee, S. P., and Islam, A., A finite range distribution of failure time, *Naval Research Logistic Quarterly*, vol. 30: 487–491, 1983.

[18] Krishnarmoorthy, S., and Rajna, P.N., Graduating childhood deaths and survivorship function with the model by Mukherjee and Islam, *Demography India*, vol. 28(1):13–19, 1999.

[19] Bhuyan, K. C., and Deogratias, R., On a probability model for child mortality pattern in North Eastern Libya, *The Turkish Journal of Population Studies*, vol. 21:33–37, 1999.

[20] Sastry, N., A nested frailty model for survival data, with an application to the study of child survival in Northeast Brazil, Journal of the *American Statistical Association*, vol. 92(438), 1997.

TIKA RAM ARYAL Ph. D.

Postdoctoral Research Fellow,
Institute of Economic Research,
Hitotsubashi University, Tokyo, Japan.
and
Central Department of Statistics,
Tribhuvan University, Kirtipur, Nepal.
Email: traryal@rediffmail.com & traryal@ier.hit-u.ac.jp