A PROBABILITY MODEL FOR INTERIOR BIRTH INTERVAL AND ITS APPLICATIONS

B.N. Bhattacharya and K.K. Singh Centre of Population Studies, Banaras Hindu University, Varanasi, India

Résumé — Un modèle de probabilité pour les intervalles génésiques intérieurs et ses applications. Nous proposons un modèle pour décrire es données selon les intervalles génésiques pour une certaine durée de mariage ou entre duex dates d'enquête. Ces intervalles s'appellent intervalles génésiques intérieurs, et les modèles diffèrent de ceux utilisés pour les intervalles fermés. Nous suggérons une procédure pour déterminer les estimations de probabilité maximum, et le modèle est appliqué à trois séries de données.

Abstract — A model has been proposed to describe the data on birth intervals that lie entirely within a segment of marriage duration or between two survey dates. Such intervals are termed as interior birth intervals, and these models are different to those of the usual closed intervals. A procedure to find out maximum likelihood estimates is outlined and the proposed model has been applied to three sets of data.

 $\it Key Words-interior birth interval, marriage duration, maximum likelihood estimate, risk of conception$

Introduction

Construction of analytical models for birth intervals is receiving increasing attention because of their relationship with population birth rate. Demographers have analyzed birth intervals in order to study fertility differentials across different populations to estimate biological parameters of reproduction such as fecundity and incidence of foetal wastages, and to develop sensitive indices for measuring changes in fertility (see Chakrabarty, 1976; D'Souza, 1973; George, 1967; George and Mathai, 1975; Srinivasan, 1966;) A large number of surveys, especially fertility surveys, conducted in various parts of the world have collected information on the fertility histories of women. However, the study of birth intervals is possible only if relatively accurate dates of marriage and successive births are available. It has been pointed out that retrospective information relating to past events is subject to memory lapses and other sources of bias that often depend on the recall period (Singh et al., 1979; Som, 1973). Thus, information on birth intervals collected through retrospective surveys, relating to births that occurred in the recent past, or through prospective surveys may be reliable. A number of probability models, dealing with straddling, forward and interior birth intervals, have been proposed by Menken and Sheps (1972); Poole (1971); Sheps et al., (1970); and Singh et al. (1978).

The present paper is concerned with the formulation of a model for the interior birth interval, which is a closed birth interval lying entirely within a given segment of marriage duration. It may also lie between two survey dates, for instance, in a prospective enquiry starting at time t_1 and continued until t_2 , the observed closed birth intervals during the time period (t_1, t_2) of length $T = t_2 \cdot t_1$ are called prospective interior birth intervals. Obviously such intervals have a distribution different from those of the usual closed birth intervals.

Model

Let T_0 , T_0+T_1 , $T_0+T_1+T_2$, ... be the time of recording successive births to a married woman since a given time t_I , where T_0 is the time of first recording, since t_I and T_i is the time between the i^{th} and $(i+1)^{th}$ recording. Suppose a woman of a given marital duration is observed between time points t_I and t_2 ($t_I < t_2$) and the time of occurrences of successive live births are recorded. The interval between i^{th} and $(i+1)^{th}$ recording ($i \ge 1$), occurring in a time period (t_1, t_2), say X_i (t_1, t_2), is defined here as interior birth interval of order i.

Perrin and Sheps (1964) have considered human reproduction as a Markovian renewal process and shown that it attains equilibrium about four to five years after marriage. Thus, if t_i is the distant point since marriage, and if we assume that the parameters of reproduction are constant in the recent past prior to t_i , the process may be assumed to be an equilibrium process at time t_1 . Further, let us assume that the parameters do not change after t_1 . Now, under these assumptions: (i) a constant risk of conception, m, (ii) one-to-one correspondence between a conception and a live birth, and (iii) a constant period of nonsusceptibility denoted by h, associated with each conception, the probability density function (p.d.f.) of the random variable $X_i(t_1,t_2)$, say $g_i(t/h)$, becomes (see Appendix part I):

$$g_{i}(t/h) = \frac{m e^{-m(t-h)}}{1 + mh} [mh + \sum_{s=0}^{i-1} A \{s+1, m (T-t-ih)\}]$$

$$- \sum_{s=0}^{i-2} A \{s+1, m(T-t-i-1 h)\} J/F^{*}_{i+1}(T/h) h < t < T-ih$$

$$= \frac{m e^{-m(t-h)}}{1 + mh} [m(T-t-i-1 h) - \sum_{s=0}^{i-2} S=0]$$

$$A \{s+1, m(T-t-i-1 h)\} J/F^{*}_{i+1}(T/h) T-ih < t < T-i-1 h$$
(2)

where the term $\sum_{s=0}^{i-2} A(s+1, m(T-t-\overline{i-1} h))$ in expressions 1 and 2 does not appear when i=1,

$$A(n,c) = 1-e^{-c} \sum_{j=0}^{n-1} \frac{c^j}{j!}$$

B. N. Bhattacharya and K. K. Singh

and $F^*_{i+1}(t/h)$ is the distribution function of the time of $(i+1)^{th}$ recording and is given by

$$F^*_{i+1}(t/h) = \{1/(1+mh)\} [m(t-ih - \sum_{s=0}^{i-1} A \{s+1, m(t-ih)\}],$$

$$ih < t < \overline{i+1} h$$

$$= \{1/(1+mh)\} [mh + \sum_{s=0}^{i} A \{s+1, (t-\overline{i+1} h) m\}$$

$$- \sum_{s=0}^{i-1} A \{s+1, m(t-ih)\}] t > \overline{i+1} h$$
(3)

If the women consist of q distinct groups, with respect to the period of nonsusceptibility, h, and p_l are their proportions in the lth group with

$$h = h_l \ (l = 1, 2, ..., q), \ o < p_l < 1 \text{ and } \sum_{l=1}^{q} p_l = 1, \text{ the p.d.f. of } X_i(t_1, t_2) \text{ is }$$

$$g_{i}(t) = \frac{\sum_{l=1}^{q} p_{l} F^{*}_{i+1} (T/h_{l}) g_{i}(t/h_{l})}{\sum_{l=1}^{q} p_{l} F^{*}_{i+1} (T/h_{l})}$$

Estimation

Distribution 4 involves 2q parameters namely m; $p_1, p_2, \ldots, p_{q-1}$; $h_1, h_2, \ldots, h_{q-1}$ and h_q . Assuming that h takes only two values, h_1 and h_2 , with respective probabilities p_1 and p_2 , $0 < h_1 < h_2$, $0 < p_1 < 1$, $p_1 + p_2 = 1$, the p.d.f. of $X_i(t_1,t_2)$ involves four parameters: m, p_1 , h_1 and h_2 . Here, for q=2, we have outlined a procedure to obtain maximum likelihood (m.l.) estimates of the parameters m and p_1 and their variance — covariance matrix for known values of h_1 and h_2 . The method can be extended on similar lines when h takes more than two values.

Let $P_{i,j}$, $j=1,2,\ldots,k$, be the probability that $X_i(t_1,t_2)$ belongs to j^{th} interval. Further, in a random sample of size n, let n_j be the number of women

falling in j^{th} interval, $\sum_{j=1}^{k} n_j = n$. The number of women classified in this

manner follows a multinomial distribution given by

$$L(n_1, n_2, ..., n_k) = n! \prod_{j=1}^{k} (P_{i,j}^{n_j}/n_j!)$$

Since $P_{i,j}$'s are functions of m, p_1 , h_1 and h_2 , the m.l. estimates of \hat{m} and \hat{p}_1 of m and p_1 for known h_1 and h_2 are solutions of two simultaneous equations:

$$\frac{\partial \log L}{\partial m} = \sum_{j=1}^{k} \frac{n_j}{P_{i,j}} \frac{\partial P_{i,j}}{\partial m} = 0$$
 (5)

$$\frac{\partial \log L}{\partial p_I} = \sum_{j=1}^k \frac{n_j}{P_{i,j}} \frac{\partial P_{i,j}}{\partial p_I} = 0 \tag{6}$$

Since the explicit solution of the equations 5 and 6 is not possible, the scoring method (Rao, 1952) can be used to obtain m.l. estimates of the parameters. The variance-covariance matrix of the estimators is given by the inverse of the information matrix.

The pilot values of the parameters, which are required for the scoring method, can be obtained by equating the first two sample moments of the observed distribution to their respective theoretical expressions (see Appendix part II). The equations can be solved by the Newton-Raphson iteration procedure.

For $p_I = 1$ the m.l. estimate of m, for known h, can be obtained by solving equation 5. In this case, $P_{i,j}$'s are obtained from the expressions A6 to A8 in the appendix by substituting $p_I = 1$ and $h_I = h$. The variance of the m.l. estimator of m is given by the expression

$$1/n \sum_{i=1}^{k} \left(\frac{\partial P_{i,j}}{\partial m} \right)^{2}.$$

Application

For the purpose of illustration, we have used the data on interior intervals from "Rural Development and Population Growth - A Sample Survey 1978" conducted under the auspices of the Centre of Population Studies, Banaras Hindu University, Varanasi, India. In accordance with the objectives of the survey, a stratified random sample of 19 villages was selected from Varanasi Tehsil and adjoining areas. This investigation consisted of a baseline survey, with the reference date March 25, 1978 and six rounds of prospective surveys, conducted during the period April, 1978 to March, 1981. The baseline survey included all 3,514 households from the sample villages. Besides other information, the baseline survey included the detailed fertility record of each married woman below 50 years of age. This record was updated in each round of prospective survey. Bhattacharya and Singh (1983) carried out the life table analysis of the birth interval data from the above survey and found that the average length of various orders of interlive birth intervals were between 36 and 45 months. Thus, the number of women with at least one interior interval lying in the three years starting from the reference date of the baseline survey and the final round of the prospective survey was small. In order to have an adequate number of women with such intervals, we have taken the intervals lying between April, 1975 and March, 1981. The married women enumerated and having an uninterrupted married life until the reference date of the final round of the prospective survey were considered. Further, the women having a marital duration shorter than five years in April, 1975, were excluded because of the assumption in the model that the start of the observational period is the distant point since marriage. In this case, T=6 years.

In the present survey, 1,022 and 293 women had experienced two or more

and three or more births respectively between April, 1975 and March, 1981. For homogeneity in fertility characteristics, these women were classified into four groups, on the basis of their marital duration as of 1975. The groups I, II, III and IV relate to women with marital durations of 5-10, 10-15, 15-20 and 20 and more years, respectively. The women with marital durations of 20 or more years are not taken into account due to the paucity of an adequate number of observations. Furthermore, to have a sufficient number of observations in each of the first three groups, the data on distribution of the length of first order interior birth intervals of women with two or more births have been considered and are given in Table 1.

The nonsusceptible period h' is the sum of the gestation period and the period of postpartum amenorrhoea (PPA). The gestation period associated with a birth is almost constant at about nine months, whereas PPA may vary among women. In the "Demographic Survey of Varanasi (Rural) 1969," conducted by the Centre of Population Studies, Banaras Hindu University, the average duration of PPA associated with a live birth was found to be nine months (0.75 years). Further, the nature of distribution of PPA was bimodal, the first and the second peaks occurring at 0-4 and 10-14 months respectively (Singh and Bhaduri, 1971), thus indicating that the population consisted perhaps of two groups in respect to PPA - one with a low average duration of PPA and the other with high average. Therefore, keeping in view that the present survey is conducted in that same locality, in the application of the distribution $X_1(t_1,t_2)$, which is based on the assumption that the duration of nonsusceptibility associated with live birth is constant, h may be taken to be 1.5 years. Similarly, for the distribution $X_1(t_1,t_2)$, where the population is heterogeneous with respect to PPA, it can be reasonably assumed that h takes only two values $h_1 = 1.00$ years and $h_2 = 1.75$ years, with respective probabilities p_1 and $1-p_1$. The observed and expected frequencies for each set of data are given in Table 1. The estimates of parameters and their estimated variances and covariances are also given in Table 1.

The distribution which assumes h_I to be the same for all women fails to explain the observed frequency in the first cell.

The computed values of chi-square under the model with variable h, are insignificant for all three sets of data. It should be noted that the computed values of chi-square for constant h are considerably larger than those for variable h. This is also in agreement with the observation of PPA that the female population consists of two groups: one with a low average value of PPA of about 0.25 years and the other with a high average of about 1.00 years. The proportion of women with a low value of PPA constitutes about 30 per cent of the total female population. This finding is in agreement with the results

B. N. Bhattacharya and K. K. Singh

Expected frequencies 0.571 INTERVAL FOR WOMEN WITH EFFECTIVE MARRIAGE DURATIONS 5-10, 10-15 AND 15-20 YEARS Constant 6.34 3.23 42.9 22.8 16.9 172.0 TABLE 1. OBSERVED AND EXPECTED DISTRIBUTIONS OF THE FIRST ORDER INTERIOR BIRTH 30.2 11.9 10.9 36.4 Variable 0.618 0.298 1.34 0.13 -0.23 15.0 172.0 17.1 31.3 41.3 29.8 21.2 6.9 9.4 of Group III frequencies Observed 172 16 28 47 30 21 15 9 Expected frequencies Constant 0.586 6.31 2.13 88.2 329.0 68.9 43.3 28.9 20.8 18.9 60.1 Variable 0.635 0.292 2.45 4.32 62.8 79.3 56.7 41.0 26.9 17.0 13.2 329.0 32.1 frequencies of Group II Observed 31 57 79 58 46 30 18 임 329 Expected frequencies Constant h 099.0 1.53 9.81 57.5 35.9 18.5 491.0 114.1 84.0 Variable 0.715 0.283 5.52 0.52 ٠. ١. 54.6 32.9 19.5 491.0 104.3 128.5 84.2 14.3 frequencies of Group I Observed 132 87 67 26 17 15 491 cov. 命介) x10-3 Var. (\$1,) x10-3 Var. (n) x10-3 Less than 1.5 - 2.02.0 - 2.53.0 - 3.54.0 - 4.5equal to 1.5 2.5 - 3.03.5 - 4.0 than 4.5 Time in greater years Total

obtained in an analysis of the pattern of PPA (Singh and Bhaduri, 1971).

The estimates of m, the live birth conception rate, obtained using distribution, are 0.715, 0.635 and 0.618, during 5-10, 10-15, 15-20 years of marriage duration respectively. In the present survey, the average age of women at return marriage was found to be 15 years. Thus, the estimated values of m for groups I, II and III relate to age intervals 20-25, 25-30 and 30-35 respectively. Accordingly, the estimated asymptotic age-specific marital fertility rates,

which are approximately
$$\sum_{i=1}^{2} p_i \frac{m}{1+m h_i}$$
 of the above three groups, are 0.344, 0.326 and 0.322 respectively

0.344, 0.326 and 0.322 respectively.

It is worth noting that the computed length of interval, based on the data of the distant past suffers from memory bias (Osman, 1984). Moreover, there is a chance of underreporting the number of children ever born. The present approach accounts for intervals of the recent past that are closer to the date of the survey, and hence may be assumed to be free from the aforementioned bias.

Acknowledgments

The authors are thankful to Dr. S.N. Singh, Director, Centre of Population Studies, Banaras Hindu University, for providing the data. The authors also wish to thank the editor and the referees for their very helpful suggestions.

References

- Bhattacharya, B.N. and K.K. Singh. 1983. On a modification of life table technique for analysis of birth interval data and its application. Janasamkhya 1(2):99-116.
- Chakrabarty, K.C. 1976. Some Probability Distributions for Birth Intervals. Unpublished Ph.D. thesis. Department of Mathematics, Banaras Hindu University, Varanasi, India.
- D'Souza, S. 1973. Interlive birth intervals of non-contraceptive population A data analytic study. Social Action 23:404-425.
- . 1974. Closed Birth Intervals: A Data Analytic Study. New Delhi, India: Sterling Publishers

B. N. Bhattacharya and K. K. Singh

- George, A. 1967. A probability model for interlive birth interval. Paper presented at the 36th session of the International Statistical Institute, Sydney, Australia.
- and A.M. Mathai. 1975. A generalized distribution for interlive birth interval. Sankhya 37B:332-342.
- Menken, J.A. and M.C. Sheps. 1972. Sampling frame as a determinant of observed distributions of duration variables. In T.N.E. Greville (ed.), Population Dynamics. New York: Academic Press.
- Osman, N.E-H. 1984. The validity of fertility trends from retrospective data. Jordan Population Bulletin of ECWA No. 24.
- Perrin, E.B. and M.C. Sheps. 1964. Human reproduction: A stochastic process. Biometrics 20:28-45.
- Poole, K. 1971. Fertility based on birth interval data. Research Triangle Institute Project No. SU-587. Research Triangle Park, North Carolina.
- Rao, C.R. 1952. Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons.
- Sheps, M.C., J.A. Menken, J.C. Ridley and J.W. Lingner. 1970. Truncation effect of closed and open birth interval data. Journal of American Statistical Association 65:678-693.
- Singh, S.N., B.N. Bhattacharya and R.C. Yadava. 1979. An adjustment of a selection bias in postpartum amenorrhea from follow-up studies. Journal of the American Statistical Association 74:916-920.
- and T. Bhaduri. 1971. On the pattern of post-partum amenorrhoea. In S.N. Singh (ed.), Proceedings of All India Seminar on Demography and Statistics, Banaras Hindu University, Varanasi, India.
- ______, R.C. Yadava and A. Pandey. 1978. A probability model for forward birth interval. Health and Population Perspective and Issues 1(4):295-304.
- Som, R.K. 1973. Recall Lapse in Demographic Enquiries. Bombay, India: Asia Publishing House.
- Srinivasan, K. 1966. An application of a probability model to the study of interlive birth intervals. Sankhya 28B:175-182.

Received May, 1983; revised January, 1986.

Appendix

I. From the assumptions as given in the model section, it can be shown that $T_i(i \ge 1)$ has a common probability density function f(t/h):

$$f(t/h) = me^{-m(t-h)}, t > h$$
 (A1)

and the probability density function of T_0 , $f_0(t/h)$:

$$f_{o}(t/h) = \frac{1-F(t/h)}{\mu} = \begin{cases} \frac{m}{1+mh} & 0 < t \le h \\ \\ \frac{m}{1+mh} & t > h \end{cases}$$
(A2)

where μ is instantaneous birth rate, which is the mean of T_i ($i \ge 1$) and F(t/h) is the distribution function corresponding to the density function f(t/h).

The distribution function of random variable $X_i(t_1,t_2)$, say $H_i(t/h)$, is

$$H_{i}(t/h) = \frac{P[T_{i} \leq t \cap \{ (T_{o} + T_{I} + \dots + T_{i}) \leq T \}]}{P[(T_{o} + T_{I} + \dots + T_{i}) \leq T]}$$
(A3)

and the corresponding density function, say $g_i(t/h)$, is

$$g_{i}(t/h) = \frac{P[(T_{o} + T_{I} + \dots + T_{i-I}) \leq T - t] f(t/h)}{P[(T_{o} + T_{I} + \dots + T_{i}) \leq T]}$$
(A4)

Since T_0 , T_1 , ..., T_i are mutually independent, the p.d.f. of $T_0 + T_1 + \dots + T_i$, say $f^*_{i+1}(t/h)$, is

$$f^*_{i+1}(t/h) = \{m/(1+mh)\} A \{i, m(t-ih)\}, ih < t < \overline{i+1} h$$

$$= \{m/(1+mh)\} [A \{i, m(t-ih)\} - A \{i+1, m(t-\overline{i+1} h)\}],$$

$$\overline{i+1} h < t < \infty$$
(A5)

and the distribution function, corresponding to $f_{i+1}^*(t/h)$, is given by expression 3.

II. Moments — when h' is the same for all women in the population, the r^{th} order moment $(r \ge 1)$ of the random variable $X_i(t_1, t_2)$ about zero is given by

 $E(X_i^r(t_1,t_2)/h) = \frac{1}{F^*_{:\perp,1}(T/h)} - \frac{r!}{m^r} \left[\sum_{u=0}^r \frac{(mh)^u}{u!} \right]$

$$-\frac{e^{-m(T-i+1}h)}{i+mh} \left\{ \sum_{u=o}^{r} (r+i-\overline{u-1}) - \frac{\{m(T-ih)\}^{u}}{u!} \right\}$$

$$+\frac{e^{-m(T-ih)}}{1+mh} \left\{ \sum_{u=o}^{r} (r+i-u) - \frac{\{m(T-i-1}h)\}^{u}}{u!} \right\}$$

$$+\frac{e^{-m(T-ih)}}{i+mh} \sum_{s=o}^{i-2} \sum_{k=o}^{s} \sum_{u=o}^{r} \frac{(mh)^{r-u} (m \overline{T-ih})^{k+u+1}}{(k+u+1)! (r-u)!}$$

$$-\frac{e^{-m(T-i+1}h)}{1+mh} \sum_{s=o}^{i-1} \sum_{k=o}^{s} \sum_{u=o}^{r} \frac{(mh)^{r-u} \{m(T-i+1}h)\}^{k+u+1}}{(k+u+1)! (r-u)!}$$

The expression for $\mu'_{i,r}$, the r^{th} $(r \ge 1)$ moment about zero of $X_i(t_1, t_2)$ when h takes q values h_1, h_2, \ldots, h_q is given by

$$\mu_{i,r}' = E(X_i^r(t_1, t_2)) = \frac{\sum\limits_{l=1}^q p_l \ F^*_{i+1} \ (T/h_l) \ E(X_i^r(t_1, t_2) \ /h_l)}{\sum\limits_{l=1}^q p_l \ F^*_{i+1} \ (T/h_l)}$$

III. From the assumption, as given in the Estimation section, $P_{i,j}$, $j=1,2,\ldots,k$, the probability that $X_i(t_1,t_2)$ belongs to j^{th} interval, is given by

$$P_{i,j} = \frac{\sum_{l=1}^{2} p_l F^*_{i+1}(T/h_l) G_i(a/h_l)}{\sum_{l=1}^{2} p_l F^*_{i+1}(T/h_l)} \qquad \text{for } j = 1$$
 (A6)

$$P_{i,j} = \int_{a_{j-1}}^{a_j} g_i^*(x) dx = \frac{\sum_{l=1}^{2} p_l F^*_{i+1}(T/h_l) [G_i(a_j/h_l) - G_i(a_{j-1}/h_l)]}{\sum_{l=1}^{2} p_l F^*_{i+1} (T/h_l)}$$

$$j = 2,3,...,k-1$$
 (A7)

$$P_{i,k} = I - P_{i,l} - P_{i,2} - \dots - P_{i,k-I}$$
 (A8)

where,

B. N. Bhattacharya and K. K. Singh

$$G_{i}(t/h) = \int_{h}^{t} g_{i}(x/h)dx = \frac{1}{F^{*}_{i+1}(T/h)} \frac{1}{1+mh} \left[mh \left\{ i-e^{-m(t-h)} \right\} + e^{-m(t-h)} \left\{ \sum_{s=0}^{i-1} A \left\{ s+1, m(T-t-i-1) h \right\} \right\} \right]$$

$$- \sum_{s=0}^{i} A \left\{ s+1, m(T-t-i-h) \right\}$$

$$- \left\{ \sum_{s=0}^{i-1} A \left\{ s+1, m(T-i-h) \right\} \right\} - \sum_{s=0}^{i-1} A \left\{ s+1, m(T-i-h) \right\} \right] h < t < T-ih$$

$$= \frac{1}{F^{*}_{i+1}(T/h)} \frac{1}{1+mh} \left[mh + \sum_{s=0}^{i-1} A \left\{ s+1, m(T-i-h) \right\} \right]$$

$$- \sum_{s=0}^{i-1} A \left\{ s+1, m(T-i-h) \right\} + e^{-m(t-h)} \sum_{s=0}^{i-1} A \left\{ s+1, m(T-t-i-h) \right\}$$

$$- e^{-m(t-h)} m(T-t-i-1-h) \right\}, T-ih < t < T-i-1-h$$
(A9)