S. N. Singh and R. C. Yadava

A Generalized Probability Model for an Equilibrium Birth Process

1. Introduction

THE probability models proposed by Dandekar (1955), Brass (1958), and Sheps (1964), Sheps and Perrin (1966), Sheps *et al.* (1969), Pathak (1966), Singh (1963, 1968), Singh and Bhattacharya (1970, 1971), Singh *et al* (1974, 1975), Das Gupta (1973) and others, are useful to explain the variation in the number of conceptions and births to a couple during a fixed period of time. A detailed account of such models can be obtained from Sheps *et al* (1969), Singh (1966), Das Gupta (1973) and others, where the authors have extensively reviewed the literature and added their contributions.

Singh and Bhattacharya (1970) derived a generalized probability distribution for the number of conceptions in a period (0, T) of length T allowing for two types of pregnancy outcomes. One of the assumptions of the model is that the female is exposed to the risk of conception at the beginning of the observational period which is not true, specially when the starting point is not marriage.

The purpose of the present paper is to derive a probability model for the Bumber of conceptions during an interval of length T assuming the start of the observational period at a considerable distance from marriage. The other assumptions are similar to those of Singh and Bhattacharya (1970). The present model may be considered as an extension of the distribution of Singh and Bhattacharya (1970) and also of the modified Poisson distribution for 'abrupt sequence' derived by Dandekar (1955). The model is derived in Section 2 where expressions for the mean and variance of the distribution are also obtained. It is applied to some observed distributions in Section 3.

2. Model

Let $X(T_0, T)$ denote the number of conceptions to a female during the period $(T_0, T_0 + T)$ where T_0 is measured from marriage, that is, the starting point is at a distance T_0 from marriage. We derive the distribution of $X(T_0, T)$ on the following assumptions :

1. The female has led a married life from marriage till the end of the observational period.

2. A conception may result either in a live birth or in a foetal death. If the conception results in a live birth, it is called complete, otherwise it is designated as incomplete.

3. The time of first conception since marriage follows an exponential distribution first with probability density function me^{-mt}, $t > 0, m > 0$ and the time between the r^{th} and $(r + 1)^{th}$ conceptions, $(r = 1, 2, \ldots)$ follows displaced exponential distributions $me^{-m(t-h_1)}$, $t > h_1$ or $me^{-m(t-h_2)}$, $t > h_2$, $(h_1 < h_2)$ according as the r^{th} conception is incomplete or complete.

   This assumption implies that if there is a conception at a point of time, then there is no other conception either in h_1 or h_2 units of time according as the conception is incomplete or complete.

4. Let θ be the probability that a conception is complete so that $(1-\theta)$ is the probability that it is incomplete.

5. T_0 is a distant point since marriage.

Under the above assumptions, the distribution of $X(T_0, T)$ becomes independent of T_0 (see Cox and Miller, 1965, pp. 340-41), henceforth we write $X(T)$ for $X(T_0, T)$. In the present case, the total number of conceptions cannot be more than n, where $n = T/h_1$ or $[T + h_1/h_1]$ according as T is a multiple of h_1 or not; $[T + h_1/h_1]$ stands for the greatest integer not exceeding $T + h_1/h_1$. The probability function P_x of $X(T)$ is given as

$$P_0 = 1 - [H_0(T) - H_1(T)]$$

$$P_r = [H_{r-1}(T) - 2H_r(T) + H_{r+1}(T)], \; r = 1, 2, \ldots, h \tag{1}$$

where

$$H_r T = \frac{m}{1 + m\{(1-\theta)h_1 + \theta h_2\}} \Bigg[\sum_{j=0}^{l_r} \binom{r}{j} \theta^j (1-\theta)^{r-j}$$

$$\int_{jh_2 + \overline{r-j}h_1}^{T} \left\{ 1 - e^{-m(t - jh_2 - \overline{r-j}h_1)} \sum_{s=0}^{r-1} \frac{m^s (t - jh_2 - \overline{r-j}\,h_1)^s}{s!} \right\} dt \Bigg] \tag{2}$$

$$l_r = \min [r, \{(T - rh_1)/(h_2 - h_1)\}].$$

Obviously,

$$H_0(T) = \frac{mT}{1 + m\{\theta h_2 + (1-\theta)h_1\}} = \frac{mT}{1 + m\bar{h}},$$

where

$$\bar{h} = \{\theta h_2 + (1-\theta)h_1\}.$$

Proof. Let the conceptions be counted from the point T_0 and let the time of r^{th} conception, for $r = 1, 2, \ldots,$ be $T_1 + T_2 + \ldots + T_r$, where T_1 is the time of first conception; from T_0 and $T_r (r > 1)$ is the time between $(r-1)^{th}$ and r^{th} conceptions. If T_0 is at a considerable distance since marriage, the above birth process becomes an equilibrium renewal process where the intervals T_2, $T_3, \ldots,$ are independent and identically distributed but T_1 has a different distribution.

The probability density function of $T_r (r > 1)$ is given as

$$\begin{aligned} f(t) &= 0 && \text{for} \quad 0 < t < h_1 \\ &= (1-\theta)me^{-m(t-h_1)} && \text{for} \quad h_1 < t < h_2 \\ &= (1-\theta)me^{-m(t-h_1)} + \theta m e^{-m(t-h_2)} && \text{for} \quad t > h_2. \end{aligned} \tag{3}$$

and that of T_1 is

$$f_1(t) = \frac{1 - F(t)}{\mu}, \tag{4}$$

where $F(t)$ and $\mu = \dfrac{1 + m\bar{h}}{m}$ are the distribution function and mean, respectively, associated with $f(t)$.

The Laplace transform of $f_1(t)$ is $\dfrac{1 - \phi(s)}{\mu s}$ where $\phi(s)$ is the Laplace transform corresponding to $f(t)$.

Since $T_1, T_2, \ldots$ are mutually independent, the Laplace transform of $T_1 + T_2 + \ldots + T_r$ is

$$\frac{(1 - \phi(s))(\phi(s))^{r-1}}{\mu s},$$

$$\frac{(\phi(s))^{r-1}}{\mu s} - \frac{(\phi(s))^r}{\mu s}. \tag{5}$$

The inverse of (5), the probability density function of $T_1 + T_2 + \ldots + T_r$, $f_r^*(t)$, say, is

$$f_r^*(t) = \frac{1}{\mu} \int_0^t f^{(r-1)}(x)\, dx - \frac{1}{\mu} \int_0^t f^{(r)}(x)\, dx$$

$$= \frac{1}{\mu} F^{(r-1)}(t) - \frac{1}{\mu} F^{(r)}(t), \tag{6}$$

where $f^{(r)}(t)$ denotes the r fold convolution of $f(t)$ with itself and $F^{(r)}(t)$, the corresponding distribution function.

The expression for $F^{(r)}(t)$ is easily obtained as

$$F^{(r)}(t) = \sum_{j=0}^{lr} \binom{r}{j} \theta^j (1 - \theta)^{r-j} \Big[1 - e^{-m(t - jh_2 - \overline{r-j}h_1)}$$

$$\sum_{s=0}^{r-1} \frac{m^s(t - jh_2 - \overline{r - j}h_1)^s}{s!} \Big]. \tag{7}$$

Now

$$P_0 = 1 - \int_0^T \frac{1 - F(t)}{\mu}\, dt$$

$$= 1 - \frac{1}{\mu} \int_0^T dt + \int_0^T F(t)\, dt$$

$$= 1 - H_0(T) + H_1(T), \tag{8}$$

and

$$P_r = \int_0^T f_r^*(t)\, dt - \int_0^T f_{r+1}^*(t) dt$$

$$= \frac{1}{\mu} \int_0^T F^{(r-1)}(t)\, dt - 2\, \frac{1}{\mu} \int_0^T F^{(r)}(t)\, dt + \frac{1}{\mu} \int_0^T F^{(r+1)}(t)\, dt$$

$$= H_{r-1}(T) - 2H_r(T) + H_{r+1}(T) \quad \text{for} \quad r = 1, 2, \ldots, n. \tag{9}$$

It is easily seen that

$$H_r(T) = \frac{mT}{1 + mh} - r + \sum_{i=0}^{r} \binom{r}{j} \theta^j (1 - \theta)^{r-j}\, \psi_{(r-1),j}\, (T - jh_2 - \overline{r - j}\, h_1),$$

$$\text{if} \qquad lr = r, \tag{10}$$

where

$$\psi_{(r-1),j}\, (T - jh_2 - \overline{r - j}\, h_1) = \frac{1}{1 + mh} \sum_{s=0}^{r-1} \sum_{k=0}^{s} e^{-m(T - jh_2 - \overline{r-j} h_1)}$$

$$\frac{m^k (T - jh_2 - \overline{r - j}\, h_1)^k}{k!}, \tag{11}$$

and

$$H_r(T) = \sum_{j=0}^{lr} \binom{r}{j} \theta^j (1-\theta)^{r-j} \Bigg[\frac{m(T - jh_2 - \overline{r-j}h_1)}{1+mh} - \frac{r}{1+mh}$$

$$+ \psi_{(r-1),j} (T - jh_2 - \overline{r-j}h_1) \Bigg],$$

if $$lr < r. \qquad (12)$$

For $\theta = 1$, that is, $h_1 = h_2 = h$, we have

$$H_r(T) = \frac{mT}{1+mh} - r + \phi_1(r-1), \qquad (13)$$

where, $$\phi_1(r-1) = \frac{1}{1+mh} \sum_{s=0}^{r-1} \sum_{k=0}^{s} e^{-m(T-rh)} \frac{m^k(T-rh)^k}{k!}. \qquad (14)$$

Thus

$$P_0 = \phi_1(0)$$

$$P_r = \phi_1(r) - 2\phi_1(r-1) + \phi_1(r-2) \quad \text{for } r = 1, 2, \ldots, n-2$$

$$P_{n-1} = n - \frac{mT}{1+mh} - 2\phi_1(n-2) + \phi_1(n-3)$$

$$P_n = -(n-1) + \frac{mT}{1+mh} + \phi_1(n-2). \qquad (15)$$

It should be emphasized here that the above expression is the same as that obtained by Dandekar (1955), but with a slight modification. Dandekar derived as

$$P_r = \phi_1(r) - 2\phi_1(r-1) + \phi_1(r-2).$$

This expression is true only for $r = 0, 1, 2, \ldots, n-2$ but does not work for $r = n-1$ and n (see equations 15).

The Mean and Variance

Since $X(t)$ takes values $0, 1, 2, \ldots, n$ with corresponding probabilities

$P_0, P_1, \ldots, P_n,$

$$E[X(T)] = \sum_{r=0}^{n} r\, P_r$$

$$= H_0(T) = \frac{mT}{1 + m\bar{h}}, \qquad (16)$$

which agrees with the general result for the equilibrium renewal process viz., mean $= T/\mu$ where μ is the mean time between renewals.

The variance of the distribution is

$$V[X(T)] = \sum_{r=0}^{n} r^2 P_r - \left(\frac{mT}{1 + m\bar{h}}\right)^2.$$

It is difficult to obtain a simple expression for $V[X(T)]$; for general T, however, an approximate expression for $V[X(T)]$ is easily obtained for large T as

$$V[X(T)] = \frac{\sigma^2 T}{\mu^3} + \left(\frac{1}{6} + \frac{\sigma^4}{2\mu^4} - \frac{\mu_3}{3\mu^3}\right) + 0(1), \qquad (17)$$

where in the present case

$$\mu = \frac{1}{m} + \bar{h}$$

$$\sigma^2 = \mu_2' - \mu^2$$

$$\mu_3 = \mu_3' - 3\mu_2' \mu + 2\mu^3,$$

where

$$\mu_2' = \theta\left[\frac{2}{m^2} + \frac{2h_2}{m} + h_2^2\right] + (1 - \theta)\left[\frac{2}{m^2} + \frac{2h_1}{m} + h_1^2\right]$$

$$\mu_3' = \theta\left[\frac{6}{m^3} + \frac{6h_2}{m^2} + \frac{3h_2^2}{m} + h_2^3\right]$$

$$+ (1 - \theta) \left[\frac{6}{m^3} + \frac{6h_1}{m^2} + \frac{3h_1^2}{m} + h_1^3 \right]. \quad (18)$$

3. Application

The model takes account of two types of pregnancy outcomes: foetal wastage and live birth. We do not possess data on number of conceptions leading to pregnancy terminations of the above two types in a fixed period of time where the start of the observational period is at a considerable distance from marriage. However, the model can be applied to data where one to one correspondence between a conception and a live birth is assumed; and the children dying within a year and those surviving more than a year are counted separately. It is well known that the length of non-susceptible period associated with an infant death is considerably smaller than that related to the child surviving at least one year. Two observed distributions of this type have been taken from 'A Demographic Survey of Varanasi (Rural)' for the application of the model.

The Demographic Survey of Varanasi (Rural) was conducted in the year 1969-70 under the auspices of Demographic Research Centre, Banaras Hindu University. **In** the survey, about 2200 households scattered in 52 villages of Varanasi Tehsil—one of the four tehsils of Varanasi District comprising nearly 1200 villages and 0.13 million households were selected following a two stage stratified random sampling procedure. The information collected in the survey includes household structure, household facilities., marriage, fertility, mortality, migration and family planning.*

In the survey, complete birth record of each eligible couple was noted. A couple is defined eligible if both the partners were alive and the female's age was less than 50 years on the reference date of the survey that is, October 1969.

The table in the appendix presents the observed distribution of number of births in the last 5 and 7 years to eligible females of age group 25-29 having at least 5 and 7 years of marriage duration. The females without any birth upto the reference date are excluded presuming them to be sterile. In the surveyed area, the average age at marriage for females is nearly 15 years and the observed distribution relates to females aged 25-29 years, hence the start of the observa-

***For detailed account of** the survey, see Singh *et al.,* 1970.

tional period may be considered at a considerable distance from marriage. For fitting the present model to the above observed distributions, we notice that $(1-\theta)$ represents the probability that a birth ends in an infant death while θ is the probability that the child survives at least one year, and h_1 and h_2 are the respective non-susceptible periods (gestation plus post-partum amenorrhea : PPA). From the data of the survey, it is observed that the average PPA associated with infant death is 2 to 3 months and it is, when the child lives for more than a year about 10-11 months. The analysis of the pattern of PPA associated with children living more than a year, by Singh and Bhaduri (1971) revealed that the nature of the curve is bimodal, the first and second peaks occurring at 2-3 months and 12-13 months respectively. Thus the population of eligible females may be considered comprising two groups : (i) those with PPA nearly 3 months irrespective of the outcome of the pregnancy; (ii) others whose PPA is 3 or 12 months according to whether or not the infant survives the first year. Let Π and $1-\Pi$ be the respective proportions for the two groups. The infant mortality rate in the area during last 5 years is about 180 per thousand. Keeping these findings in view, we take $h_1 = 1$ year, $h_2 = 1.75$ years $\theta = 0.2$ $\Pi = 0.4$.

The parameter m is estimated by equating the observed mean with the theoretical mean which is given by

$$\Pi \frac{mT}{1+mh_1} + (1-\Pi)\frac{mT}{1+m\bar{h}}.$$

The expected frequencies are shown in columns (3) and (5) of the table. The values of χ^2 in both the cases are insignificant at 5 per cent level. The model describes both the observed distributions though it is derived under some simplifying assumptions relating to human fertility. The important characteristic of the present model is that it takes into account the situation that a large number of females are not exposed to the risk of conception at the start of the observational period, and hence is more suitable for analysing data on number of births or conceptions relating to age segments of the reproductive period, or the preceding T years from the reference date of any retrospective enquiry.

References

1. Brass, W., 1958, The distribution of births in human populations, *Population Studies*, **12**, 51-72.
2. Cox, D. R. and Miller, H. D., 1965, *The Theory of Stochastic Processes*, London.

3. Dandekar, V. M., 1955, Certain modified forms of binomial and Poisson distribution, *Sankhya*, **15**, 237-250.
4. Das Gupta, P., 1973, A stochastic model of human reproduction : some preliminary results, *Theoretical Population Biology*, **4**, 466-490.
5. Pathak, K. B., 1966, A probability distribution for the number of conceptions, *Sankhya*, **28B**, 213-218.
6. Perrin, E. B. and Sheps, M. C., 1964, Human reproduction—A stochastic process, *Biometrics*, **20**, 28-45.
7. Sheps, M. C. and Perrin, E. B., 1966, Further results from a human fertility model with a variety of pregnancy outcomes, *Human Biology*, **38**, 180-193.
8. ———, Menken, J. A. and Radick, A. P., 1969, Probability models for family building—An analytical review, *Demography*, **6**, 161-183.
9. Singh, S. N., 1963, Probability models for the variation in the number of births per couple, *J. Amer. Stat. Assoc.*, **58**, 721-727.
10. ———, 1966, Some probability distributions utilized in human fertility, *Seminar Volume in Statistics*, (Banaras Hindu University), 74-84.
11. ———, 1968, A chance mechanism of variation in the number of births per couple, *J. Amer. Stat. Assoc.*, **63**, 209-213.
12. ———and Bhattacharya, B. N., 1970, A generalized probability distribution for couple fertility, *Biometrics*, **26**, 30-40.
13. ———, Yadava, R. C. and Bhaduri, T., 1970, A demographic survey of Varanasi (Rural) 1969-70, *Prajna*, **16**, 176-189.
14. ———and Bhattacharya, B. N., 1971, On some probability distributions for couple fertility, *Sankhya*, **33B**, 315-322.
15. ———and Bhaduri, T., 1971, On the pattern of postpartum amenorrhea, *Proceedings of All India Seminar on Demography and Statistics* (Banaras Hindu University), 284-298.
16. ———, Bhattacharya, B. N. and Yadava, R. C., 1974, A parity dependent model for number of births and its applications, *Sankhya*, **36B**, 93-102.
17. ———, ———and ———, 1975, A fertility model and its applications, *Demography India*, **4**, 443-448.

Appendix

TABLE—DISTRIBUTIONS OF ELIGIBLE FEMALES AGED 25-29 YEARS ACCORDING TO THE NUMBER OF BIRTHS DURING THE LAST 5 AND 7 YEARS

Number of births	*Last 5 years*		*Last 7 years*	
	Observed number of women	*Expected number of women*	*Observed number of women*	*Expected number of women*
(*1*)	(*2*)	(*3*)	(*4*)	(*5*)
0	20	25.8	3	6.9
1	155	150.4	58	61.7
2	200	189.2	166	162.7
3	45	54.9	151	135.8
4 and above	5	4.7	34	44.9
Total	425	425.0	412	412.0
$\hat{m}$	0.605		0.620	
χ^2	3.865		6.840	
d.f.	3		3	