J. Biosci., Vol. 3 Number 2, June 1981, pp. 207-214. © Printed in India.

A conception dependent probability distribution of couple fertility

I.J.SINGH
Rural Health Training Centre, Chiraigaon-Block, Chiraigaon (Sarnath), Varanasi 221 007
Department of Preventive and Social Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221 005

MS received 8 October 1979; revised 3 March 1980

Abstract. A probability distribution for the number of conceptions during a specified period of time is derived assuming that the fecundability as well as the proportion of incomplete conceptions vary between conceptions. A procedure for finding the estimate of some of the parameters has been outlined. The probability model was tested with the date from the Varnasi survey.

Keywords. Fecundability; complete conception; rest period; fertility.

Introduction

Several workers have derived probability models for describing the variations in the number of conceptions during a given time interval (Dandekar, 1955; Brass, 1958; Singh, 1963, 1964, 1968; Singh and Bhattacharya, 1970, 1971; Sheps and Perrin, 1966). In these models, the fertility parameters such as fecundability as well as the proportion of incomplete conceptions (a conception is called complete if it results in a live birth; otherwise, it is called incomplete) are assumed to be constant throughout the observation period. But empirically, it has been observed that these parameters vary from couple to couple and depend on the age and conception or parity for the same couple. Recently, Singh et al. (1974) proposed a model for the number of births during a few years of married life assuming fecundability to vary with parity. Singh and Singh (1978) extended the above model for conceptions introducing two types of pregnancy outcomes, i.e., complete and incomplete conceptions. In all the above models, the proportion of incomplete conceptions is taken as constant throughout the period of observation. But Gulick (1971) reported that the risk of incomplete conceptions varies according to the order of conceptions and this conception linked risk phenomenon is not geographical but ‘biological’. In addition, the age of the mother, income, environmental conditions, educational status and other variables contribute to the increase in incomplete conceptions. These special conception-risks are always present and must be taken into account in derivation of the probability distribution because such a model would be more akin to the human fertility phenomenon and provide a better estimate of fecundability.

The purpose of this paper is to present a probability distribution model for the number of conceptions during a specified period of time, on the assumptions that the fecundability as well as the proportion of incomplete conceptions depend on conception.

The probability model

Let $X(t)$, $O<t<T$, denote the number of conceptions during the period $(0, t)$ of length t. $\theta_i (i = 1,2, \ldots)$ is the probability that the ith conception is complete so that $(1-\theta i)$ is the probability that it is incomplete. The probability distribution of $X(T)$ is derived under the following assumptions:

(i) The female has led a married life throughout the period $(0, T)$.

(ii) Probability that the first conception occurs during the small time interval $(t, t + \Delta t)$ is $m_0 \Delta t + 0(\Delta t)$, where m_0 is the conception rate in the beginning, $m_0 > 0$, $t > 0$.

(iii) After a conception, there is no possibility of another conception for constant times h_1 and h_2 where h_1, and h_2 are the rest periods (duration of pregnancy and post partum amenorrhoea) associated with incomplete and complete conceptions respectively.

(iv) Given $X(t) = r$, $r = 1,2,3,\ldots$, the conditional probability of a conception during the time interval $(t, t + \Delta t)$ is $m_r \Delta t + 0(\Delta t)$, if the rth conception has occurred prior to $(t - h_2)$ in case of complete and has occurred prior to $(t - h_1)$ when it is incomplete conception, zero otherwise, where m_r is the conception rate after the rth conception.

(v) The rest periods and the time intervals of fecundable state are mutually independent.

(vi) Either the female is exposed to the risk of conception throughout the interval $(0, T)$ except during the rest period; or she is not exposed to this risk at any time during the interval (O, T). Let a and $(1-\alpha)$ be the respective probabilities.

As explained in Singh and Singh (1978), the maximum number of conceptions by a female during the time interval $(0, T)$ cannot exceed n, where $n = [T/h_1]$ if $[T/h_1]=T/h_1$; otherwise, it is equal to $[T/h_1]+1$, where $[T/h_1]$ stands for the greatest integer not exceeding T/h_1 $h_1 < h_2$.

Under the assumptions 1,2,3,4,5 and 6, the distribution function $H_{r+1}(T)$ can be obtained as

$$H_{r+1}(t) = a \prod_{i=1}^{1_r} (1-\theta_i) \left[B(1-e^{-m_s\{t-rh_1)} + \sum_{k_1=1}^{1_r} \frac{\theta_{k_1}}{(1-\theta_{k_1})} \left\{ B(1-e^{-m_s\{t-h_2-\overline{r-1}h_1)} \right\} \right.$$
$$+ \left\{ \sum_{k_1 \neq k_2 = 1}^{1_r} \left(\prod_{j=k_1}^{k_2} \frac{\theta_j}{(1-\theta_j)} \right) \right\} \left\{ B(1-e^{-m_s\{t-2h_2-\overline{r-2}h_1)} \right\} + \cdots\cdots + $$

$$\left\{\sum_{k_1\neq k_2\neq\cdots\neq k_{1_{r-1}}=1}^{1_r}\left(\prod_{j=k_1}^{k_{1_{r-1}}}\frac{\theta_j}{(1-\theta_j)}\right)\right\}\left\{B(1-e^{-m_s(t-\overline{r-1}h_2-h_1)}\right\}+$$

$$\left\{\sum_{k_1\neq k_2\neq\cdots\neq k_{1_r}=1}^{1_r}\left(\prod_{j=k_1}^{k_{1_r}}\frac{\theta_j}{(1-\theta_j)}\right)\right\}\left\{B(1-e^{-m_s(t-rh_2)}\right\}\Bigg] \quad \ldots\ldots\ldots (1)$$

Where B = $\sum_{s=0}^{r}\left(\prod_{\substack{y=0\\y\neq s}}^{r}m_y\right)\Big/\prod_{\substack{y=0\\y\neq s}}^{r}(m_y-m_s)$

The above model is derived considering all $\theta i's$ (i= 1,2,3, . . .,r) and m_y's (y = 0,1,2, . . .,r) to be distinct. However, if $\theta_1 = \theta_1$, $\theta_2 = \theta_3 \ldots = \theta_r = \theta$ and $m_y = m$ for y = 1,2,3, . . .,r, except m_0, the above model reduces to the following form:

$$H_{r+1}(t) = a\left[(1-\theta_1)\sum_{j=0}^{1_{r-1}}\binom{r-1}{j}\theta^j(1-\theta)^{r-j-1}\cdot W+\theta_1\sum_{j=1}^{1_r}\binom{r-1}{j-1}\theta^{j-1}(1-\theta)^{r-j}\cdot W\right]$$

where

$$W = \left[\left(\frac{m}{m-m_0}\right)^r(1-e^{-m_0(t-jh_2-\overline{r-j}h_1)}-m_0\left\{\sum_{s=0}^{r-1}\frac{m^s}{(m-m_0)^{s+1}}\cdot\left(1-e^{-m(t-jh_2-\overline{r-j}h_1)}\cdot\right.\right.\right.$$

$$\left.\left.\left.\sum_{u=0}^{r-s-1}\frac{\{m(t-jh_2-r-jh_1)\}^u}{u!}\right)\right\}\right] \quad \ldots\ldots\ldots\ldots (2)$$

If we replace m by dm_0 where d is constant, equation (2) assumes the following form:

$$H_{r+1}(t) = a\left[(1-\theta_1)\sum_{j=0}^{1_{r-1}}\binom{r-1}{j}\theta^j(1-\theta)^{r-j-1}W'+\theta_1\sum_{j=1}^{1_r}\binom{r-1}{j-1}\theta^{j-1}(1-\theta)^{r-j}W'\right].$$

where

$$W' = \left[1-\left(\frac{d}{d-1}\right)^r\cdot e^{-m_0(t-jh_2-\overline{r-j}h_1)}+e^{-dm_0(t-jh_2-\overline{r-j}h_1)}\cdot\sum_{s=0}^{r-1}\frac{d^s}{(d-1)^{s+1}}\cdot\right.$$

$$\left.\sum_{u=0}^{r-s-1}\frac{\{dm_0(t-jh_2-\overline{r-j}h_1)\}^u}{u!}\right]. \quad \ldots\ldots\ldots (3)$$

Hence the probability of r conceptions during the time interval (0, T) is

$$P[X(T)=r] = H_r(T)-H_{r+1}(T),\ r=1,2,3\ldots\ldots\ldots(n-1).$$

$$P[X(T)=n] = H_n(T),\ H_1(T)=a(1-e^{-m_0T}),\ H_0(T)=1\ \ldots\ldots(4)$$

Derivation of the model

Let successive conceptions occur at times (Z_1, Z_2,. . . ,Z_r, $Z_r + 1$, . . .) so that ($Z_{r+1} - Z_r$) is the time interval between the rth and (r + 1)th conception. This interval of time is the sum of two types of interval: (a) a time interval, denoted by S_r, during which no additional conception can occur, that is the rest period following the *r*th conception and (b) time interval T_r, which is the time between resumption of fecundable state after the termination of *r*th conception and the time of $(r + 1)$th conception. If T_0 is the time of first conception from marriage, then $T_0 = Z_1$, and $S_0 = 0$. From the assumptions 1 and 2, T_0 follows an exponential distribution (see Singh, 1964), with distribution function

$$P\left[T_0 \leq t\right] = 1 - e^{-m_0 T} \qquad \text{.........................(5)}$$

Now ϵ_r and η_r are the two quantities defined by

$$\epsilon_{r+1} = \sum_{i=0}^{r} T_i \, , \; \eta_r = \sum_{j=1}^{r} S_j \, , \; \eta_0 = 0, r > 0 \qquad \text{..........(6)}$$

It is assumed that each of the two sequences $\{T_i, i \geq 0\}$ and $\{S_j, j \geq 1\}$ are mutually independent random variables. Hence $Z_r+1 = \varepsilon_{r+1}+\eta_r$ would be a set of independent random variables. Since, ε_{r+1} is the sum of $(r+1)$ displaced exponentially distributed random variables, the distribution function of ε_{r+1}, denoted by $F_{r+1}(t)$ can easily be obtained as

$$F_{r+1}(t) = P\left[\epsilon_{r+1} \leq t\right] = \sum_{s=0}^{r} \frac{m_0 m_1 \ldots\ldots m_r \int_0^t e^{-m_s \cdot x} \cdot dx}{\prod\limits_{\substack{y=0 \\ y \neq s}}^{r} (m_y - m_s)} \qquad \text{.........(7)}$$

where all m_y's are distinct (see Singh *et al.*, 1974; Singh and Singh, 1978). From assumption 3, S_j follows two-point distribution, given by

$$\left.\begin{array}{l} P\left[S_j = h_2\right] = \Theta_j, \; 0 \leq \Theta_j \leq 1 \\ P\left[S_j = h_1\right] = (1-\Theta_j) \, , \; j = 1, 2, \ldots\ldots, r \end{array}\right\} \qquad \text{.....................(8)}$$

and hence the distribution function of S_j denoted by $G_r(t)$, as obtained by Singh and Singh (1979) is

$$G_r(t) = P\left[\eta_r \leq t\right] = \prod_{i=1}^{1_r} (1-\Theta_i) \left[1 + \sum_{k_1=1}^{1_r} \frac{\Theta_{k_1}}{(1-\Theta_{k_1})} + \sum_{k_1 \neq k_2 = 1}^{1_r} \left\{ \prod_{j=k_1}^{k_2} \frac{\Theta_j}{(1-\Theta_j)} \right\} + \right.$$

$$\sum_{k_1 \neq k_2 \neq k_3 = 1}^{1_r} \left\{ \prod_{j=k_1}^{k_3} \frac{\theta_j}{(1-\theta_j)} \right\} + \cdots\cdots + \sum_{k_1 \neq k_2 \cdots \neq k_{1_{r-1}} = 1}^{1_r} \left\{ \prod_{j=k_1}^{k_{1_{r-1}}} \frac{\theta_j}{(1-\theta_j)} \right\} + \sum_{k_1 \neq k_2 \cdots \neq k_{1_r} = 1}^{1_r} \left\{ \prod_{j=k_1}^{k_{1_r}} \frac{\theta_j}{(1-\theta_j)} \right\} \Bigg]$$

$$\text{where } 1_r \quad \min \left[r, \frac{t - rh_1}{h_2 - h_1} \right] \quad \cdots\cdots\cdots (9)$$

Now the distribution function of $Z_r + 1$, denoted by $H_r + 1\ (t)$ is

$$H_{r+1}(t) = \int_0^t F_{r+1}(t-x)\, dG_r(x) \quad \cdots\cdots\cdots(10)$$

Using equations (7) and (9) in equation (10), on simplification and considering also the assumption (6), the final expression is obtained.

Application of the model

For showing the utility of the probability model, we must have accurate data for the number of conceptions. Due to non-availability of such data for a specified period of time, the model can be applied to the data on the number of births. For example, its application may be seen, where, one to one correspondence between a conception and a live birth is assumed and the children dying within a year and those surviving for more than a year are accounted separately. In this connection, it should also be noted that the rest period following an infant death is considerably shorter than that occurring when the child surviving at least for one year. A similar situation applies to incomplete and complete conceptions. Therefore, for illustration, the present model is applied to the data collected in the Varanasi Survey 1969-70. A detailed account of the survey is given by Singh *et al.* (1970).

For this suvey, an analysis of birth intervals was done according to the life table techniques, which was similar to that of Sheps (1965) and it revealed that the average period of time in the case of first few births, with the exception of first are approximately the same and these intervals for two consecutive births are comparatively smaller than the interval of first birth from marriage for those females who were married between the ages 13 to 16 years. Similar results have been obtained by Jain (1964) and others. Gulick (1971) reported that the risk of death of a first born child is larger than that for the second, third, fourth or fifth. In accordance with this empirical observation, it is desirable to take $\theta_1 = \theta_1$, $\theta_2 = \theta 3$ $=, \ldots, = \theta_r = \theta$ and $m_y = dm_0 (y = 1,2,3, \ldots, r)$ for illustrative purpose.

Again, if the sample of couples is taken from a population which is mixture of two types of populations, *viz.,* (a) the population of females, which have longer duration of post partum amenorrhoea when the new born survives more than a year and have shorter duration of post partum amenorrhoea when the new born dies within a year and (b) the population of females, who have shorter duration of post partum amenorrhoea and equal length of intervals whether the new born dies within a year or survives more than a year. Therefore, the probability model may be expressed as

$$P\left[X(T)=r\right] = \pi\left[P\left\{X(T)=r/h_1,h_2\right\}\right] + (1-\pi)\left[P\left\{X(T)=r/h'_1,h'_2\right\}\right] \qquad \ldots\ldots\ldots\ldots(11)$$

where = $h'_1 = h'_2$, = h_1 and π and (1 — π) are the proportion of subgroups.

The present model is applied for the number of births during a period of ten years. When any distribution is applied to the observed data, reasonable values for parameters such as h_1, h_2, θ_1, θ and π can be assumed. Hence h_1 = 1.00 year (9 months for gestation and 3 months for post partum amenorrhoea), h_2 = 1.75 years (9 months for gestation and 12 months for post partum amenorrhoea), π = 0.50 (Singh and Singh 1978) and θ_1 = 0.75, θ = 0.85 (Gulick, 1971) are assumed. The estimates of the remaining parameters *a, d* and m_0 are calculated by equating the mean, variance and relative frequencies of females with zero birth of the observed distribution to their theoretical expression: i.e.,

$$\bar{x} = \sum_{r=0}^{n} r.P_r = aY_1(m_0,d) \quad (\text{say}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$s^2 = \sum_{r=0}^{n} r^2.P_r - \left\{aY_1(m_0,d)\right\}^2 = Y_2(a,m_0,d) \quad (\text{say}) \ldots (13)$$

$$f_0 = 1-a+a\,e^{-m_0T} \qquad \ldots\ldots\ldots (14)$$

where $\bar{x}$, s^2 and f_0 are the mean, variance and proportion of couples having no birth in the observed distribution respectively. Y_1 (m_0, d) and Y_2 (a,m_0, d) are the function of m_0, *d* and *a*, m_0, *d* respectively.

For large *T*, approximate expression for the mean of the distribution similar to those of Brass (1958) and Singh and Singh (1978), can be obtained as

$$\bar{x} = a\left[\pi\left\{1+\frac{dm_0T'}{1+dm_0\bar{h}(1-\bar{h}/2T')}\right\} + (1-\pi)\left\{1+\frac{dm_0T''}{1+dm_0h_1(1-h_1/2T'')}\right\}\right]$$

$$\text{where } T' = T-\bar{h}-1/m_0,\ T'' = T-h_1-1/m_0,\ \bar{h} = \frac{(\theta_1+\theta)h_2}{2} + \frac{\{(1-\theta_1)+(1-\theta)\}h_1}{2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(15)$$

Now m_0 can be obtained from (14) for an assumed value of *a*. By using the equation (15) and the values of m_0 and *a,* an estimate of *d* is obtained. A set of such estimates of parameters *a*, m_0 and *d* denoted by *a*, m_o and *d* which satisfy the equation (13) is taken as a first approximation. Further, the estimated values of these parameters may be refined according to the equations (12), (13) and (14) respectively. The values of estimates are

$$\hat{a} = 0.977\ , \hat{m}_0 = 0.339\ , d = 1.63\ .$$

On the basis of these values, the expected frequencies of column 3 of table 1 are calculated. For $\theta_i = \theta$ ($i = 1,2,3, \ldots, r$) the model reduces to that given by Singh and Singh (1978). The same model is also applied using the estimated values of

$$\hat{a} = 0.977\ , \hat{m}_0 = 0.339\ , d = 1.58\ .$$

The expected frequencies in column 4 of table 1 are calculated on the basis of these values. The values of x^2 shows that the proposed model is a better approximation for the situation. However, other functional form in regard to conceptions, may also produce the better approximation.

Table 1. Correlation of the observed and expected number of women with the number of births during 10 years of marriage.

No. of births	Number	Expected number	
		$\theta_i = 0.75$, $\theta = 0.85^a$	$\theta = 0.85^b$
0	36	36.0	36.0
1	62	60.0	62.3
2	143	138.9	145.1
3	165	175.5	172.6
4	101	104.6	100.4
5	41	29.2	
6 and over	2	7.8	33.6
Total	550	550.0	550.0
λ^2	—	1.70	3.00
Degrees of freedom		2	2

[a] According to Gulick (1971).
[b] Present model.

Acknowledgements

The author is extremely thankful to Drs S. N. Singh and R. C. Yadava, Department of Mathematics and Statistics, Banaras Hindu University, Varanasi for their valuable suggestions.

References

Brass, W. (1958) *Popul. Stud.*, **12,** 51.
Dandekar (1955) *Sankhya*, **15,** 237.
Gulick, F. A. (1971) *Parity, contraception and infant mortality: A note on some parallel relationships,* Proceedings of All India Seminar on Demography and Statistics, Banaras Hindu University, Varanasi. p. 21.

Jain, S. P. (1964) *J. Family Welfare,* **10,** 16.
Sheps, M. C. and Perrin, E. B. (1966) *Human Biology,* **38,**180.
Sheps, M. C. (1965) *Popul. Stud.,* **19,** 65.
Singh, S. N. (1963) *J. Am. Stat. Assoc,* **58,** 721.
Singh, S. N. (1964) *Sankhya,* **B26,** 89.
Singh, S. N. (1968) *J. Am. Stat. Assoc,* **63,** 209.
Singh, S. N. and Bhattacharya, B. N. (1970) *Biometrics,* **26,** 33.
Singh, S. N. and Bhattacharya, B. N. (1971) *Sankhya,* **B33,** 315.
Singh, S. N., Bhattacharya, B. N. and Yadava, R. C. (1974) *Sankhya,* **B36,** 93.
Singh, S. N. and Singh, I. J. (1978) *Health and Population—Perspectives and issues,* **1,2,**153.
Singh, S. N., Yadava, R. C. and Bhaduri, T. (1970) *Prajna,* **16,**176.