S. Mitra

Probabilities of Childbirth Distribution of Completed Family Size

1. The Problem

IN an earlier paper, I had developed a method to determine the distribution of women by completed family size (number of live children born) from annual probabilities of live childbirths. In the derivation of the results, a year was regarded as the unit of time in which a birth may or may not occur and it was further assumed that these annual probabilities were independent of one another. Although, algebraic expression of the distribution function is easy to obtain in terms of annual probabilities, the actual derivation, in numerical terms, depends on the availability of, and simplifications of terms involving thirty or more probabilities. For operational simplicity, one may assume that over the reproductive period of say 15 to 44 years of age, the probabilities of childbirth remain constant within each five year interval. That is to say, estimates of the average probabilities within such intervals could be regarded as sufficient for an understanding of the pattern of their distribution. This approach is compatible with the fact that the comparable data on age-specific fertility rates, when they are available, are usually obtained by five year intervals. Further, it lends itself to the algebraic treatments necessary for the solution of the converse problem, namely, the derivation of probabilities of childbirth from the distribution of completed family size. A method for estimating such probabilities has been outlined in this paper.

2. Probability Functions

For the sake of continuity, as well as for subsequent use, certain results from the earlier work are presented here. For the six five-year intervals let p_i denote the annual fixed probabilities within each interval (i= 1, 2,... 6). The probability that a woman will not bear any child during the entire reproductive period is

$$P(0) = \Pi (1-p_i)^5, \tag{1}$$

of only having one live birth is

$$P(1) = 5A_1 P(0), \tag{2}$$

where

$$A_1 = \sum_i \frac{P_i}{1-p_i}. \tag{3}$$

In general, the probability of r live births is given by (for $r \geqslant 1$)

$$P(r) = \frac{5}{r}[A_1 P(r-1) - A_2 P(r-2) + \ldots + (-1)^{r-1} A_r P(0)], \tag{4}$$

where

$$A_j = \sum_i \left(\frac{p_i}{1-p_i}\right)^j. \tag{5}$$

3. Analysis of Equations

It is easy to see that the values of A_j can be successively obtained from the distributions of women by number of live births. For example,

$$A_1 = \frac{P(1)}{5P(0)}, \tag{6}$$

where P (1) and P (0) are the proportions of women with only one live birth and none respectively. Once A_1 is obtained, A_2 can be solved from the expression

$$P(2) = \frac{5}{2}[A_1 P(1) - A_2 P(0)] \tag{7}$$

and so on. Estimates of p_i can then be obtained by solving the non-linear equations shown in (5). It is well known that solutions of such equations are tricky, and not necessarily unique and identifiable, and whenever possible, can be obtained as approximations by iterative methods. In this particular instance, the fact is that p_i's are never greater than 0.5 (in fact, less than 0.3), so that the inequality

$$0 \leqslant \frac{p_i}{1-p_i} < 1 \tag{8}$$

holds, facilitating the search for solutions by establishing certain boundary conditions. For simplicity in writing, let

$$x_i = \frac{p_i}{1-p_i},$$

so that

$$p_i = \frac{x_i}{1-x_i}. \tag{9}$$

Because, $0 \leqslant x_i \leqslant 1$ (in fact, $0 \leqslant x_i \leqslant .4$), x_i^j rapidly converges to zero as j increases. As a result, for a large j, $A_j = \Sigma x_i^j$ is dominated by x or by the maximum of the x_i's. Accordingly, an upper bound of x is provided by the equation

$$x^u = (A_j)^{1/j} \tag{10}$$

for large j. Similarly, a lower bound of x can be obtained from

$$x^l = \frac{A_j}{A_{j-1}}. \tag{11}$$

The rare possibility of more than one maximum (when the annual probability of childbirth in the age interval 20-24 is the same as that in the interval (25-29) can be checked by comparing the j-th power of (11) with A_J. That is to say, the condition

$$2\left(\frac{A_j}{A_{j-1}}\right)^j < A_j \qquad (12)$$

is necessary for two maxima (the case of more than two maximums can be safely ignored), in which case the upper bound of these maxima can be obtained from

$$x^u - (A_j/2)^{1/j} \qquad (13)$$

in most cases, though there is one maximum, usually in the age interval 20-24, which corresponds to x_2. In diminishing order of magnitude, the remaining x's can be arranged as x_3, x_4, x_1, x_5 and x_6.

It may be pointed out that the boundaries of x_2 as shown in (10) and (11) can be narrowed further by trying to establish the boundaries of x_3 or the second maximum. For example, the upper boundary of x_3 can be obtained from

$$x_3^u = \left[A_j - \left(\frac{A_j}{A_{j-1}}\right)^j\right]^{1/j} \qquad (14)$$

for the largest j, since A_j/A_{j-1} provides the lower boundary for the first maximum. Similarly, a lower boundary of x_3 is provided by the maximum of

$$x_3^l = \frac{A_j-(x_2^u)^j}{A_{j-1}-(x_2^u)^{j-1}} \qquad (15)$$

for $j = 2, 3, \ldots$ Observe that the largest value of x_2 produces the smallest value of (15) for all j, as the derivative with respect to x_2 of

$$x_3 = \frac{A_j - x_2^j}{A_{j-1} - x_2^{j-1}} \text{ is}$$

$$\frac{dx_3}{dx_2} = \frac{x_2^{j-2}}{A_{j-1} - x_2^{j-1}} [(j-1)\, x_3 - jx_2], \tag{16}$$

which is uniformly negative since the maximum of the x's

$$x_2 > (j-1)\, x_3/j. \tag{17}$$

The lower boundary of x_3, obtained from (15) can be used to reduce the value of A_j, so that the j-th root of

$$A_j - (x_3^l)^j \tag{18}$$

sets a new upper limit of x_2 which is smaller than (10). This upper limit of x_2 can, in turn, be used in (15) to determine the new lower limit of x_3 and the process can be repeated till no important changes in the limits are observed.

Next, the values of A_j and A_{j-1} can be adjusted by the two boundary values of x_3 and the minimum of

$$\frac{A_j - x_3^j}{A_{j-1} - x_3^{j-1}} \tag{19}$$

can be used as the lower limit of x_2. Note that (19) increases until

$$x_3 < \underline{j-1}\, x_2, \tag{20}$$

and decreases thereafter. Accordingly, the lowest value of (19) may be obtained for any of the two boundaries of x_3, but in either cases, the value will be larger than (11)

4. Estimate of x_2

A trial solution of x_2 may be obtained by averaging (10) and (11), or by averaging the limits when the boundaries have been narrowed. Other alternatives are also available, one of which is developed in the following. An examination of (10) and (11) reveals that both limits will converge asymptotically to the actual x_2 as j increases (see Figure 1), where the

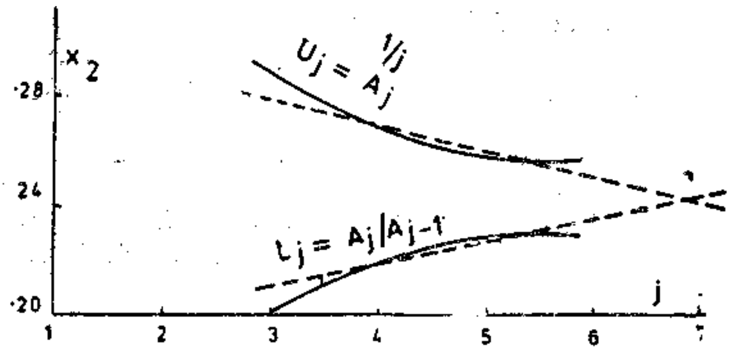

Fig. 1

curves drawn freehand were generated by 1965 US age specific fertility rates. The asymptote, or the solution of x_2 is then estimated by the point of intersection of the two straight lines passing through the two final consecutive values of

$$U_j = (A_j)^{1/j}, \qquad (21)$$

and

$$L_j = A_j/A_{j-1}, \qquad (22)$$

where U and L stand for the upper and lower limits of x_2 respectively. This solution can be written as

$$x_2 = U_j - \frac{(U_{j-1} - U_j)\,(U_j - L_j)}{(U_{j-1} - L_{j-1}) - (U_j - L_j)}. \qquad (23)$$

The estimate of x_2 thus obtained, can next be used to generate

$$B_j = A_j - x_2^j\,,\ \ j = 1, 2 \ldots \qquad (24)$$

so that B_j's can be treated in the same manner as $A_j's$ to produce the estimate of the second maximum or x_3 and the process can be repeated with C_j, D_J, etc. However, because of the sensitive nature of the relationships among numbers raised to higher powers, the successive estimates of x values will depend too much on the number of significant digits left correct in B_j, C_j . . . , etc. It is recommended that the initial values of A_j may be obtained correct upto a certain number of significant digits so that in the end, the final solutions will be correct at least upto the second or third significant digits.

Because of the loss of significant digits, particularly for higher values of j, one may end up with inconsistent U and L values for $B)$, $C,$, etc. For example, the inequalities

$$U_j < U_{j-1}, \; L_j > L_{j-1} \text{ and } U_j > L_j \tag{25}$$

must hold all the time. The reversal of any or all of these inequalities can be caused by significant loss of accuracy in, say, ***Ct*** values which may, in turn be attributed to similar loss in B_j's.

Reasonably accurate estimates appear when x_2 is obtained from A_3, A_t and A_5, x_3 from B_2, B_3 and B_4, x_4 and x_1 from C_1, C_2 and C_3 and D_1, D_2 and D_3 respectively. In each case, the three consecutive A_j, B_j, .. values will generate pairs of U and L values, necessary for the estimate of the corresponding x value. Finally, without caring too much for refinement,
is simply obtained from E_2 as

$$x_5 = \sqrt{E_2}, \tag{26}$$

so that finally,

$$x_6 = E_1 - x_5. \tag{27}$$

The selections of specific A_j, B_j. . . values are naturally, guided by the considerations of their reasonable accuracy, as well as the largest j that narrows down the two limits as much as possible.

5. Applications

The age-specific fertility rates of U. S. and Puerto Rico for the year 1965 were earlier used (Mitra, *ibid)* to generate the distribution of women by number of live births. The latter distributions have been used to examine the extent to which the methods presented in this paper, can reproduce those initial rates which were regarded as approximations of respective probabirities. The results are shown in Table 1, where A_j values defined

TABLE l-ESTMATION OF PROBABILITIES OF CHILD BIRTH FROM DISTRIBUTIONS OF WOMEN BY NUMBER OF LIVE BIRTHS : U.S. AND PUERTO RICO, 1965

j	Age GROUP	*Age specific Fertility Rates*	*j Children Born to Number of women*	A_j	*Estimated Probability*
(1)	(2)	(3)	*(4)*	*(5)*	*(6)*
		U.S.			
0			43		
I	15-19	.0713	146	.6824200	.0779
2	20-24	.1968	236	.1171320	.1949
3	25-29	.1625	142	.0237411	.1655
4	30-34	.0950	176	.0051832	.0928
5	35-39	.0464	96	.0011738	.0374
6	49-44	.0128	42		.0160
7+			19		
		Puerto Rico			
0			12		
1	15-19	.1042	60	.9708900	.1085
2	20-24	.2574	139	.2176459	.2583
1	25-29	.1911	202	.0594176	.1976
4	30-34	.1170	212	.0180023	.1142
5	35-39	.0885	169	.0057557	.0899
6	40-44	.0417	106		.0346
7+			100		

in Eq. (5) are successively obtained from Equations (6), (7) and (4). The number of live births are shown per one thousand women past reproductive age interval, that is forty-five years of age or over.

The estimated probabilities shown in column (6) and for U. S. and Puerto Rico compare quite well with the respective age specific fertility rates, where the latter were used as probabilities to derive $A_j's$ and thereafter, the distribution of women by number of children born. These results demonstrate the solutions and, therefore, can be used to translate the child bearing pattern of women with completed fertility into a series of probabilities that these women experienced during their reproductive period. In this connection it may be pointed out that the above procedure is by no means restricted to women of 45 years of age or over but can be applied to distributions that covers a fraction of the reproductive period. That is to say, the distributions of women by number of children born during the age interval, say 15-40 can be treated in the same manner to estimate the first five probabilities. That way, one can compare, in probabilistic terms, the fertility experiences of different cohorts of women, and to extend it even further, among different groups (socio-economic or others) of the same or different cohorts of women.

6. Discussion

The derivation of probabilities of childbirth at different ages was shown to depend on the solutions of non-linear equations which are reasonably simple in forms. In this paper, only five equations were used for solving six unknowns which, although somewhat unusual, did produce the desired solutions. Admittedly, additional equations could have produced even better estimates, but these were not used for reasons of simplicity, and also because of the fact that the data needed to generate those equations may not always be available or reliable. In any case, the users of this method may, if they so desire and when the data so permit, introduce additional equations in the system. However, the examples used in this paper, seem to justify the adequacy of fewer equations than the number of unknowns. To this end, one may like to compare the A_j values generated from the data with those reproduced by the solutions, since in actual situations, the estimated probabilities will have nothing to compare with, as age-specific birth rates, when they are available, may not be regarded as the best substitutes for those probabilities. The following table presents such a comparison between the observed and estimated *Aj* values which appear to be quite close to one another.

TABLE 2—VALUES OF *Aj* DERIVED FROM THE ESTIMATED PROBABILITIES COMPARED WITH THOSE OBTAINED FROM THE DISTRIBUTION OF WOMEN BY NUMBER OF CHILDREN BORN : U.S. AND PUERTO RICO, 1965.

Values of A_j				
j	U.S.		Puerto Rico	
	Observed	Estimated	Observed	Estimated
(1)	(2)	(3)	(4)	(5)
1	.68242	.68242	.97089	.97089
2	.11713	.11736	.21765	.21891
3	.02374	.02373	.05942	.05956
4	.00518	.00515	.01800	.01789
5	.00117	.00115	.00576	.00566

7. Summary

Given the probabilities of childbirth at different ages and given the independence of these probabilities, the distribution by number of live births of women of completed family size can be generated quite easily. This was done earlier with the additional simplifying assumption that there are basically six different probabilities corresponding to six traditional five year intervals beginning at age fifteen.

The converse problem, namely, the derivation of these probabilities from the generally available distribution of women by number of children born, is somewhat involved. The evaluation of these probabilities rests on the solution of a set of non-linear equations which are generated from the given distribution. Method outlined here for their solutions produced satisfactory results. With certain modifications, it can be extended to cases where the distributions refer to women who are yet to complete their child-bearing period. Estimates of such probabilities will also provide the location as well as the magnitude of differences among different populations or among subgroups of the same population.

Reference

Mitra, S., Probability measures of distributions of women by aumber of childbirths, *Demography India*, **II**(2), 221-226.,