S. Mitra

Probability Measures of Distributions of Women by Number of Childbirths

1. Statements

FROM a theoretical standpoint, a distribution of women by completed family size can be derived if the probabilities of conception or of childbirth are known for specific intervals throughout the reproductive period. Even though, a single year may be regarded as a time interval during which a birth may occur once and no more than once (excluding multiple births), a set of yearly probabilities will render the derivation of the distribution of completed family size somewhat involved, because of operational complications in handling thirty or so probability values. The derivations can be considerably simplified if one assumes that over the reproductive period of say 15 to 45 years of age, the probabilities of childbirth remain constant within each five year age interval. Although a year is still regarded as the unit of time in which a birth may or may not occur, and further that these annual probabilities are independent of one another, the assumption of the constancy of the probabilities over a five year age interval has the added advantage that in most cases, fertility data are available in the form of five year age specific birth rates, which as such, or estimates derived thereof, can be substituted for the average annual probabilities for the respective intervals. Also, the assumption of independence in this context does not seem to be inappropriate in view of the fact that the age-specific rates are themselves affected by factors like marital status, attitude towards family planning etc. The idea stems from the consideration that while for an individual the annual probabilities may not be independent of one another, the assumption of independence may still hold for a large group. In any event, the distributions derived from such assumptions can be tested for internal consistencies or can be compared with actual distributions when such data are available.

2. Derivation of Probability Measures

For the six five yearly intervals *let* p_i, denote the fixed annual probabilities $(i = 1, 2,$ $3. 4, 5, 6)$. The probability that a woman will not bear any child during the entire ıcproductive period is given by

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

and the probability of only one child being born is

$$P(1) = \sum_j {}^5C_1 p_j (1-p_j)^4 \prod_{i \neq j} (1-p_i)^5$$

since the child can be born in any of the six intervals j will range from 1 to 6. The expression can be rewritten as :

$$P(1) = \sum_j {}^5C_1 \left(\frac{p_j}{1-p_j}\right) p(0)$$

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

where

$$A_1 = \Sigma \frac{p_j}{1-p_j} \tag{3}$$

In order to derive the probability measure for two children, one has to note that these children may be born in any one or in any two of the six age intervals. Accordingly,

$$P(2) = \sum_j {}^5 C_2\, p_j^2\, (1-p_j)^3 \prod_{j \neq i} (1-p_i)^5$$

$$+ \sum_{j<k} ({}^5C_1)^2 p_j p_k (1-p_j)^4 (1-p_k)^4 \prod_{j,\, k \neq i} (1-p_i)^5$$

$$= 10 \sum_j \left(\frac{p_j}{1-p_j}\right)^2 p(0) + 25 \sum_{j<k} \left(\frac{p_j}{1-p_j}\right)\left(\frac{p_k}{1-p_k}\right) P(0)$$

$$= \left[10\left(\Sigma \frac{p_j}{1-p_j}\right)^2 + \frac{25}{2!} \sum_j \left(\Sigma \frac{p_j}{1-p_j}\right)\left(\Sigma \frac{p_k}{1-p_k} - \frac{p_j}{1-p_j}\right)\right] P(0)$$

removing the restriction $j<k$. Letting

$$A_2 = \Sigma \left(\frac{p_j}{1-p_j}\right)^2 \tag{4}$$

the expression for $P(2)$ simplifies into

$$P(2) = \frac{P(0)}{2!} (25A_1^2 - 5A_2)$$

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

because of (2) and (3).

For three children who may be born in any one, two or three of the six five year intervals, the corresponding expression

$$P(3) = \sum_j {}^5C_3\left(\frac{P_j}{1-p_j}\right)^3 P(0) + \sum_{j,k} {}^5C_2\,{}^5C_1\left(\frac{P_j}{1-p_j}\right)^2\left(\frac{p_k}{1-p_k}\right)P(0)$$

$$+\sum_{j<k<1}\left({}^5C_1\right)^3\left(\frac{P_j}{1-p_j}\right)\left(\frac{p_k}{1-p_k}\right)\left(\frac{p_1}{1-p_1}\right)P(0) \quad (6)$$

The factor of $P(0)$ in the second expression of (6) can be written as

$$50\Sigma\left(\frac{p_j}{1-p_j}\right)^2\left(A_1-\frac{p_j}{1-p_j}\right) = 50\left(A_1 A_2 - \sum_j\left(\frac{P_j}{1-p_j}\right)^3\right) \quad (7)$$

Similarly, the factor of $P(0)$ of the last expression of (6) can be written as

$$\frac{125}{3!}\sum_{j,k}\left(\frac{P_j}{1-p_j}\right)\left(\frac{p_k}{1-p_k}\right)\left(A_1-\frac{p_j}{1-p_j}-\frac{p_k}{1-p_k}\right) \quad (8)$$

after removing the restriction $j < k < 1$. Writing

$$A_3 = \Sigma\left(\frac{p_j}{1-p_j}\right)^3 \quad (9)$$

and simplifying (7) and (8) in the same manner (6) is reduced to

$$P(3) = P(0)\left[10A_3+50(A_1A_2-A_3)+\frac{125}{3!}\{A_1(A_1^2-A_2)-2(A_1A_2-A_3)\}\right]$$

$$= \frac{5P(0)}{3!}(25A_1^3-15A_1A_2+A_3)$$

which can be further simplified as

$$P(3) = \frac{5}{3}\{A_1P(2)-A_2P(1)+A_3P(0)\} \quad (10)$$

Although the technique of generating function does not seem to be applicable in determining the general form of $P(r)$, it is apparent (and as can be shown) that for $r \geqslant 1$

$$P(r) = \frac{5}{r}\{A_1P(r-1) - A_2P(r-2) + \ldots (-1)^{r-1}A_rP(0)\} \quad (11)$$

where $$A_r = \Sigma\left(\frac{p_j}{1-p_j}\right)^r \quad (12)$$

and for $r > 5$, one has to remember that all of the r children cannot be born in any one five year age interval.

3. Applications

The results obtained earlier have been experimented upon 1964 U.S. and Puerto Rican age specific fertility rates assuming that those can be approximated as probabilities for the corresponding five year age intervals. It must be remembered that since these rates apply to different cohorts of women, the probability measures derived from them will have to be interpreted like the total fertility rate which is obtained by summing the age-specific rates. That is to say, provided a cohort of women experienced a similar set of age-specific rates during their reproductive period, their completed fertility distributions could be obtained from (1) and (12) when these rates are assumed as approximations for respective probabiliti es.

TABLE 1
U.S. (1965) AND PUERTO RICAN (1964) AGE-SPECIFIC BIRTH RATES

	Age-specific birth rate (per woman)	
Age-group	*U.S.*	*Puerto Rico*
(1)	*(2)*	*(3}*
15-19	.0713	.1042
20-24	.1968	.2574
25-29	.1625	.1911
30-34	.0950	.1170
35-39	.0464	.0885
40-44	.0128	.0417
Total fertility rate	2.9240	3.9995

SOURCE: (1) U.S. National Center for Health Statistics, Vital Statistics of the United States, 1965, Vol. 1, Natality, 1967.
(2) Department of Health, 'Annual Vital Statistics Report' 1965, San Juan, Puerto Rico.

The data were chosen to demonstrate the divergence of the probability measures under differing fertility conditions. The latter is evident from the difference in the two total fertility rates obtained by summing the five year age-specific fertility rates and multiplying the results by five. This rate shows the number of children a woman 45 years old or over, can expect to bear under fertility conditions depicted by a given series of age-specific rates. Thus, in Puerto Rico, the expected number is four as against less than three in the United States.

The distribution by completed family size has been shown next for 1000 women along with the corresponding A_r values.

TABLE 2
COMPLETED FAMILY SIZE EXPECTED IN U.S. AND PUERTO RICO AND VALUES OF A_r

Number of children	*Number of women*		*Values of A_r*	
	U.S.	*Puerto Rico*	*U.S.*	*Puerto Rico*
(1)	*(2)*	*(3)*	*(4)*	*(5)*
0	43	12		
1	146	60	.682	.971
2	236	139	.117	.218
3	242	202	.024	.059
4	176	212	.005	.018
5	96	169	.001	.006
6	42	106	.000	.002
7	14	55		.001
8+	5	45		
Total	1000	1000		
Average	2.92	4.02		

The pattern of the above distributions seems logical, and the average numbers of children per woman compare favorably with the total fertility rates, as they should. It may be noted that these distributions can not be compared with those obtained for a generation of women, since the former are based on cross sectional rather than longitudinal fertility history.

It is interesting to observe the rapidly declining trend of A_r and this tendency is even stronger where total fertility rate is low. This goes on to suggest that in estimating the formula (11) of .P(r), terms beyond the third or the fourth may be ignored without any significant loss of accuracy.

In the foregoing analysis, all women have been assumed to have identical probability of having a baby in any age interval. Although there are admittedly important exceptions to this assumption, one factor that can be particularly singled out is the childlessness due to either nonfecundity or deliberate avoidance of conception. This will be helpful because for these women, the probability of giving birth to a child is always zero regardless of the status of other factors that affect fertility. This factor can be taken into consideration by making a straightforward assumption that the proportion of childlessness due to the aforementioned reasons, remains constant at all ages, and then modifying the age-specific rates accordingly.

In other words, if this proportion of childlessness is denoted by c, then the age-specific fertility rates of the remaining women will be inflated by $1/(1-c)$, and the values of $P(r)$ obtained in the usual manner. Later, these $P(r)$ values will have to be deflated by the factor $1 - c$, except that for $r = 0$, the constant c is also added to the deflated $P(0)$. The effect on the distribution of women by number of children can be seen in Table 3 for the United States for $C = .05$.

TABLE 3

1000 $P(r)$ VALUES FOR $C = .05$

	U.S. (1965)								
r	0	7	2	3	4	5	6	7	8+
1000 $P(r)$	84	123	311	339	177	103	47	17	9

The values of 1000 $P(r)$ can be compared with the corresponding figures shown in Col. (2) of Table 2. For zero parity, the number is higher in Table 3, as could be expected. The smaller numbers in parity one, two and three are compensated by an increase in the remaining family size categories:

Discussion

The probability measures shown in (1) and (11) for the distribution of women by number of children born can be generalized to develop similar measures for age-groups of different sizes. The appearances of the number five as an exponent in (1) and a factor in (11) for the five year age-groups, make it clear that for a size-class n, the number 5 will have to be replaced by n in these expressions with appropriate adjustments for the limits of i in (1). In particular, for single year age-groups, equations (1), (11) and (12) will assume the following forms.

$$P(0) = \prod_{i=1}^{30} (1-p_i). \tag{13}$$

$$P(r) = \frac{1}{r} \sum_{i=1}^{r} (-1)^{i-1} A_i P(r-i). \tag{14}$$

$$A_r = \sum_{i=1}^{30} \left(\frac{p_i}{1-p_i}\right)^r. \tag{15}$$

Understandably, the single year probabilities can be expected to produce better results and should be used when such data are available. However, expressions in terms of five year age groups can not only be used in most countries where birth registration statistics are adequate, but they also appear to lend themselves to algebraic treatments leading to estimates of p_is from the distribution of women by number of children born. This will be attempted in a subsequent investigation.