Amarendra Mishra

On the Fitting of a Mathematical Model to the Statistics of Age at First Marriage

Introduction

FOR a person who is married, let x be a variable which assumes the value zero if the person is married before attaining 20 years of age, one, if he or is married between 20-24 years, two, if between 25-29 years and so on. Our attempt here is to find a suitable mathematical model for the frequency distribution compiled in this manner.

The variable x can also be regarded as a number of failures preceeding the first success, if we consider the cases of first marriage only during a particular year in a particular region i.e. the marriages of those persons only who were bachelors (or spinsters) before their marriage. Uspensky (1937) discussed a problem for a sequence of independent trials with a constant probability p in each trial. But discussing the distribution of first births Mukerji (1965) did not assume constant probability of success and independence of trials. He assumed that the probability p_i of getting a success in the $(i + 1)$th trial when it is known that first i trials resulted in failure, $i = 0, 1, 2, \ldots$, increases as i moves from zero to a certain value s and decreases monotonically as i moves from s to a value $t(\geqslant s)$ and then remains constant thereafter. (See also Srivastava and Prasad, 1971; Mishra, 1976). We make the same assumption as Mukerji in our case.

Suggested Model

The probability that the person is married in the i-th age group (i.e. $x = i$) is given by

$$P(x = i) = P(x \geqslant i) \cdot P(x = i/x \geqslant i)$$

i.e. this probability can be expressed as the product of the probabilities that the same did not marry in the preceding i age-groups and the same marries in the i-th age-group given that he or she did not marry in the preceding i age-groups. Writing

$$P(x = i/x \geqslant i) = p_i$$

and

$$1 - p_i = \bar{p}_i.$$

We have

$$P(x = 0) = p_0 \qquad \text{as } P(x \geqslant 0) = 1$$

$$P(x = 1) = \bar{p}_0 p_1$$

$$P(x = 2) = \bar{p}_0 \bar{p}_1 p_2$$

$$P(x = 3) = \bar{p}_0 \bar{p}_1 \bar{p}_2 p_3$$

$$\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots$$

$$P(x = k) = \bar{p}_0 \bar{p}_1 \bar{p}_2 \bar{p}_3 \ldots \bar{p}_{k-1} p_k$$

$$P(x > k) = \bar{p}_0 \bar{p}_1 \bar{p}_2 \bar{p}_3 \ldots \bar{p}_{k-1} \bar{p}_k$$

The different probabilities $p_0, p_1, p_2 \ldots$ are defined in the following manner.

$$p_0 = \alpha$$

$$p_1 = \alpha + \beta$$

$$p_2 = \alpha + r\beta$$

and

$$p_1 = \alpha + (r - \mu)\beta \text{ for all } i = 3, 4, \ldots$$

α, β, r and μ are the four parameters of the model. Obviously $s = 2$ if $r > 1$ and $s = 1$ if $r \leqslant 1$. The value of t is 3.

Maximum Likelihood Estimates of the Parameters

If $f_0 f_1 f_2 \ldots f_{k-1} f_{k+}$ are the respective frequencies for $x = 0, 1, 2, \ldots k-1, k+$, the likelihood function L is given by

$$L = \alpha^{f_0}(1-\alpha)^{N-f_0}(\alpha+\beta)^{f_1}(1-\alpha-\beta)^{N-f_0-f_1}(\alpha+r\beta)^{f_2}(1-\alpha-r\beta)^{N-f_0-f_1-f_2}$$

$$(\alpha + \overline{r-\mu}\beta)^{\sum_{3}^{k-1} f_i}(1-\alpha-\overline{r-\mu}\beta)^{T} \tag{1}$$

where $N = f_0 + f_1 + f_2 + \ldots$ and $T = f_4 + 3f_5 = 4f_6 + \ldots$

$$\begin{aligned}\log L = {} & f_0 \log\alpha + (N-f_0)\log(1-\alpha) + f_1\log(\alpha+\beta) + \\ & + (N-f_0-f_1)\log(1-\alpha-\beta) + f_2\log(\alpha+r\beta) + \\ & + (N-f_0-f_1-f_2)\log(1-\alpha-r\beta) + \\ & + \sum_{j=3}^{k-1} f_j \log(\alpha+\overline{r-\mu}\beta) + T\log(1-\alpha-\overline{r-\mu}\beta).\end{aligned} \tag{2}$$

The estimating equations are

$$\begin{aligned}\frac{\partial \log L}{\partial\alpha} = 0 = {} & \left[\frac{f_0}{\alpha} \quad \frac{N-f_0}{1-\alpha}\right] + \left[\frac{f_1}{\alpha+\beta} \quad \frac{N-f_0-f_1}{1-\alpha-\beta}\right] + \\ & + \left[\frac{f_2}{\alpha+r\beta} - \frac{N-\sum_0^2 f_i}{1-\alpha-r\beta}\right] + \\ & + \left[\frac{\sum_{j=3}^{k-1} f_j}{(\alpha+\overline{r-\mu}\beta)} - \frac{T}{(1-\alpha-\overline{r-\mu}\beta)}\right]\end{aligned} \tag{3}$$

$$\begin{aligned}\frac{\partial \log L}{\partial\beta} = 0 = {} & \left[\frac{f_1}{\alpha+\beta} - \frac{N-f_0-f_1}{1-\alpha-\beta}\right] + \left[\frac{f_2}{\alpha+r\beta} - \frac{N-\sum_{j=0}^{2} f_j}{1-\alpha-r\beta}\right]r \\ & + \left[\frac{\sum_{j=3}^{k-1} f_j}{(\alpha+\overline{r-\mu}\beta)} - \frac{T}{(1-\alpha-\overline{r-\mu}\beta)}\right](r-\mu)\end{aligned} \tag{4}$$

$$\frac{\partial \log L}{\partial r} = 0 = \left[\frac{f_2}{\alpha + r\beta} - \frac{N - \sum_{j=0}^{2} f_j}{(1 - \alpha - r\beta} \right] +$$

$$+ \left[\frac{\sum_{j=3}^{k-1} f_j}{(\alpha + r - \mu\beta)} - \frac{T}{(1 - \alpha - r - \mu\beta)} \right] \beta \quad (5)$$

$$\frac{\partial \log L}{\partial \mu} = 0 = \left[\frac{\sum_{j=3}^{k-1} f_j}{(\alpha + r - \mu\beta)} - \frac{T}{(1 - \alpha - r - \mu\beta)} \right] (-\beta). \quad (6)$$

Note from above that as (6) = 0 and $\beta \neq 0$ the expression within the bracket is zero and hence the expression within the first middle bracket of (5) is also zero. This implies that the first bracket of (4) and hence the same of (3) are also zero. The set of above four estimating equations is thus reduced to the following set of equations.

$$\frac{f_0}{\alpha} - \frac{N - f_0}{1 - \alpha} = 0 \quad (7)$$

$$\frac{f_1}{\alpha + \beta} - \frac{N - f_0 - f_1}{1 - \alpha - \beta} = 0 \quad (8)$$

$$\frac{f_2}{\alpha + r\beta} - \frac{N - \sum_{j=0}^{2} f_j}{(1 - \alpha - r\beta)} = 0 \quad (9)$$

$$\frac{\sum_{j=3}^{k-1} f_j}{(\alpha + r - \mu\beta)} - \frac{T}{(1 - \alpha - r - \mu\beta)} = 0. \quad (10)$$

It is apparent from above that the expected frequencies would be exactly equal to the observed frequencies for x = 0, 1 and 2. The difference between the two sets of frequencies would start from x = 3 onwards.

Application of the Model

In the present section we propose to demonstrate the application of the model developed in the preceding sections. The model is used to represent the distribution of marriages according to age-groups in Australia from 1965 to 1969 for both the sexes. These distributions are available in the 'Official Year Book of Australia' a yearly publication of Australian Bureau of Statistics, Canberra, Australia. The reason for selecting Australia is mainly due to the fact that Australia is one of the few countries in the world which has maintained a systematic and reliable statistics on age at those marriages before of which the marital status of a person is bachelor (or spinster). The reason for selecting the above period is the ready availability of data. Since the numbers involved were large, the total number of persons married in each calender year were prorated at various age-groups such that the total was 1000. The transformation leaves the shape of original distribution unaltered and therefore the model fitted to the transformed distribution can be assumed to express the original distribution.

The observed and the expected frequencies of persons married in different age-groups are compared with the help of χ^2. The maximum likelihood estimates of the parameters α, β, r and μ are also given in the following tables.

TABLE 1—GOODNESS OF FIT TO THE OBSERVED DISTRIBUTION OF FIRST MARRIAGES OF MALES ACCORDING TO THEIR AGE-GROUPS, AUSTRALIA, 1965-69

($s = 2$, $t = 3$)

Variate x	Year 1965		1966		1967		1968		1969	
	$\alpha = 0.07000$		0.09100		0.08300		0.07900		0.07700	
	$\hat{\beta} = 0.49022$		0.49756		0.52402		0.55509		0.56980	
	$\hat{r} = 1.17899$		1.17669		1.12357		1.08184		1.06537	
	$\hat{\mu} = 0.19934$		0.26935		0.32840		0.25033		0.30434	
	O	E	O	E	O	E	O	E	O	E
0	70	70	91	91	83	83	79	79	77	77
1	521	521	535	535	557	557	584	584	597	597
2	265	965	253	253	242	242	229	229	223	223
3	81	79	67	66	60	59	59	58	56	53
4	35	36	29	30	29	30	24	27	22	26
5	14	16	13	14	14	15	13	12	13	12
6	7	7	6	6	7	7	6	6	6	6
7+	7	6	6	5	8	7	6	5	6	6
χ^2_3	0.49508		0.31991		0.25981		0.63391		0.86853	

O — Observed frequency E — Expected frequency

TABLB 2—GOODNESS OF FIT TO THE OBSERVED DISTRIBUTION OF FIRST MARRIAGES OF FEMALES ACCORDING TO THEIR AGE-GROUPS, AUSTRALIA, 1965-69

($s = 1, \quad t = 3$)

Variate *x*	*Year* 1965		1966		1967		1968		1969	
$\hat{\alpha}$ =	0.31400		0.33200		0.31100		0.30800		0.30900	
$\hat{\beta}$ =	0.45276		0.45543		0.48145		0.49838		0.49708	
$\hat{\chi}$ =	0.70068		0.72456		0.74960		0.72969		0.75958	
$\hat{\mu}$ =	0.42907		0.37955		0.45634		0.44822		0.43218	
	O	*E*	*O*	*E*	*O*	*E*	*O*	*E*	*O*	*E*
0	314	314	332	332	311	311	338	308	309	009
1	526	526	526	526	546	546	558	558	557	557
2	101	101	94	94	96	96	90	90	92	92
3	27	26	24	23	22	21	21	19	21	19
4	14	15	11	12	11	12	10	11	8	10
5	7	8	6	6	6	6	5	6	6	6
6	5	5	4	3	3	4	3	3	3	3
7+	6	5	3	4	5	4	5	5	4	4
	x_3^2 = 0.43012		x_2^{2*} = 0.12681		x_2^{2*} = 0.13095		x_2^{2*} = 0.46810		x_2^{2*} = 0.61053	

O — Observed frequency *E* — Expected frequency

*χ^2 is calculated after pooling the frequencies of $\chi = 6$ and 7+ as they are less than 5.

References

1. Mishra, Amarendra, 1976, Evaluation of the impact of fertility control from birth order statistics, *Demography India*, **5** (1 and 2), 236-243.
2. Mukerji, S., 1965, On the fitting of a mathematical model to the statistics of a first birth, *Sankhya, Series B* **27** (3 and 4), 265-270.
3. Srivastava, M. L. and Prasad, Rudranand, 1971, A note on Mukerji's model of the distribution of first births, *Sankhya, Series B*, **33** (3 and 4), 323-330.
4. Uspensky, J. V., 1937, *Introduction to Mathematical Probability*, McGraw-Hill Book Company, New York and London.