M. M. M. Farahani

Estimation of fecundability in Homogeneous-Heterogeneous Population

Introduction

In the past, several probability models have been presented to study the mean value of fecundability—Gini (1924) defined the monthly probability of conception in the absence of contraception, outside the gestation time and the unsusceptible period following the termination of a pregnancy.

With Gini's definition, if all women have the identical constant monthly fecundability p (a homogeneous population) then the probability of first conception occurring at month x is :

$$g(x) = pr(X = x) = p(1-p)^{x-1} \qquad x = 1, 2, \ldots \tag{I}$$

Later on Potter and Parker (1964) have presented a heterogeneous population and they suggested the type I geometric distribution as a useful model with the following assumptions :

(i) Each couple has the same constant fecundability p until pregnancy.

(ii) Fecundability is constant for a given woman but it is distributed between them with a Pearson type I curve, i.e. a Beta distribution of parameters a and B.

(iii) The number of couples is large and conception is a random event. In this case if we assume the fecundability of women is constant value p during the period of observation then X (the random month of conception has the conditional geometric distribution.

$$g(x; p) = pr\{X = x \mid p\} = p(1 - p)^{x-1} \qquad x = 1, 2, \ldots \tag{2}$$

and if p varies between women with Beta density function

$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \tag{3}$$

where $0 < p \leqslant 1$ and $\alpha, \beta > 0$ then the unconditional distribution will be as follows :

$$g(x) = pr\{X = x\} = \frac{1}{B(\alpha, \beta)} \int_0^1 p^{\alpha} (1 - p)^{\beta+x-2} \, dp \tag{4}$$

$$= \frac{\alpha}{\alpha + \beta} \qquad x = 1$$

$$= \frac{\alpha\beta(\beta + 1) \ldots (\beta + x - 2)}{(\alpha + \beta)(\alpha + \beta + 1) \ldots (\alpha + \beta + x - 1)} \qquad x \geqslant 2.$$

The α and β are not directly known, but the mean and variance of the waiting times of those becoming pregnant in terms of the same constants α and β are known. Potter and Parker (1964) used method of moment for finding α and β and later on Majumdar and Sheps (1970) developed maximum likelihood estimators that make more effective use of the data. Also see Sheps and Menken (1972) and Das Gupta *et al.* (1974).

Several authors have derived estimators of the mean and variance of conception delay under the assumption that fecundability follows a Beta distribution, specially in simulation, but when results are compared with those obtained when facundability is assumed to be homogeneous and equal to estimated mean, it can be shown that the mean delay calculated using a Beta distribution is generally longer by 2 to 4 months, and the variance is multiplied by a factor of at least four (Jacquard and Leridon, 1973).

The purpose of the present paper is to provide a model having a mix of both homogeneous and heterogeneous populations. As an illustration, this model is applied to two groups of data of the Princeton Fertility study and Hutterite data.

The Model

Let X be the random month of conception. The distribution of X is then obtained under the following assumptions :

(1) All women have identical constant monthly fecundability ρ.

(2) This fecundability is divided into two parts, one is constant between women 'a' (homogeneous population), and the other part varying between women 'p'. (Heterogeneous population)

$$\rho = a + p.$$

(3) Fecundability is constant over time for a woman but p varies among women with a modified Beta distribution

$$f(p) = \frac{1}{b^{\alpha+\beta-1} B(\alpha, \beta)} p^{\alpha-1} (b - p)^{\beta-1} \tag{5}$$

where $b = 1 - a, \alpha, \beta > 0$ and $0 < p < 1 - a$.

(4) Conception is a random event, conditional on fecundability. With these assumptions X has the conditional geometric distribution

$$g(x, p) = pr\{X = x \mid p\} = (a + p)(b - p)^{x-1} \qquad x = 1, 2, \ldots \tag{6}$$

where $b = 1 - a$. The p.d.f. $f(p)$ is given by

$$f(p) = \frac{1}{b^{\alpha+\beta-1} b(\alpha, \beta)} p^{\alpha-1} (b - p)^{\beta-1} \tag{7}$$

where $0 < p < b$, $\alpha, \beta > 0$ and $b = 1 - a$. The mean and the variance of p are

$$E(p) = \bar{p} = \frac{b\alpha}{\alpha + \beta} \tag{8}$$

$$V(p) = \sigma_p^2 = \frac{b^2 \alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} . \tag{9}$$

Then we find that unconditional distribution of number of menstrual intervals needed for the first conception is

$$g(d) = pr\{X = x\} = \int_0^b g(x, p) f(p)\, dp$$

$$= \frac{1}{b^{\alpha+\beta-1}\, B(\alpha, \beta)} \int_0^b (a + p)\, p^{\alpha-1}\, (b - p)^{\beta+x} - 2\, dp \qquad (10)$$

or

$$g(x) = \frac{a\, B(\alpha, x + \beta - 1) + b\, B(\alpha + 1, x + \beta - 1)}{B(\alpha, \beta)}\, b^{x-1} \qquad (11)$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\, \Gamma(\beta)}{\Gamma(\alpha + \beta)} = \int_0^1 x^{\alpha-1}\, (1 - x)^{\beta-1}\, dx \qquad (12)$$

and

$$\Gamma(\alpha) = (\alpha - 1)\, \Gamma(\alpha - 1) = \int_0^\infty x^{\alpha-1}\, e^{-x}\, dx. \qquad (13)$$

In (11) if we put $a = 0$, this will become (4) hetrogeneous population. From (11) we have

$$g(1) = a + b \frac{\alpha}{\alpha + \beta} \qquad (14)$$

and

$$g(x + 1) = b\, \frac{(x + \beta - 1)\,(ax + a\beta + \alpha)}{(x + \beta + \alpha)\,(ax + a\beta + \alpha - a)}\, g(x). \qquad (15)$$

Maximum likelihood estimates of α, β and a. For estimation of α, β and a maximum likelihood estimators are used for uncensored data. The Logarithm of likelihood function (11) is

$$\log L = \text{Const} + \sum_{x=1} = fx \log g(x)$$

where f_x is observed frequencies in month x. Let N be total women and F_x is the cumulative observed frequencies. Then Scores, Rao (1952) for the M. L. estimators of α, β and a are :

$$S_1 = \frac{\partial \log L}{\partial \alpha} = \frac{N}{a\beta + \alpha - a} + \sum_{x=0}^{t} \left[\frac{N - F_x}{ax + a\beta + \alpha} - \frac{N - F_x}{x + \beta + \alpha} - \frac{N - F_x}{ax + a\beta + \alpha - a} \right] = 0$$

$$S_2 = \frac{\partial \log F}{\partial \beta} = \sum_{x=0}^{t} \left[\frac{a(N - F_x)}{ax + a\beta + a} - \frac{N - F_x}{x + \beta + \alpha} \right] + \sum_{x=1}^{t} \left[\frac{N - F_x}{x + \beta - 1} - \frac{a(N - F_x)}{ax + a\beta + \alpha - a} \right] = 0$$

and

$$S_3 = \frac{\partial \log L}{\partial a} = \frac{\sum_{x=1}^{t} (N - F_x)}{1 - a} + \sum_{x=0}^{t} \frac{(N - F_x)(x + \beta)}{ax + a\beta + \alpha} - \sum_{x=1}^{t} \frac{(N - F_x)(x + \beta - 1)}{ax + a\beta + \alpha - a} = 0.$$

Estimates of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{a}$ are derived using the above formula and Newton-Raphson iteration procedure (Pennington, 1970).

Sources of Data

The semi heterogeneous model has been fitted to two distributions relating to first conception, taken from the Princeton Fertility study and Hutterite data. Details of the data are given in the respective papers.

Results

The observed and expected frequencies of conception for Hutterite data and Princeton Fertility Study are shown in Table 1 and Table 2, respectively. The model is found to fit in with the observed data more closely as compared to other models.

TABLE 1-OBSERVED AND EXPECTED FREQUENCIES OF CONCEPTION BY MONTH, HUTTERITE DATA

Month of conception	*Observed frequency*	*Expected frequency*		
		A	*B*	α = 1, β = 3.5 *a* = .1
1	103	92.3	91.34	102.6
2	53	62.4	61.91	63.1
3	43	43.6	43.37	42.0
4	27	31.3	31.19	29.5
5	30	23.0	23.04	21.6
6	9	17.3	17.26	16.2
7	12	13.2	13.25	12.4
8	9	10.2	10.31	9.7
9	6	8.0	8.13	7.7
10	8	6.4	6.50	6.2
11	10	5.2	5.25	5.0
12	5	4.2	4.29	4.1
13-15	9	8.7	8.92	8.5
16-18	7	5.1	5.32	4.9
19-24	7	5.3	5.59	4.8
25 +	4	5.8	6.33	3.7
Total	342	342.0	342.00	342.0

A —Taken from Majumdar and Sheps (1970) for Type I Geometric distribution.
B —Taken from Singh and Bhaduri (1972) for continuous time.

TABLE 2—OBSERVED AND EXPECTED FREQUENCIES OF CONCEPTION BY MONTH, PRINCETON FERTILITY STUDY

Month of conception	*Observed frequency*	*Expected Frequency*		
		A	*B*	α = 1, β = 1.56 *a* = .01
1	380	357.0	328.79	380.0
2	153	173.8	176.87	166.4
3	94	100.9	107.61	93.4
4	45	65.2	71.07	59.8
5-6	93	78.1	86.20	72.0
7-9	51	60.1	66.17	56.9
10-12	46	32.2	34.87	31.6
13-24	68	49.4	51.15	51.9
25-48	20	23.5	21.92	27.6
49 +	8	17.8	13.35	18.4
Total	958	958.0	958.00	958.0

A —Taken from Majumdar and Sheps (1970) for type I Geometric distribution.
B —Taken from Singh and Bhaduri (1972) for continuous time.

References

1. Das Gupta, P. and Hicklman, L., 1974, Estimation of parameters of Type I Geometric distribution from truncated observations on conception delays. *Mathematical Biosciences, 22, 75-94.*
2. Gini, C, 1924, Premier recherches sur la fecondabilite de la femme, pp. 889-892. In : *Proceedings of the International Mathematics Congress,* Toronto.
3. Jacquard, A. and Leridon, H., 1973, Simulating human reproduction : How complicated should a model be? pp. 241-249. In : Dyke B. *et al., Computer Simulation Human Population Studies.* Academic Press, Inc.
4. Majumder, H. and Mindel, C. Sheps, 1970, Estimators of a type I Geometric distribution from observations on conception times. *Demography,* 7, 349-360,
5. Mindel, C. Sheps and Jane, A. Menken, 1972, On estimating the risk of conception from censored data, pp. 167-200. In : Greville, T.N.E., *Population Dynamics.* Academic **Press** Inc.
6. Pennington, Ralph H., 1970, *Computer Methods and Numerical Analysis.* The Macmillan Company, London.
7. Potter, R. and Parker, M. P., 1964, Predicting the time required to conceive. *Population Studies,* 18,99-116.
8. Rao, Calyampudi R., 1952, *Advanced Statistical Methods in Biometric Research.* Wiley, New York,
9. Singh, S. N. and Bhaduri, T., 1972, Maximum likelihood estimates for the parameters of a continuous time model first conception. *Demography,* 9, 249-25.
10. Westoff, Charles F., Robert, G. Pootler, Jr., Phillip, C. Sagi and Elliot, G. Mishler, 1961, *Family Growth in Metropolitan America.* Princeton University Press.