S. M. Shah and K.S. Shah

Estimating of Parameters of Continuous Distribution of Time Between Marriage and First Conception

Introduction

THThe duration of time between two successive births or between marriage and first birth is an indicator of the level of fertility of a couple. For analyzing data on the first conception leading to a live birth, Potter and Parker (1964) and Singh (1961, 1967) suggested the Type I Geometric distribution as a useful model. Potter, Parker and Singh estimated the parameters using the first two moments of the conception months. Majumdar and Sheps (1970) pointed out the limitations of such estimates and gave a method for obtaining maximum likelihood estimates, which possess certain optimal properties. But it is difficult to compute these estimates without the help of a computer.

Singh (1961, 1964) proposed a continuous probability distribution based on another set of assumption for the above situation. He outlined a method to obtain best asymptotically normal estimates of the parameters. These estimates are obtained after several iterations starting from any set of consistent estimates.

Singh proposed the following model (1972).

$$g(t, a, h) = \frac{a h^{a}}{(h + t)^{a+1}} \qquad a > 0, \quad h > 0$$

$$= \frac{a}{h}\left[1 + \frac{t}{h}\right]^{-(a+1)} \tag{1}$$

This distribution is a particular case of beta distribution.

$$B(r, k, g, b) = \frac{1}{\beta(k, q)} \cdot \frac{b^{k} \cdot t^{k-1}}{(1 + bt)^{k+q}} \qquad (2)$$

where $b = 1/h$, $k = 1$ and $a = q$.

In this paper, the gamma distribution is suggested for the distribution of time of the first conception. Further, it is shown how gama distribution is obtained from (1) as a limiting distribution. The gamma distribution is fitted to the observed data collected from the case card records of leading maternity hospitals of Anand and Anand Municipal Hospital, Anand. For sake of comparison, the gamma model is also fitted to Hutterite (Sheps, 1967, pp. 129-132) data.

The Model

Let T be the period from marriage to the first conception leading to a live birth, when the female is exposed to the risk of conception. The distribution of T as given in Singh (1961, 1964) is derived under the following assumptions:

(*a*) The number of coitions during any time interval $(0, t)$ of length t is a random variable and follows the Poisson distribution with parameter $t\lambda_1$ where λ_1 is a positive constant.

(*b*) Coitions are mutually independent and p_1 the probability of a coition resulting in conception is constant.

(*c*) Conceptions are mutually independent and p_2, the probability of a conception resulting in a live birth is constant.

Under the assumptions (*a*), (*b*) and (*c*) the number of live births follows a Poisson distribution with parameter $\lambda = \lambda_1 p_1 p_2$ if conception are assumed to be instantaneous i.e. the related periods of temporary sterility (gestation and post-partum amenorrhea) are zero. In the case of a first birth, it is enough to assume that a conception not resulting in a live birth is instantaneous.

In Singh (1961), the parameters are assumed to be constant for the simplicity of the derivation of the model. Singh also assumed that λ follows a Pearson Type III distribution with parameters a and h.

Let $g(t, a, h)$ be the probability density function of T, under the above assumptions. Then $g(t, a, h)$ is given by

$$g(t, a, h) = \frac{ah^{a}}{(h + t)^{a+1}} \qquad a > 0, \quad h > 0$$

[Singh's Model, 1972]

$$= \frac{a}{h}\left[1 + \frac{t}{h}\right]^{-(a+1)}. \qquad (3)$$

We write the above distribution (3) as a particular case of beta distribution as follows:

$$B(t, k, q, b) = \frac{1}{\beta(k, q)} \frac{b^k \cdot t^{k-1}}{(1 + bt)^{k+q}} \tag{4}$$

where $b = 1/h$, $k = 1$ and $a = q$.

Now we shall consider the limiting case of equation (4) when $b \to 0$, $q \to \infty$ such that $bq = c$, a constant.

$$g(t, c, k) = \lim_{\substack{b \to 0\, q \to \infty \\ bq = c}} B(t, k, q, b)$$

$$= \frac{t^{k-1}}{\sqrt{k}} \lim_{\substack{b \to 0\ q \to \infty \\ bq = c}} \frac{\sqrt{k + q}}{\sqrt{q}} \frac{b^k}{(1 + bt)^{k+q}}$$

$$= \frac{t^{k-1}}{\sqrt{k}} \lim_{q \to \infty} \frac{\sqrt{2\pi}\,(k + q - 1)^{k+q-1/2}\, e^{-(k+q-1)}}{\sqrt{2}\,\pi(q - 1)^{q-1/2}\, e^{-(q-1)}} \cdot \frac{c^k}{q^k\left(1 + \frac{tc}{q}\right)^{k+q}}$$

$$= \frac{c^k}{\sqrt{k}}\, t^{k-1}\, e^{-ct} \quad k > 0,\ c > 0, \tag{5}$$

$$0 < t < \infty$$

which is the required gamma distribution. For $k = 1$, we get exponential distribution.

Fitting of the Model

For fitting the model (5), its parameters are estimated by the method of moments and by the method of maximum likelihood.

The Method of Moments :

The rth moment μ'_r of (5) is given by

$$\mu'_r = \frac{k(k + 1) \ldots (k + r - 1)}{c^r} \tag{6}$$

from which one obtains

$$k = \frac{\mu_1'^2}{\mu_2' - \mu_1'^2}, \quad c = \frac{\mu_1'}{\mu' - \mu_1'^2} \tag{7}$$

Replacing μ'_1 and μ'_2 by sample moment m_1 and m_2 respectively, the estimates of the parameters k and c are obtained as

$$\hat{k} = \frac{m_1^2}{m_2 - m_1^2}, \quad \hat{c} = \frac{m_1}{m_2 - m_1^2}. \tag{8}$$

In the usual notation, the asymptotic covariance matrix V of $(\hat{k}, \hat{c})$ is

$$V = \begin{bmatrix} \text{var}\,(\hat{k}) & \text{cov}\,(\hat{k}, \hat{c}) \\ \text{cov}\,(\hat{k}, \hat{c}) & \text{var}\,(\hat{c}) \end{bmatrix} \tag{9}$$

where

$$\text{var}\,(\hat{k}) = \frac{2k(k+1)}{N} \tag{10}$$

$$\text{var}\,(\hat{c}) = \frac{c^2(2k+3)}{NK} \tag{11}$$

$$\text{cov}\,(\hat{k}, \hat{c}) = \frac{2c(k+1)}{N}. \tag{12}$$

The Method of Maximum Likelihood

Let $T_1, T_2, T_3, \ldots T_n$ be a random sample of size n from the population with the probability density function (5). Then the likelihood function given by

$$L = \left(\frac{c^k}{\sqrt{k}}\right)^n \Pi\, t_i^{k-1}\, e^{-c\Sigma t_i} \tag{13}$$

Equating the partial derivatives of log L *w.r.t.* c and k to zero, we obtain the following equations for estimating c and k.

$$\frac{\partial \log L}{\partial c} = \frac{nk}{c} - \sum t_i = 0. \tag{14}$$

$$\frac{\partial \log L}{\partial k} = n \log c - n\,\frac{1}{\sqrt{k}}\,\frac{\partial \sqrt{k}}{\partial k} + \sum \log t_i = 0. \tag{15}$$

From (14) we obtain

$$\hat{c} = \frac{\hat{k}}{A} \tag{16}$$

where $A = \Sigma t_i/n$ Arithmatic mean of the observations. From (15) we obtain

$$\eta = y \tag{17}$$

where
$$\eta = \log k - \frac{1}{\sqrt{k}} \frac{\partial \sqrt{k}}{\partial k}$$

and
$$y = \log A - \log G$$
$$G = \text{Geometric mean}$$
$$= (\Pi t_i)^{1/n}.$$

Greenwood (1960) has tabulated the values of $\eta(k)$ against η. The value of k satisfying the equation (17) is obtained from the Tables of Greenwood (1960) by dividing the value of $\eta(k)$ corresponding to the observed value y by the observed value y. Using this value of k, the M. L. estimate of c is obtained from equation (16).

In the usual notation the asymptotic covariance matrix, I^{-1}, of the maximum likelihood estimates is given by (19) where

$$I = \begin{bmatrix} E\left(-\frac{\partial^2 \log L}{\partial k^2}\right) & E\left(-\frac{\partial^2 \log L}{\partial k \cdot \partial c}\right) \\ E\left(-\frac{\partial^2 \log L}{\partial k \cdot \partial c}\right) & E\left(-\frac{\partial^2 \log L}{\partial c^2}\right) \end{bmatrix} \tag{18}$$

$$I^{-1} = \begin{bmatrix} V(\hat{k}) & \text{cov}(\hat{k}, \hat{c}) \\ \text{cov}(\hat{k}, \hat{c}) & V(\hat{c}) \end{bmatrix} \tag{19}$$

where

$$V(\hat{k}) = \frac{1}{n\,p} \tag{20}$$

$$V(\hat{c}) = \frac{c^2}{\eta}\left(\frac{1}{k} + \frac{1}{k^2\,p}\right) \tag{21}$$

$$\text{cov}(\hat{k}, \hat{c}) = \frac{c}{knp} \tag{22}$$

$$p = \frac{d^2 \log \sqrt{k}}{\partial k^2} - \frac{1}{k}$$

$$\eta = \log k - \frac{d}{dk} \log \sqrt{k}$$

$$\frac{d\eta}{dk} = \eta' = \frac{1}{k} - \frac{d^2}{dk^2} \log \sqrt{k} = -p.$$

Using the large sample theory, estimates of $V(\hat{k})$, $V(\hat{c})$ and cov $(\hat{k}, \hat{c})$ are obtained by replacing k and c respectively, by $\hat{k}$ and $\hat{c}$ in equations (20), (21) and (22).

The value of $1/p$ i.e. $-(1/\eta')$ at $k = \hat{k}$ is obtained by equating $1/\eta'$ by the first difference approximation to $1/y'$ where

$$\frac{1}{y'} = y^{-1}\left[\frac{y_1 k_1 - y_0 k_0}{y_1 - y_1} - \hat{k}\right]$$

where $y_0 < y < y_1$ and $y_1 - y_0 = 0.01$ and y is the observed value of η at $k = \hat{k}$.

Application

The Data :

The continuous model has been fitted to two distributions relating to first conception taken from Anand data and Hutterite data (Sheps, 1967 : 129-132).

The Results :

For illustration the following calculations are given for the data in Table 1. The first two raw moments are $m_1 = 5.1692$ and $m_2 = 47.3384$. The moment estimates of k and c obtained with the help of the equation (8) are $\hat{k} = 1.2959$, $\hat{c} = 0.2507$. The expected frequencies based on these values of $\hat{k}$ and $\hat{c}$ are calculated and are given in column 3 of Table 1. Substituting estimates of $\hat{k}$ and $\hat{c}$ in the expressions (10), (11) and (12) we obtain the estimates

$$V(\hat{k}) = 0.0458, \qquad V(\hat{c}) = 0.0021$$
$$\text{cov}\,(\hat{k}, \hat{c}) = 0.0088 \quad \text{and} \quad r = 0.8979,$$

where r is the coefficient of correlation.

The value of maximum likelihood estimates $\hat{k}$ and $\hat{c}$ are obtained with the help of Tables of $\rho\,[\log \rho - (\sqrt{\rho'}/\sqrt{\rho})]$ given by Greenwood (1960). The values are given below. $\hat{k} = 1.4962$ and $\hat{c} = k/\bar{x} = 0.2894$. The expected frequencies in column 4 of Table 1 are obtained with the help of these values of $\hat{k}$ and $\hat{c}$. From equations (15), (16) and (17) we obtained the estimates

$$V(\hat{k}) = 0.0285, \qquad V(\hat{c}) = 0.0014$$
$$\text{cov}\,(\hat{k}, \hat{c}) = 0.0055, \quad \text{and} \quad r = 0.8594.$$

The same procedure has been followed in the other case (Table 2). The estimates by both methods, for the two sets of data, with the corresponding standard errors are given in Table 3.

TABLE 1—THE OBSERVED AND EXPECTED FREQUENCIES OF CONCEPTION BY MONTH, ANAND TOWN

Time from marriage to first conception (in months)	*Observed frequency*	*Expected frequency*	
		Method of moment $\hat{k}$ = 1.2959 c= .2507	*Method of Maximum likelihood* k-1.4962 $\hat{c}$= .2894
(1)	*(2)*	*(3)*	*(4)*
0-2	32	34.50	30.60
2-4	40	31.40	33.50
4-6	18	22.30	23.90
6-8	12	15.10	16.13
8-11	11	9.80	10.27
10-12	5	6.30	6.30
12-14	5	4.20	3.9o
14-16	3	2.40	2.40
18	2	1.70	1.40
22	1	1.70	1.40
24	1	0.60	0.20
Total	130	130.00	130.00
x^2 (4d.f)		4.6874	4.3686

TABLE 2—OBSERVED AND EXPECTED FREQUENCIES OF CONCEPTION BY MONTH, HUTTERITE DATA

Time from Marriage to first conception (in months)	*Observed frequency*	*Expected frequency* — *Method of Moments* aa=4.81, b = 14.30	based on a" = 4.81, $\bar{h}$= 16.51	based on 5) k = .5845, $\hat{c}$=.1350	*Method of maximum likelihood* $\hat{a}$a = 3,40, b=9.19	*based on(I)* a=3.26, $\hat{h}$=10.0	*based on\5)* k =.8678. $\hat{c}$=.2005
0-1	103	86.1	84.32	114.26	92.3	91.34	87.22
1-2	53	61.2	60.45	47.23	62.4	61.91	53.83
2-3	43	44.4	44.13	35.12	43.6	43.37	41.90
3-4	27	32.7	32.74	25.33	31.3	31.19	31.60
4-5	30	24,5	24.64	19.65	23.0	23.04	28.30
5-6	09	18.6	18.81	12.41	17.3	17.26	18.36
6-7	12	14.3	14.52	16.89	13.2	13.25	16.83
7-8	09	11.1	11.34	11.25	10.2	10.31	12.85
8-9	06	8.7	8.94	9.!6	8.0	8.13	9.23
9-10	OS	6,9	7.12	7.83	6,4	6.50	6.80
10-11	10	5.5	5.71	5.95	5.2	5.25	8.25
11-12	05	4.5	4.62	5.47	4.2	4.29	5.40
12-15	09	9.1	9.42	11.49	8.7	8.92	11.00
15-18	07	5.2	5,40	6.84	5.1	5.32	6.24
18-24	07	5.0	5.28	7.371	5.3	5.59	6.261
24-48	04	4.2	4.56	4.75	5.8	6.33	2.93
Total	342	3420	342.00	342.00	342.0	342.00	342.00
x^2		(13df)= 19.5	(13df) = 18.74	(12df) = 16.3913	(13df = 17.2	(!3df)- 15.92	(12df)= 15.90

a—Taken from Majumdar and Sheps (1970) for Type *I* **Geometric distribution.**

TABLE 3-ESTIMATES AND THEIR STANDARD ERRORS FOR THE CONTINUOUS MODEL

Estimate	*Method of moments*	*Method of maximum likelihood*
Anand Town	N= 130.	N= 130.
	c: 0.2507 ±0.0458	c: 0.2894 ±0.0374
	k: 1.2959±0.2I40	k: 1.4962±0.1688
	r: 0.8979	r: 0.8594
Hutterite **Data**	***N*= 342**	***N* = 342**
	c: 0.1350 ±0.0195	c: 0.2005 ±0.0171
	k: 0.5845 ± 0.0735	k: 0.8678± 0.0571
	r: 0.9037	r: 0.8888

From Tables 1 and 2 it is seen that the model derived in (5) **gives** a better fit.

References

1. Greenwood, J. Arthar, 1960, Aids for fitting the gamma distribution by Maximum likelihood, *Technometrics, 2,* 55-65.
2. Majumdar, H., and Sheps, M, C., 1970, Estimates of Type 1 geometric distribution from observations on conception times. *Demography, 1,* 349-360.
3. Potter, Robert G. and Parker, M.P., 1964, Predicting the time required to conceive, *Popula lation Studies,* **18,** 99-116.
4. Sheps, M. C., 1967. Uses of stochastic models in the evaluation of population policies I. *Theory aud Approaches to Data Analysis, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* 4, Berkeley, University of California Press.
5. Singh, S. N., 1961, Dissertation submitted for the degree of *Ph D.* at the University of California, Berkeley.
6. ___, 1964, On the time of first birth, *Sankhya,* 26 B, 95-102.
7. ____ 1967, On the first success. *The Journal of Scientific Research,* B.H.U., 18, 242-252.
8. ___and Bhaduri, T., 1972, Maximum likelikood estimates for the parameters of a continuous time model for first conception. *Demography, 9,* 249-255.