Prithwis Das Gupta

Estimation of the Parameters of a Censored Gamma Exponential Distribution and its Application to Data on Conception Delays

BRASS (1958) has proposed a model in which he assumes that the waiting time for a conception has a gamma exponential distribution. Singh and Bhaduri (1972) have estimated the parameters of this distribution by both the method of moments and the method of maximum likelihood. In a follow-up study of a group of women, there is always the possibility that some women will not become pregnant in the limited period of time they are observed. In the present note, we deal with this censoring problem in the gamma exponential distribution. We derive expressions for the estimates by the method of moments and the method of maximum likelihood, and these are applied to the longitudinal data collected in the Menstrual and Reproductive History Study at the University of Minnesota (Treloar *et al.*, 1967) on the number of months required to conceive for the first time (counting from marriage).

The Model

Let the probability that a woman will conceive between t and $t+\delta$ be $p\delta + 0(\delta)$. Therefore, her waiting time x for the first conception after marriage follows an exponential density

$$g^*(x\,;p) = pe^{-px}, \quad 0 < x < \infty; \quad p > 0. \tag{1}$$

Let us assume that p has a gamma density given by

$$h(p) = \frac{a^{\nu}}{\Gamma(\nu)} p^{\nu-1} e^{-ap}, \quad 0 < p < \infty; \quad a, \nu > 0, \tag{2}$$

so that the unconditional distribution of the waiting time for the first conception is

$$g(x) = \int_0^{\infty} g^*(x\,; p)\, h(p)\, dp = \nu a^{\nu}/(a+x)^{\nu+1}, \quad 0 < x < \infty; \quad a, \nu > 0. \tag{3}$$

Let N be the total number of women under observation, and let the censoring point be t (months, say). There are n women who become pregnant before this point, so that $(N-n)$ women should conceive after t months. Let $x_1, x_2, \ldots, x_n$ be the waiting times of the women who conceive in the first t months ($x_i \leqslant t$, $i = 1, 2, \ldots, n$). If g_t is the probability that a woman will conceive after t months, then

$$g_t = \int_t^{\infty} g(x)\, dx = a^{\nu}/(\alpha + t)^{\nu}. \tag{4}$$

Method of Maximum Likelihood

The likelihood function can be written as

$$\log L = \sum_{i=1}^{n} \log g(x_i) + (N-n) \log g_t, \tag{5}$$

from which it follows that

$$\frac{\partial \log L}{\partial a} = \nu \left[\frac{N}{a} - \sum_{i=1}^{n} \frac{1}{a+x_i} - \frac{N-n}{a+t} \right] - \sum_{i=1}^{n} \frac{1}{a+x_i} = 0, \tag{6}$$

and

$$\frac{\partial \log L}{\partial \nu} = \frac{n}{\nu} + N \log a - \sum_{i=1}^{n} \log (a + x_i) - (N-n) \log (a+t) = 0. \tag{7}$$

From (7), we obtain

$$\nu = n \Big/ \left[\sum_{i=1}^{n} \log (a + x_i) + (N - n) \log (a + t) - N \log a \right]. \qquad (8)$$

Substituting (8) into (6), we get an equation of the form $U(a) = 0$, which can easily be solved by iteration to obtain $\hat{a}$. We then find the estimate $\hat{\nu}$ by substituting $\hat{a}$ into (8). We note that when $t \to \infty$ — i.e., when the women are observed until all of them get pregnant—we have $n = N$, so that the above equations reduce to (3.3) and (3.4) in Singh and Bhaduri (1972).

Method of Moments

If we ignore the women who do not conceive in the first t months, the probability law corresponding to (3) should be written as

$$g^0(x) = g(x)/(1 - g_t), \quad 0 < x \leqslant t; \qquad a, \nu > 0. \qquad (9)$$

The method of moments equates the first two raw moments μ_1 and μ_2 of the distribution (9) to the corresponding observed moments m_1 and m_2 obtained from the data. Denoting $g_t/(1 - g_t)$ by H_t, it is possible to show that

$$\mu_1 = (a - H_t \nu t)/(\nu - 1), \ \nu \neq 1, \qquad (10)$$

and

$$\mu_2 = [2a^2 - H_t \nu t\{t(\nu - 1) + 2a\}]/[(\nu - 1)(\nu - 2)], \ \nu \neq 1, 2. \qquad (11)$$

An alternative expression for μ_1 is given in Brass (1958, p. 70), and for μ_2 in Das Gupta (1973, p. 108). When $t \to \infty$, so that $H_t \to 0$, eqs. (10) and (11) reduce to the moments of $g(x)$ in (3), and these are given in Singh and Bhaduri (1972, p. 251). Equating m_1 and m_2 to (10) and (11), we obtain

$$\begin{aligned} a &= [(\nu - 1)m_1 t - (\nu - 2)m_2]/(t - 2m_1), \\ \nu &= [2(m_2 - m_1^2) - H_t \nu t(t - 2m_1)]/(m_2 - 2m_1^2). \end{aligned} \qquad (12)$$

Eqs. (12) are of the form $a = U(\nu)$ and $\nu = V(a, \nu)$ respectively. They can be combined into a single equation in one variable, $\nu = W(\nu)$, from which the moment estimate $\hat{\nu}$, and hence $\hat{a}$, can be obtained by iteration. When $t \to \infty$, so that $H_t \to 0$, eqs. (12) reduce to those in (3.13) in Singh and Bhaduri (1972).

When the goodness of fit is judged in terms of not only the observed and expected frequencies up to the censoring point t but including also those at the open end, the solution of a and ν from the first moment (10) and the tail-end probability (4) should give better results. In other words, we solve a and ν from

$$m_1 = (a - H_t \nu t)/(\nu - 1), \qquad \nu \neq 1, \tag{13}$$

and

$$(N - n)/n = g_t. \tag{14}$$

Substituting (14) in (13), we obtain

$$\nu = (m_1 + a)/[m_1 + t(N - n)/n]. \tag{15}$$

Eq. (15) can now be substituted in (14) to form an equation $U(a) = 0$, which can be solved by iteration.

Application to Data

The Menstrual and Reproductive History Study of the University of Minnesota Health Service provides us with the records of women's waiting times in terms of the number of months between marriage and first conception. The censoring point we have considered is 50 ($= t$) months, and the observed frequencies of conception by number of months are shown for 1,127 women in the first two columns of Table 1.

The last two columns in Table 1 give the expected frequencies as obtained from the maximum likelihood method and the method of moments respectively. Both the methods yield values of χ^2 that are nonsignificant at the 1 per cent level. Nevertheless, the fits are not excellent, due largely to an inflated observed frequency of conception in the first month. A relatively high frequency in the first month has also been observed in other studies (see Das Gupta and Hickman, 1974). The complications in the first month thus make it difficult to provide a satisfactory model for the waiting time. However, the objective of this note is not so much to establish the suitability of the gamma exponential model for the waiting time data as to indicate the method of handling the problem of censoring inherent in most data (in the form of an open-ended interval) when this model is used.

References

1. Brass, W., 1958, The distribution of births in human populations, *Population Studies,* **12,** 51-72.
2. Das Gupta, P., 1973, *A Stochastic Model of Human Reproduction,* Population Monograph Series 11, Institute of International Studies, University of California, Berkeley (reprinted by Greenwood Press, Westport, Connecticut, in 1976).
3. ____ and Hickman, L., 1974, Estimation of the parameters of a type I geometric distribution from truncated observations on conception delays, *Mathematical Biosciences,* **22** (Winter), 75-94,
4. Singh, S. N. and Bhaduri, T., 1972, Maximum likelihood estimates for the parameters of a continuous time model for first conception, *Demography,* 9, 249-255.
5. Treloar, A. E. Boynton, R. E., Behn, B.G. and Brown, B. W., 1967, Variation of the human menstrual cycle through reproductive life, *International Journal of Fertility, 12,* 77-126.

Appendix

TABLE—1 OBSERVED AND EXPECTED FREQUENCIES OF CONCEPTIONS BY MONTHS AND THE ESTIMATED PARAMETERS; MRHS DATA, UNIVERSITY OF MINNESOTA

Waiting Time (months)	*Obs. Freqs.*	*Exp. Freqs.*		*Waiting Time (months)*	*Obs. Freqs.*	*Exp. Freqs.*	
		Max. L. Method	*Moment[b] Method*			*Max. L. Method*	*Moment Method*
0-1	73	54	52	15-20	99	87	89
1-2	52	50	48	20-25	61	68	70
2-3	51	46	44	25-30	74	54	56
3-4	40	42	41	30-35	39	44	46
4-5	32	39	38	35-40	24	36	38
5-6	32	37	36	40-45	37	31	32
6-7	22	34	34	45-50	32	26	27
7-8	25	32	31	$>50\ (=t)$	270	273	270
8-9	24	30	30				
9-10	24	28	28	Total	1,127	1,127	1,127
10-11	28	26	26				
11-12	22	24	25	χ^2 (20 d.f.)[a]	—	35.32	35.95
12-13	20	23	23	$\hat{a}$	—	26.87	31.16
13-14	26	22	22	$\hat{v}$	—	1.349	1.493
14-15	20	21	21				

a. The value at the 1 per cent level is 37.6.
b. From first moment and tail-end probability.