On the Most Plausible Value of Gestation Period: An Application of Stochastic Model

Brijesh P. Singh¹, Kushagra Gupta² & Kaushalendra K. Singh³

¹Assistant Professor (Statistics), Faculty of Commerce & DST-CIMS, Banaras Hindu University, Varanasi, U.P., India.
²Biostatistician, Cytel Statistical Software and Services Pvt. Ltd., Pune
³Professor, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, U.P., India.

Abstract

In human reproduction process, analyzing the birth interval is a crucial and important aspect. Demographers show their keen interest in the study of birth intervals even in the modern era when age at marriage and education level is has increased up to a certain level which is quite satisfactory. For estimation of fertility parameters i.e. fecundability, researchers have shown keen interest in the interval between marriage and first conception since sixties under various set of assumptions considering discrete as well as continuous probability distributions. Since the data of conception cannot be observed directly, thus, demographers used the data of length of first birth and subtracted a fixed value of gestation period i.e. 9 months from the duration of first birth interval. Biswas (1972) have proposed a linear model with non linear variants in his study to estimate the gestation duration and has shown that the most likely value of duration of gestation is 40 weeks (≈280 days). In this paper an attempt has been made to estimate the gestation duration by adding a parameter for gestation period in the model proposed by Singh (1964). In this study 9.4 months gestation period is obtained as the most probable value, which is almost 282 days that is supported by earlier study done by Biswas (1972).

Keywords: Gestation duration estimation, Reproduction process, Pareto distribution, Simulation study, Maximum Likelihood estimates, Stochastic model.
1. INTRODUCTION

First birth interval plays a significant role in determination of fertility level of the society because the length of first birth interval can be considered as start of parenthood, i.e. couple start their reproductive liability. First birth interval also has important for analyses of fertility because first birth interval is free from post partum amenorrhea (PPA) and the females usually do not use any type of contraception before delivering first birth. Several authors have proposed stochastic models relating to first live birth interval under different sets of assumption. These models have been developed for two different purposes. Firstly, models have been used to assist in interpreting observed data using the parameters of the model fitted to the data. Secondly, models have been proposed for the reproductive process to study the interaction of underlying variables.

For the study of parameter of fertility, stochastic models are further categorized into two parts (i) discrete time model and (ii) continuous time model. In discrete time models, the waiting time for first conception has been described by geometric distribution (Gini, 1924; Henry, 1961) by considering the assumption that fecundability is constant till the woman conceives. Further, Henry (1953) and Vincent (1961) used with some modifications to study the natural fertility. Potter and Parker (1964) and Singh (1961, 1967) generalized the geometric distribution by incorporating presence of foetal losses before the first live births. The demographers are still working on these models to achieve their goals regarding inclusions of social parameters. Singh et-al. (2012) explained the social parameters which changes over time such as number of months required for conception.

In continuous time models, Singh (1964) proposed an exponential and modified exponential time model by considering that fecundability varies from female to female for the same situation. These models are based on the common assumption that all the females under study are susceptible to conception at the time of marriage. This interval is greatly influenced by the social customs prevalent in different societies as well as certain physiological constraints such as females visit their parents for considerable periods of time even after marriage which introduces an inoperative period (termed as period of temporary separation in the time of first conception) (Saxena, 1969; Yadava, 1971, Bhattachaya et al., 1988). Chakraborty (1976) proposed a model for the time of first birth assuming that only a proportion of females are exposed to the risk of conception at marriage. On the other hand, Pathak and Prasad (1977) developed a continuous time model for the first conceptive delay taking the period of adolescent sterility as a random variable which follows an exponential distribution. Further, Pathak (1978) derived a discrete time distribution for the time of first conception with provision for adolescent sterility. The assumption in the above models that the component of adolescent sterility continues to play indefinitely was later modified by Nair (1983). Mishra (1982, 1984) suggested a truncated model for
first birth interval assuming that only a proportion of females are exposed to the risk of conception at marriage.

The data of conception cannot be observed directly and due to this fact demographers used the data on first birth and subtracted a fixed value of gestation period i.e. 9 months from the duration of first birth. This data is further used to get the distribution of pattern of first conception utilizing the discrete or continuous time model discussed above. Biswas (1972), proposed a linear model with non linear variants in his study to estimate the gestation duration. But all this type of study faces the problem of getting the reliable data on conception. In this paper an attempt has been made to estimate the gestation duration by adding a parameter in the model proposed by Singh (1964) for the same data set. In this study an attempt has been made to obtain the maximum likelihood (ML) estimators of scale and shape parameters of the model for the appropriate value of location parameter.

2. 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 (1964) is derived under the following assumptions:

a) The number of coition during any time interval $(0, t)$ of length $t$ is a random variable which follows the Poisson distribution with parameter $\lambda t$; where $\lambda t$ 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 $\lambda t = p_1 p_2 \lambda t$ in time interval $(0, t)$, if conceptions are assumed to be instantaneous, i.e., the related periods of temporary sterility are zero. In the case of a first birth it can be assumed that a conception not resulting in a live birth is instantaneous. Under these conditions the waiting time for first conception leading to a live birth follows an exponential distribution.

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0, \quad \lambda > 0 \quad (2.1)$$

In this connection, it should be noted that the above assumptions, which lead to mathematical simplicity, are strong and many details including abortions and miscarriages which make up a significant proportion of the total pregnancy history are ignored. Singh also assumed that $\lambda$, which may be defined as average conception rate,
follows a Pearson Type III distribution with parameters $a$ and $b$. Under these assumptions, Singh (1964) obtained $g(t, a, b)$, the probability density function of $T$, as follows:

$$f(t \mid a, b) = \frac{ab^a}{(b + t)^{a+1}}, t > 0, a > 0, b > 0$$ (2.2)

This distribution a kind of Pareto distribution, where $b$ is scale parameter and $a$ is shape parameter.

3. PROPOSED MODEL

Generally, we have the data on first births, and if we need a distribution of first conception generally we subtract nine months from the time duration of first birth which seems to be illogical in some cases. Thus, a displaced exponential distribution for the time of first birth by taking ‘$g$’ as gestation duration has been proposed and which can be written as:

$$f^*(t) = \lambda \exp(-\lambda (t - g)); \quad t > g, g > 0, \lambda > 0$$ (3.1)

where $\lambda$ is conception rate and $g$ is gestation duration. Under the same assumption that $\lambda$ follows Pearson Type III distribution with parameters $a$ and $b$.

$$f(\lambda) = \frac{b^a \lambda^{a-1} \exp(-\beta \lambda)}{\Gamma a} ; \quad \lambda > 0, a, b > 0$$ (3.2)

The probability density function of $t$ can be obtained as:

$$f^*(t \mid a, b, g) = \frac{b^a}{\Gamma a} \int_0^\infty \lambda^a \exp(-\beta \lambda) \exp(-\lambda (t - g)) d\lambda$$ (3.3)

After simplification we get;

$$f^*(t \mid a, b, g) = \frac{ab^a}{(t - g + b)^{a+1}}, \quad t > g, a > 0, b > 0, g > 0$$

or

$$f^*(t \mid a, b, g) = \left(\frac{a}{b} \right)^{\alpha} \left(1 + \frac{t - g}{b}\right)^{-\alpha}, \quad t > g, a > 0, b > 0, g > 0$$ (3.4)

The above closed form of the distribution is known as Pareto Type II distribution. Here $g$ is the location parameter, $b$ is scale parameter and $a$ is shape parameter.
4. ESTIMATION PROCEDURES

The estimation of the parameters, there are several methods like moment method, ML estimation, etc. The parameters \( a \) and \( b \) are estimated with the help of BAN method earlier in Singh (1964). In the current analysis we have utilized the previous estimates to obtain the estimates of \( g \) by ML method and it is minimum order statistics which changes from sample to sample. For simplicity, analytical procedure has been applied and by initial guess value of \( g = 9 \) months have been considered in fitting the data on first conception of females. For the better and close fitting the value of \( g \) is taken below 9 as well as above 9 to get the optimum or best fit in this region for the same data of first conception used in Singh (1964). The ML method is described below:

Maximum likelihood estimation begins with writing a mathematical expression known as the Likelihood Function of the sample data. Loosely speaking, the likelihood of a set of data is the probability of obtaining that particular set of data, given the chosen probability distribution model. This expression contains the unknown parameters of the model. The values of the parameters which maximize the sample likelihood are known as the Maximum Likelihood Estimates or MLE’s.

The likelihood function for the Pareto distribution with parameters \( a \) and \( b \) for the sample variable \( T=\{t_1,t_2,t_3,\ldots,t_n\} \) is given by:

\[
L(a,b) = \prod_{i=1}^{n} \frac{ab^a}{(t_i - g + b)^{a+1}}; \quad t > g, a > 0, b > 0, g > 0
\]  

(4.1)

\[
\ln(L) = n \ln(a) + n a \ln(b) - (a+1) \sum_{i=1}^{n} \ln(t_i - g + b); \quad t > g, a > 0, b > 0, g > 0
\]  

(4.2)

Partially differentiating with respect to \( a \) is given by:

\[
\frac{\delta \ln(L)}{\delta a} = \frac{n}{a} + n \ln(b) - \sum_{i=1}^{n} \ln(t_i - g + b)
\]  

(4.3)

Similarly, partially differentiating with respect to \( b \) is given by

\[
\frac{\delta \ln(L)}{\delta b} = \frac{na}{b} - (a+1) \sum_{i=1}^{n} \frac{1}{(t_i - g + b)}
\]  

(4.4)

Equating the above two equations with zero to maximize the \( \ln(L) \) one can obtain the estimates of \( a \) and \( b \). For maximizing the \( \ln(L) \), equating the above to zero the value
of \( a \) can be estimated. The estimate of \( b \) is obtained by lower bound and maximizing \( L \). Thus, we have

\[
\hat{b} = \min_i (t_i - g)
\]

and

\[
\hat{a} = \frac{n}{\sum_{i=1}^{n} \ln(t_i - g + \hat{b}) - n \ln(\hat{b})}
\]

\( (4.5) \)

\( (4.6) \)

5. DATA AND APPLICATION

The data on first birth has been taken from the Demographic Survey of Banaras Tehsil, 1956 (Singh, 1956). In the present study we take the data set which includes the couple with age of female 15 years at return marriage and their marital duration is 15 years and more. This data has been utilized to get the estimates of parameters at different gestation duration and fitting of the distribution and further used for validation of the gestation duration.

6. RESULTS AND DISCUSSION

Table 1 is the simulation study based on the 100 random samples of size 1000 each. It gives the mean, variance, minimum, maximum and coefficient of variation of ML estimates of scale and shape parameters for the different prefixed values of \( g \). In this simulation study, initial guess for scale and shape parameters of the model have been taken from the estimates obtained earlier by Singh (1964). For the estimates of scale, highest mean value is obtained at \( g = 9.6 \) months and lowest average is obtained for \( g = 9.4 \) months. Variability of the estimates of scale parameter is lowest for \( g = 9.4 \) months and highest for \( g = 8.8 \) months. The mean value of estimates of shape parameter is highest at \( g = 9.2 \) months and lowest at \( g = 9.4 \) months and variance of shape parameter is lowest at \( g = 9.4 \) months and highest at \( g = 8.6 \) months. The coefficient of variation provides more precise results for the estimates of scale and shape parameters at different gestation duration. The table clearly shows that maximum coefficient of variation (C.V.) for scale and shape parameter is obtained at \( g = 8.6 \) months and minimum C at \( g = 9.4 \) months. This explains that the gestation duration at 9.4 months is most appropriate among the other gestation durations. Table 1 also shows that the coefficient of variation is going high again as \( g \) increases above \( g = 9.4 \) and below \( g = 8.8 \). Thus, here we have considered the variation close to 9 months as generally used in this type of previous analysis. Here for validation of the estimates of \( g \), the value of \( g = \{8.6, 8.8, 9.0, 9.2, 9.4\} \) has been considered for fitting of the data.

Table 1: Summary of Parameters of the Proposed Model using Simulation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>At g (months)</th>
<th>Mean Value</th>
<th>Variance</th>
<th>Min</th>
<th>Max</th>
<th>C.V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>8.6</td>
<td>11.2982</td>
<td>3.4266</td>
<td>8.0983</td>
<td>16.6511</td>
<td>16.3839</td>
</tr>
<tr>
<td>Shape</td>
<td>8.8</td>
<td>4.6461</td>
<td>0.4275</td>
<td>3.4317</td>
<td>6.5651</td>
<td>14.0723</td>
</tr>
<tr>
<td>Scale</td>
<td>9</td>
<td>11.5019</td>
<td>3.5093</td>
<td>7.8113</td>
<td>17.4808</td>
<td>16.287</td>
</tr>
<tr>
<td>Shape</td>
<td>9.2</td>
<td>4.7277</td>
<td>0.4105</td>
<td>3.4949</td>
<td>6.7575</td>
<td>13.5517</td>
</tr>
<tr>
<td>Scale</td>
<td>9.4</td>
<td>4.7797</td>
<td>0.3707</td>
<td>3.6165</td>
<td>6.5831</td>
<td>12.7387</td>
</tr>
<tr>
<td>Shape</td>
<td>9.6</td>
<td>4.7721</td>
<td>0.338</td>
<td>3.6165</td>
<td>6.5831</td>
<td>12.1834</td>
</tr>
</tbody>
</table>

Table 2: Fitting of the Distribution for different values of Gestation periods of females with return age at marriage 15 years

<table>
<thead>
<tr>
<th>Time in Months for First Birth</th>
<th>Observed</th>
<th>EST at g=8.6</th>
<th>EST at g=8.8</th>
<th>EST at g=9.0</th>
<th>EST at g=9.2</th>
<th>EST at g=9.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>g-24</td>
<td>47</td>
<td>35.6</td>
<td>18.55</td>
<td>77.5</td>
<td>64.9</td>
<td>51.13</td>
</tr>
<tr>
<td>24-48</td>
<td>96</td>
<td>105.59</td>
<td>115.25</td>
<td>81.1</td>
<td>88.58</td>
<td>96.54</td>
</tr>
<tr>
<td>48-72</td>
<td>50</td>
<td>45.85</td>
<td>49.68</td>
<td>36.58</td>
<td>39.45</td>
<td>42.32</td>
</tr>
<tr>
<td>72-96</td>
<td>17</td>
<td>22.19</td>
<td>23.89</td>
<td>18.21</td>
<td>19.43</td>
<td>20.66</td>
</tr>
<tr>
<td>96-120</td>
<td>11</td>
<td>11.67</td>
<td>12.49</td>
<td>9.79</td>
<td>10.34</td>
<td>10.96</td>
</tr>
<tr>
<td>120-144</td>
<td>5</td>
<td>6.55</td>
<td>6.98</td>
<td>5.6</td>
<td>5.86</td>
<td>6.2</td>
</tr>
<tr>
<td>144-168</td>
<td>5</td>
<td>3.88</td>
<td>4.12</td>
<td>3.37</td>
<td>3.49</td>
<td>3.7</td>
</tr>
<tr>
<td>168+</td>
<td>8</td>
<td>7.66</td>
<td>8.04</td>
<td>6.85</td>
<td>6.95</td>
<td>7.48</td>
</tr>
<tr>
<td>Total</td>
<td>239</td>
<td>239</td>
<td>239</td>
<td>239</td>
<td>239</td>
<td>239</td>
</tr>
</tbody>
</table>

| Chi Square Value              | 6.84849  | 49.7512      | 20.9449      | 9.65953      | 3.10206      |
| p-value for Chi Square        | 0.2322   | <0.0001      | 0.0008       | 0.0855       | 0.6843       |
Table 2 contains the data on time of first birth with females return marriage of 15 years. First two columns contain month wise intervals and corresponding number of couples having births within that interval termed as observed. The other columns are the fitted frequencies at different value of the gestation duration based on the corresponding parameters of scale and shape shown in Table 1. The Table also contains the chi-square values at different gestation durations. From the Table it is clear that at $g=9.4$ the value of chi-square is minimum. Thus, $g=9.4$ months should be taken as the most probable value of gestation duration.

**Figure 1:** Observed and estimated curve for first birth interval at different values of gestation duration

Figure 1 represents the fitting of the distribution. The Figure shows the observed as well as the expected values considering the model of the data on time of first birth with females return marriage of 15 years. The observed and expected values are quite close at $g=9.4$ months and highest departure of expected value from the observed value has been obtained at $g=8.8$ months.

**7. CONCLUSION**

The current study is an attempt to estimate an appropriate duration of gestation using distribution of conception by taking gestation duration as a random variant. The

proposed distribution has three parameters which are estimated by ML method. To know the appropriate measure of gestations duration, a simulation based study has been done. The fitting of the distribution indicates that the most likely estimates of $g$ should be 9.4 months ($\approx 282$ days). In an earlier study which is done by Biswas (1972) has also shown that the most likely value of duration of gestation is 40 weeks ($\approx 280$ days). This duration is also suggested by biologists as well as medical practitioners. Thus, the above suggested value of gestation period may be used for further analysis of fertility pattern and behavior.

REFERENCES:


