Survival Modeling of First Birth Interval After Marriage

Amusan Ajitoni Simeon, Zarina Mohd Khalid.

Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru Malaysia
asamusan2@live.utm.my

Abstract: This is a data exploratory analysis of retrospective cross-sectional study of pattern of first birth interval after marriage in Nigeria. The data for the study are extracted from the published reports of the National Demographic and Health Survey 2009 edition. Fertility as a major component of population change is influenced by first birth interval after marriage, since the interval is positively correlated with the cumulative number of children a woman would have at the end of her reproductive life. Studies have described this interval using non-parametric methods which lack features to project the estimates further. This paper is designed to fill the gap by attempting to fit a parametric model to data on the first birth interval among women of reproductive age in Nigeria. Four parametric models whose various curves and estimates are compared with non-parametric values are considered, namely Inverse Gaussian, Log-logistic, Weibull and Burr Type XII. The best model appears to be Inverse Gaussian based on the Akaike Information Criterion of lowest value of 116617.6. Quantile-quantile plots also identify Inverse Gaussian as model whose data points clustered much around straight line. All other curves give credence to the Inverse Gaussian model as the one that describes the data better than the rest. However, the study does not rule out adaptability of Log-logistic distribution to model waiting time to first birth after marriage since it also behaves similarly to Inverse Gaussian distribution.


Keywords: Marriage, Women, First Birth Interval, Non-parametric, Parametric, Models, Estimates

1. Introduction

The length of waiting time to first birth after marriage of a woman is of major interest to population experts, since the interval is positively associated with the cumulative number of children a woman would have at the end of her reproductive life. In the same vein, postponement of first birth after marriage as conceptualized by Timaus (Moultrie, Say, & Timæus, 2012) also suggests the importance of this waiting time for its implication on the cumulative fertility. This waiting time to first birth after marriage is sometimes referred to ‘first birth interval’. Short first birth interval after marriage of a woman leads to rapid transition to higher parity and consequently to high fertility, particularly when the first birth is a female child (Global, 2007; Mace & Sear, 1997; Moultrie et al., 2012; Yohannes, Wondafrash, Abera, & Girma, 2011)

Among several indicators to measure fertility pattern in human population, FBI after marriage is important. This is because data on FBI are not affected by recall biases, being the first event of marriage life of a woman. The data are also free from erratic fluctuations of safe fertility period of breastfeeding and post-partum amenorrhea resulting from previous life births, which are associated with data from other birth intervals (Amin & Bajracharya, 2011; Lloyd, 2004).

Fertility which is recognized by demographers as a major component of population dynamics remains high in Africa by the standards of the rest of the world. However, the belated fertility transition has been observed to be underway in the region with remarkable progress in countries like South Africa, Botswana, and Zimbabwe (Bongaarts, 2008; Bremner et al., 2010; Garenne, Tollman, Kahn, Collins, & Ngwenya, 2001; Moultrie et al., 2012). Birth intervals have been noticed to be responsible for the decline in TFR in these countries. For example, in South Africa TFR has rapidly declined to 2.3 births per woman which is a significant drop from 6.5 in the 1960s as a result of birth postponements, an attitudinal change which is largely independent of ages and parities of the women (Bremner et al., 2010; Moultrie et al., 2012).

Many countries in Asia were able to reduce their fertility through government policies. For instance, China and Vietnam have witnessed declines in their TFR due to stringent government policies that discourage early and arranged marriage and at the same time encourage birth postponement at any parity of the women (Banister, 1987; Löfstedt, Ghilagaber, Shusheng, & Johansson, 2005).

Nigerian population remains the highest in Africa, placing tenth in the world despite current fertility decline in Africa. The current estimate put the figure at over 170 million and the current total
fertility rate of 5.7 births per woman compared to the overall total fertility rate (TFR) of 5.2 births in Africa, is considered relatively high (Commission, 2009).

The importance of efforts to achieve significant fertility decline in Nigeria cannot be overemphasized, as it has been observed that poor socioeconomic development in the sub-region has been attributed to high fertility and as such, all hands must be on deck to understand the phenomena that could accelerate the pace of fertility decline in the country. However, no meaningful success can be achieved along this line without understanding the characteristics of first birth interval which has not been researched into exhaustively unlike other birth intervals. Even though, there have been several studies on First Birth Interval (FBI) after marriage by different authors for various countries employing various statistical techniques to describe it and its determinants, nevertheless its distribution still remains unexplored. For example, a stochastic model was developed by Singh (Pathak, Singh, & Singh, 2006). In this model, assumptions of pregnancy loss due to pregnancy related complications and condition of cultural practice of physical separation of spouses for a certain period of time in India were imposed.

The assumption of physical separation is at variant to the traditional cultures in Africa. In fact, traditional African cultures place great values on childbearing and parenthood immediately after marriage (Commission, 2009; Feyisetan & Bankole, 2002), and marriage is believed to be onset of exposure of the woman to the risk of becoming pregnant and subsequently should deliver a baby. Kazembe in Kenya (Kazembe, 2009) proposed a sequential ordinal model, a semi parametric model using data from Kenya 2008 DHS. Here, the intervals were grouped into categories to reduce it to sequential ordinals and subsequently employed a logistic regression to evaluate the influence of covariates on the interval.

The work of (Logubayom & Luguterah, 2013) also described FBI with data from 2008 DHS of Ghana using non-parametric and semi parametric approaches, Kaplan Meier and Cox’s Proportional Model respectively to do this. The study estimated the median FBI after marriage in Ghana to 30 months. In a similar vein in Huanning county China, the median FBI which was around 34 months in the 1950s which reduced to less than 15 months in the 1980s (Löfstedt et al., 2005) as a result of government policy. Ordinary linear regression model of FBI data was fitted by (Amin & Bajracharya, 2011) to examine the impacts of some covariates on the interval. (Keiding, Hojbjer, Sørensen, & Slama, 2012) employ Current duration approach to estimate the time to pregnancy by a married woman.

In the light of the above, many of the techniques employed to describe First Birth Interval were distribution free whose estimates cannot be projected into the future. It appears there is paucity of parametric approach to examine this interval. This study is designed to fill the gap. There is the need to identify the distribution of this interval in order to describe it effectively with the characteristics of the distribution.

This study aims at carrying out data exploration analysis on the First Birth Interval after marriage among Nigerian women, in order to identify a distribution that best describes the interval.

The rest of the paper is organized in the following order; 2.-methods, 3.-results, 4.-discussion and conclusion.

2. Material and Methods

This is data exploratory analysis on retrospective cross-sectional study of pattern of FBI after marriage in Nigeria. The data for the study were extracted from the published reports of the National Demographic and Health Survey 2009 edition. The survey was carried out in June 2008 by ICF Macro Calverton, Maryland in collaboration with the Nigeria National Population Commission (Commission, 2009)

Through a multi-stage probability sampling procedure a total 33,380 representative samples of national population of women of reproductive age were selected for interview in the survey. T is the waiting time or survival time to first birth not less than 7 months and not more than 120 months into marriage 15,363 respondents who went into marriage for the first time without a child and whose records were complete of information needed form the sample for this paper, out of which, 1296 survival time were rightly censored.

This paper employs non-parametric and parametric methods to assess the data on the waiting time to first birth after marriage (FBI). Generally, the distribution of survival time to first birth “T” is described by the survival function, cumulative distribution function and hazard function. Here, T is continuous and non-negative. The cumulative distribution function of T (cdf) with the probability density function (pdf) is given as

\[ F(t) = P(T \leq t) = \int_0^t f(x) \, dx \]

Where \( F(t) \) is the probability that a woman has birth on or before time t. That is, cumulative distributional function.

Probability that a woman gives birth after time t is given by the survival function as follows;
where $z = \frac{\log(t) - \mu}{\sigma}$; $\mu = \log\text{mean}$ and $\sigma = \log\text{scale parameter}$

Inverse Gaussian’s pdf

$$f(t|\mu, \sigma) = \frac{\lambda}{2\pi\sigma^3} \exp\left\{\frac{\lambda}{2\sigma^2} (t - \mu)^2\right\}$$


Inverse Gaussian is also known as Wald distribution. It is popular for modeling non-negative positively skewed data. $\mu$ is the scale parameter while $\lambda$ is the shape parameter

The Burr Type XII distribution has the pdf as thus;

$$f(t|\alpha, \beta, k) = \frac{k}{\beta} \left(\frac{t}{\beta}\right)^{k-1} \left(1 + \left(\frac{t}{\beta}\right)^k\right)^{-\alpha - 1}$$

The distribution is a 3-parameter family distribution, where $\alpha$ is the scale parameter, $\beta$ is the first shape parameter and $k$ is second shape parameter.

The Maximum Likelihood Estimates of these distributions are obtained through matlab distribution fitting tools. The Akaike Information Criterion (AIC) provides a way to select a model from a set of models. It is based on information theory which seeks a model that has a good fit to the data with a few parameters. The formula is given as $AIC = -2(\log\text{likelihood}) + 2k$, (Lawless, 2011; Posada & Crandall, 2001), where $k$ is the number of free parameters in the model and likelihood is the probability of the data given a model. The lower is the value of AIC the better is the model. Quantile-quantile plots examine the distributions to identify the plot whose points cluster much around the straight line, which is adjudged better than the one whose points scatter away from the straight line

3. Results

There are a total of 15,363 data points, out of which 1296 (8.4%) are rightly censored. Table 1 shows the table of the non-parametric estimates, through the Kaplan-Meier (Product-Limit) method, of the mean and median survival time to first birth after marriage. Their respective standard errors and confidence intervals are also displayed in the table 1 below. The mean and median waiting times to first birth after marriage by women in Nigeria are respectively given as 28.8 and 20.0 months by the analysis.
Table 1: Product- Limit Estimates of Means and Medians for Survival Time

<table>
<thead>
<tr>
<th></th>
<th>Mean Estimate</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
<th>Median Estimate</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>28.832</td>
<td>0.1990</td>
<td>28.4420</td>
<td>29.2230</td>
<td>20.0000</td>
<td>0.1700</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>20.0000</td>
<td>0.1700</td>
<td>19.6680</td>
<td>20.3320</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As displayed below, Table 2 presents the estimates, including their standard errors and confidence intervals, of the parameters of the models under investigation. All the estimated parameters lie within 95% confidence interval, an indication that they are significant (p<0.05). The goodness of fit for selection of the best model criteria as given by the Akaike Information Criterion (AIC) statistic is presented at the last column of this table, with Weibull having the largest value of 10,739.8, while Inverse Gaussian model has 116617.2, the least.

Table 2: Parameter Estimates of Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
<th>Log-likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>$\lambda$ =31.6268</td>
<td>0.20039</td>
<td>31.2340</td>
<td>32.2004</td>
<td>-60367.9</td>
</tr>
<tr>
<td></td>
<td>$\alpha$ =1.38753</td>
<td>0.0084</td>
<td>1.3710</td>
<td>1.4041</td>
<td></td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>$\mu$ =28.7405</td>
<td>0.1942</td>
<td>28.359</td>
<td>29.1211</td>
<td>116,617.2</td>
</tr>
<tr>
<td></td>
<td>$\lambda$ =45.4598</td>
<td>0.5319</td>
<td>44.4174</td>
<td>46.5022</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>$\mu$ =3.0447</td>
<td>0.0061</td>
<td>3.0329</td>
<td>3.0566</td>
<td>-58968.8</td>
</tr>
<tr>
<td></td>
<td>$\sigma$ =0.4176</td>
<td>0.0029</td>
<td>0.4119</td>
<td>0.4232</td>
<td></td>
</tr>
<tr>
<td>Burr</td>
<td>$\alpha$ =10.2434</td>
<td>0.1229</td>
<td>10.0026</td>
<td>10.4842</td>
<td>-58342.2</td>
</tr>
<tr>
<td></td>
<td>$\nu$ =5.8071</td>
<td>0.1787</td>
<td>5.4569</td>
<td>6.1573</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\kappa$ =0.2074</td>
<td>0.0086</td>
<td>0.1905</td>
<td>0.2244</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Mean Survival Time To First Birth For The 4 Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Standard Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>28.8667</td>
<td>0.1541</td>
<td>28.5646</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>28.7405</td>
<td>0.1942</td>
<td>28.359</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>28.5042</td>
<td>0.2964</td>
<td>27.9232</td>
</tr>
<tr>
<td>Burr Type XII</td>
<td>59.8018</td>
<td>11.4223</td>
<td>262.4749</td>
</tr>
</tbody>
</table>

Table 3 above shows the table of the estimated mean survival time to first birth with their respective standard errors and confidence intervals using the estimated parameters of the distributions under investigation. All the means estimated are in the neighborhood of 28.8 months, the value gotten from the non-parametric method, except for the Burr type XII distribution whose estimated mean is as large as 59 months with a large standard error of 11.4 months.

Table 4: Median Survival Time To First Birth For The 4 Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Median</th>
<th>Standard Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>24.2848</td>
<td>0.1927</td>
<td>23.9071</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>22.0032</td>
<td>0.1738</td>
<td>21.6625</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>21.0041</td>
<td>0.1259</td>
<td>20.7573</td>
</tr>
<tr>
<td>Burr</td>
<td>18.0994</td>
<td>0.4672</td>
<td>19.3898</td>
</tr>
</tbody>
</table>
As shown in the following, figures 1a and 1b depict the shapes of the Kaplan-Meier survival curve and the cumulative hazard curve respectively for the time to first birth after marriage data in Nigeria. The Kaplan-Meier curve is a right continuous step function which steps down only at the uncensored observations.

Probability density curves of all the distributions under investigation are mounted over the histogram plot of the data on ‘first birth interval’ after marriage are displayed in figure 2. This is visualizing how perfect each of the distributions describes the data.

The Non-Parametric curve which is in red color conforms to the shape of the histogram plot. The Inverse Gaussian is in blue, while log-logistic is purple color. These two distributions conformity to the data is similar. However, Weibull is in pink color, with the lowest peak. Furthermore, it is observed that Burr Type XII distribution in black color has the highest peak and covers the highest peak of the histogram but fails to cover part of the data as the waiting time to first birth becomes large.

In a similar vein, figure 3 displays the survival curves of all the distributions overlaid on the Kaplan-Meier curve. The Non-Parametric curve, in red, consistently lies over the K-M curve, while other distributions take their paths around the curve showing their degree of closeness to the Kaplan-Meier curve.
Probability plot is a graphical method for determining whether sample data conform to a distribution. Figure 4 shows the probability plots of all the distributions with Burr’s plot in black color having a wider deviation away from the probability plot particularly at the end, while Weibull deviates upwards as shown in pink color. However, Inverse Gaussian plot in blue color consistently lies on the data points plot with a little deviation at the end. The figure also clearly shows the closeness of Log-logistic plot in purple color to that of Inverse Gaussian in relation to their deviation at the end.

The cumulative probability curves for all the distributions including that of non-parametric in red color are shown in Figure 5. All the curves except for the Burr’s are asymptotic to unity at the end as failure time becomes increasingly large.

Figure 6 displays the cumulative hazard curves for all the distributions and for the Non-parametric in red color. The curves show that Burr and Weibull in black and pink colors respectively do not closely follow the paths described by the data and by the Non-Parametric curve. However, Inverse Gaussian and Log-logistic in blue and purple colors respectively mimic the path traced out by the data and the non-parametric curve.

Figures 7a-d display the plots of quantile of first birth interval data against quantile of each of the four distributions. In figure 7a above, quantile-quantile plot shows the clustering of the data points on the straight line with Inverse Gaussian distribution. While in figure 7b, the points are not as much as clustered along the straight line with Log-logistic distribution as compared to figure 7a of the Inverse Gaussian. However, figure 7c below shows the quantile plot of FBI data against the quantile of Weibull distribution. The clustering of the points on the straight line is somewhat sino-sidal in shape. Figure 7d is the quantile plot of FBI data versus the quantile of Burr Type XII distribution.
Discussions

The study explores four parametric models to fit the waiting time to first birth after marriage among women of reproductive age in Nigeria, using the 2008 Demographic and health Survey data. The Non-parametric estimates of the mean and median survival time to first birth after marriage are found to be 28.8 and 20.0 months respectively. These estimates are not too distant away from the estimates obtained for Ghana using similar data from 2008 Ghana Demographic and Health Survey. In Ghana, (Logubayom & Luguterah, 2013) Logubayom reported that majority (74%) of the women had their first birth within 36 months of marriage and the mean waiting time to first birth among Ghanaian women was 30 months.

In survival analysis the means are obviously skewed to the right because of the presence of large values at the extreme, the best choice of descriptive measure is obviously the median. Log-logistic model gives the closest value of median time to first birth of 21 months to that of Kaplan-Meier estimate of 20 months. This model also produces the narrowest confidence interval for the estimated median among all the four models investigated. This is closely followed by Inverse Gaussian model using these features as displayed in Table 4. However, Burr gives the widest confidence interval with a large value of standard error accompanying the estimated 18 months of median waiting time to first birth, an indication that the model might not be appropriate for the data.

The quantile-quantile plot is a major diagnostic for checking model adequacy. The more closely the plot pattern is to the straight line the more evidence there is in support of the model (Tableman, 2008). The plot of quantile of the FBI data against the Inverse Gaussian distribution appears to lie closer to the straight line than the remaining quantile-quantile
plots as shown in figure 7a. Furthermore, the quantile-quantile plot of the data against the log-logistic distribution in figure 7b also seems to describe the data but not as adequate as the Inverse Gaussian, while Burr Type XII shows wider deviation from the data as shown in figure 7d.

Probability plot is another graphical technique for assessing whether a sample data conforms to a hypothesized distribution, based on subjective visual examination of the data. The difttool in matlab provides the avenue for this, such that if the hypothesized distribution adequately describes the data, the plotted points fall approximately on a straight line, and if the plotted points deviate significantly from the straight line, especially at the ends, then the hypothesized distribution is not appropriate. Figure 4 shows that Inverse Gaussian lies consistently on the plotted data points with a little deviation at the end, which is an indication of the distribution’s conformity to the data. However, Burr Type XII shows a significant deviation at the end as clearly shown in the same figures 4.

The model selection criteria of minimum value of AIC identifies Inverse Gaussian as the model that best describes the data with the minimum value of 116,617.2, the value which is followed by Burr’s value of 116,690.4 as given in table 2. This result of Burr low value of AIC buttresses the sensitivity of the model to large values as shown in the discrepancy between the mean survival time of 59 months and its corresponding median value of 18 months.

In the light of series of graphical comparisons made in the data analysis which lend support to the AIC value, Inverse Gaussian appears to describe the data better than the rest of the models. However, adoption of Log-logistic distribution to model waiting time to first birth after marriage is not completely ruled out by this paper more so that the model has one time been employed to model competing risk to first birth in an Iranian population by Ayatollah (Shayan, Ayatollahi, & Zare, 2011). Furthermore, adaptability of Burr Type XII distribution to model waiting time to first birth may also be investigated possibly through data transformation.

There may be some factors affecting this interval. Investigating the influence of these factors on the interval may be appropriate to understand further the dynamics of time to first birth after marriage.

References


4/15/2014