Research Journal of Mathematical and Statistical Sciences ISSN 2320–6047
Vol. **3(8),** 1-4, August **(2015)** *Res. J. Mathematical and Statistical Sci.*

Projection of Sex Ratio at Birth in U.P. India using MCMC Technique in Bayesian Procedure

Pandey Abhishek, Singh O.P. and Kumar U.
Department of Statistics, Udai Pratap Autonomous College, Varanasi, Uttar Pradesh, INDIA

Available online at: www.isca.in, www.isca.me
Received 15th July 2015, revised 28th July 2015, accepted 10th August 2015

Abstract

In the present paper we have projected Sex Ratio at Birth in Uttar Pradesh using Linear regression model, assuming lower and upper limit from the past estimates of Fraction Of Female at Birth from the available Sample Registration System data available from 1999 to 2009. Parameters of the model have been estimated using MCMC (Monte Carlo Markov Chain) Technique in Bayesian Procedure. We have assumed Non-informative prior distribution to implement the Bayesian approach for the parameter estimation.

Keywords: Linear Regression Model, Bayesian methodology, Sex Ratio at Birth, Fraction of Female at Birth(ffab), Bayesian Methodology, Non-informative prior.

Introduction

Projections are conditional statements gives the idea about future population. So it helps the state government and government of country to make future planning to meet the enhanced requirement of increased population.

On the basis of projected population planning commission and policy makers can make policies and program for the people. In general, population projections are treated as predictions and are never to be termed as final population.

At present time there are two main approaches in statistics viz. conventional and Bayesian approach for data analysis. Analyzing data by Bayesian procedure is a new approach. Its popularity and faith in the people of various discipline has been increased since last twenty years. Difficult situations can be handled by BUGS due to its flexibility and general approach. In the present work we have used the Bayesian approach for the purpose of data analysis.

We compare the relative strength of number of males and females in a population by a sex ratio. It is defined as under-

$$Sex\ ratio\ at\ birth = \frac{Number\ of\ male\ babies\ born\ during\ a\ year}{Number\ of\ female\ babies\ during\ the\ same\ year} X1000$$

Sample Registration System has provided the data of sex ratio at birth for very short duration. Data of Sex Ratio at Birth is provided by Registrar General of India from 1999 to 2009 (centered by three year moving average).The estimated values of Sex Ratio at Birth from 1999-2009 are shown in table-1.

Congdon P[1]., Dyson T. et al[2], Gelman A et al[3], Gilks et al[4], Gill J[5], altogether gives the ideology about the Bayesian Method and about Monte Carlo Markov Simulation that how to simulate and analyze the samples from our observed data, and forecast the result in a given confidence interval. Rahul et al[6], Rahul, Pandey G.S. et al[7], Rahul, Singh G.P. et al[8] suggested the method Projecting Population applied to states of India and India as a whole using Time series data using a suitable model and running a program in WinBUGS. Registrar General Of India[9], provides report on growth and nature of futuristic population of India and its states. Spiegelhalter eta al[10], teaches the methodology of running of WinBUGS software through various worked out examples.

Objective: The objective of the present paper is to examine the past and futuristic trends in Sex Ratio at Birth(SRB) in Uttar Pradesh. Time series estimates SRB for the Uttar Pradesh has collected from various SRS Statistical Reports of the year ranging from 1999 to 2009.Our objective is to project Sex Ratio at Birth for each age group separately using the linear regression model in the Bayesian frame work and from these estimates to compute Sex Ratio at Birth for the future.

Methodology

Bayesian Methods: Bayesian method provides new technique of analyzing the data. This method of analyzing data got enormous popularity in the various discipline. At first our attempt is to make a probabilistic model that is considered to explain properly the underlying mechanism of the system based on our past study and procedure of collecting samples. After that our aim is to formulate appropriate prior distributions unknown quantities of the model. Baye's rule is applied after observing the past data to get the posterior distributions for these desirable parameters, which depends on the conditional probability distributions given the observed data. The rule may be expressed symbolically as follows –

$$P(\theta/x)=\frac{P(\theta)xP(x/\theta)}{P(x)}=\frac{prior\ x\ likelihood}{Marginal}=\frac{P(\theta)xP(x/\theta)}{\int P(\theta)xP(x/\theta)d\theta} \quad [1]$$

Here**,** θ is the set of unobserved quantities of interest/parameters, P(θ) is the prior distribution of θ, P(x/ θ) is the probability distribution of data x given prior distribution and information of θ which is popularly called likelihood function of data x, and P(θ/x) is called posterior distribution of parameters/unobserved quantities of interest θ. As soon as we obtain posterior estimates of the parameter θ, we can use this distribution to provide estimates of parameter θ.

Model: Let us suppose that $A_{2[i]}$, fraction of female at birth in Uttar Pradesh in the year ti (i = 1, 2,…, 20) where i starts from 1999and lasts up to 2051.where ti takes values 1999, 2000, 2001,…….2009, 2011, 2016,2021, 2026, 2031, 2036, 2041, 2046, 2051. The data of fraction of female at birth (ffab) is given in the Table 2. We have used the Linear regression model for the projection of fraction of female at birth.

$$(Fraction\ of\ \ female\ at\ birth)_t = a+bt+\varepsilon_t\ ;\ where\ ll < \ ffab < ul \quad [2]$$

Where time is represented by t, ll is the lower asymptote and ul is the upper asymptote of ffab . "ll" is lower asymptote taken as 96% 0f the minimum of ffab (A2)U.P. and "ul" is the upper asymptote taken as 104% of maximum of f_{fab}(A2) of Uttar Pradesh. Each parameters are projected using this model in the Bayesian framework and from these estimates Sex Ratio at Birth is evaluated. For the Bayesian Model we assume that (Fraction of female at birth)$_t$ ~ N(eta$_t$,tau).Where eta$_t$ is the deterministic part of the model, again the prior distribution of intercept "a" and slope "b" are assumed to be non-informative N(0,0.0000001) and τ is given prior Gamma(0.001, 0.001). All these priors are non-informative providing limited information and we do not have information of specific nature of their probability distribution. A more rigorous discussion on the choice of non informative priors is available in Win BUGS manual by Spiegel halter, Thomas, Best, and Gilks[10].

Tools: Posterior distributions of Bayesian method involves complicated mathematical terms. Most of them can be handled by Monte Carlo Markov chain simulation method. The Markov Chain Monte Carlo (MCMC) method is a repetition procedure of generating samples from our distribution. We have used this method for handling the difficulties which arises due to typical mathematical terms that involves expected value of the function of a random variable. The calculation can be made much easier by generating large number of independent samples by simulation procedure from the (complex) distribution of the random variable. After that we take the mean of obtained values of the function from these sample points. Win BUGS (Bayesian inference Using Gibbs Sampling for Windows) is a freely available software that helps us to find out the estimates of unobserved quantities of ultimate interest by using MCMC process. This procedure requires running a number of chains starting with one chain initially which can be increased up to three (default) or more for each parameters. It requires large number of iterations to reach to the stationary distribution. If we further update the model then it is supposed that the samples are drawn randomly from the posterior distribution of the parameters. In WinBUGS there are number of inbuilt functional tools that checks the convergence of the chains. Generally one can use multiple diagnostics on a single chain. In WinBUGS we can run multiple chains simultaneously for each parameter. We have used some of the diagnostics available with the WinBUGS that is briefly described below. For convergence of MCMC simulations we run a number of chains in it. WinBUGS provides dynamic trace plot of the chains while updating the model. When we cannot see sufficient mixing of chains even after lots of updates, it indicates lack of convergence of the chains. The bgr-diagnostics calculates the modified form of Gelman-Rubin convergence statistic; see Brooks and Gelman (1998)[3]**.** Green running plots are of the statistic in which the width of the central 80% interval of the pooled, Blue running plots of the average width of the 80% intervals within the and the red plot shows their ratio *R* (= pooled / within) are provided by WinBUGS. Brooks and Gelman (1998)[3] told that we should be concerned with convergence of *R* to 1, and pooled and within interval widths should converge to got stability. In WinBUGS we can get smooth density plots of the chains. The density curve takes bell (normal) shape when the chains approach to stationary. The absence of convergence of the chains indicates lack of normality. There is another diagnostic tool available inside BUGS namely Auto-correlation. When the chains converge to the stationary distribution then autocorrelation decreases with the increase in the lags. The basis to reach the convergence of the chain is also provided by it. A detailed discussion on the diagnostics can be found in Gill[5].

When it seems that chains have converged, then this simulation procedure can be continued for a further number of iterations to obtain the samples that can be used for posterior inference. The accuracy of our posterior estimates will increase when we generate and include more samples in the iteration process. After running the adequate number of updates and got satisfied by of history of chains, we can exclude the previous samples. Summary statistics can only be obtained from the further generated samples.

Analysis: Table-1 given below is observed data of Sex Ratio at Birth Uttar Pradesh as reported by SRS (Sample Registration System, Registrar General of India). It includes the data continuous 11 SRS reported data starting from 1999 to 2009. Table-2 given below is the represents the fraction of female at birth evaluated by dividing number of female babies born in a year divided by number of male babies born in the same year.

Table-1
Sex Ratio of U.P., India

Year	srb
1999	1152
2000	1149
2001	1157
2002	1172
2003	1164
2004	1160
2005	1144
2006	1135
2007	1140
2008	1144
2009	1149

Table -2
Fraction of female at birth(U.P., India)

Year	ffab
1999	0.464668
2000	0.465241
2001	0.463519
2002	0.460335
2003	0.462076
2004	0.462943
2005	0.466382
2006	0.468368
2007	0.467235
2008	0.466382
2009	0.465241
2011	NA
2016	NA
2021	NA
2026	NA
2031	NA
2036	NA
2041	NA
2046	NA
2051	NA

To obtain the Bayes estimates of parameters of the model described in equation (2) a program was written in the language of WinBUGS. After running this program in WinBUGS we summarize the estimates after discarding 15,000 initial updates. There was no indication of lack of convergence of chains. Number of iteration required after achieving convergence can be decided by seeing MC error of the parameters. Simulation process should be continued till MC error is reduced by 5% or more for each parameter. Thus 30,000 updates were run after the initial burn in. MC error for each parameter was found below 5%. In addition to controlled MC error, the Karnel density of each parameter was also found to be well in normal shape. It was also seen that the auto correlations for different lags had declining trend with increase in lag. The bgr diagnostic for all parameters were close to one. After being confirmed with all the diagnostic about convergence of the parameters, the estimates (posterior mean which is considered as Bayes estimate under quadratic loss function) were obtained which are given in table-2. Table-3 shows Projected value of Sex Ratio at Birth. Figure-1 shows the density function of the estimates which is found bell shaped for the most of the estimates. Figure-2 shows bgr-dignostic plot for all the estimates. Which shows convergence. Figure-3 shows the Time series plots found for all the estimated values of the parameters. of the chains. Figure-4 shows Auto Correlation plots found for most of the estimates which shows continuous decline with the lag.

Table-3
Projected Value of Sex Ratio at Birth and other parameters

year	sexratio 2.5%	sexratio.mean	sexratio 97.5%
2011	1193	1153	1116
2016	1200	1154	1112
2021	1211	1155	1103
2026	1225	1156	1092
2031	1240	1157	1081
2036	1257	1158	1069
2041	1274	1159	1057
2046	1291	1160	1045
2051	1308	1161	1032

other parameters	25%	mean	97.50%
sigma	0.009644	0.01408	0.02078
a	0.4545	0.4642	0.4736
b	-0.01175	-6.43E-04	0.0103
tau	2316	5656	10750

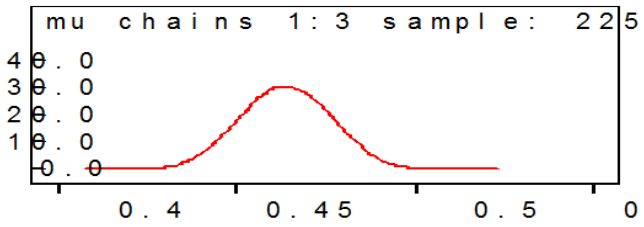

Figure-1
Karnel Density for all parameters

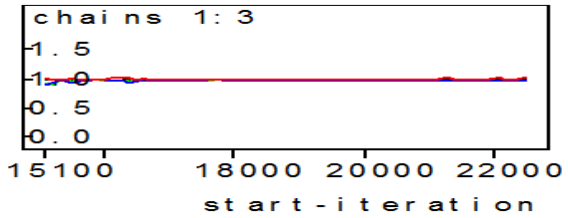

Figure-2
Gelman-Rubin(Bgr-dignostics) for all parameters

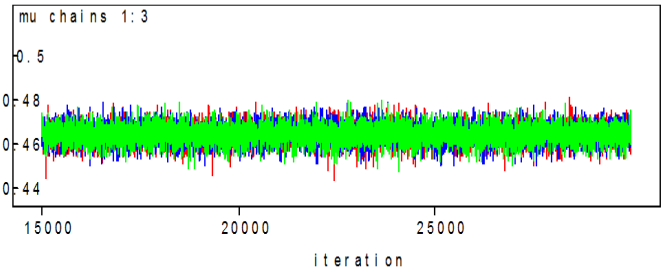

Figure-3
Time series plot for parameters

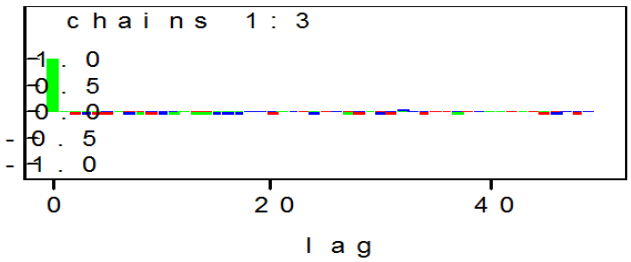

Figure-4
Auto correlation for all parameters

Table-3 shows the estimated values of SRB in 95% confidence interval. SRB in the 2.5% column shows approx linear increment with time. SRB in the 97.5% column declines slowly. But the mean value of SRB approximately constant with the passage of time

Conclusion

We used a Bayesian approach, implemented in WinBUGS, to check the suitability of Linear regression model for the growth of Sex Ratio at Birth data obtained from different SRS reports of UP . Our main focus was to develop the methodology and program for Bayesian Projection. The estimated values of the parameters of proposed model are shown in table-3. The table shows interval estimates (95% Highest Posterior Density) for all the mentioned age groups and projections in different years. The model suggests that the SRB of U.P will increase in 2.5% region and will remain nearly constant for the mean values .SRB will decline for the 97.5% region. In all the Bgr-plots for all the age groups found to converge. Some efforts needed by people and government of India for checking sex selective abortion so that to achieve sex equality in population.

References

1 Congdon P., Bayesian Statistical Modeling, John Wiley and Sons, Chichester, U.K.**, (2001)**

2 Dyson T., India's Population-The Future, in T. Dyson R. Cassen and L. Visaria, Eds., Twenty-First Century India: Population, Economy, Human Development, and the Environment. Press, Oxford, 74-10, **(2004)**

3 Gelman A., Carlin J.B., Stern H.S. and Rubin D.B., Bayesian Data Analysis, Second Edition, Champman and Hall/CRC, **(2003)**

4 Gilks W.R. Richardson S. and Spiegelhalter D.J., Eds., "Markov chain Monte Carlo in Practice, Champman and Hall, London, U.K., **(1996)**

5 Gill J., Bayesian Methods: A Social and Behavioral Sciences Approach, Champman and Hall/CRC., **(2002)**,

6 Rahul and Singh O.P., Population Projection Of Indian Using MCMC Technique in Bayesian Theory (2011-2051), published by LAP LAMBERT Academic Publishing, Germany., **(2012)**

7 Rahul, Pandey G.S. and Singh O.P., Population Projection of India Using MCMC Technique: A Win BUGS Implementation, Demography India, New Delhi, India, **36(1),** 145-154, **(2007)**

8 Rahul Singh G.P. and Singh O.P., Population Projection of Kerala Using Bayesian Methodology, *Asian Journal of Applied Sciences,* **2(4),** 402-413, (**2009**)

9 Registrar General of India, Population Projection for India and States 2001-2026",Office of the Registrar General and Census Commissioner, India, **(2006)**

10 Spiegelhalter D.J. Thomas, A. Best, N.G. and Gilks W.R., BUGS: Bayesian inference Using Gibbs Sampling, Version 0.5. MRC Biostatistics Unit, Cambridge, **(1996)**