Research Journal of Mathematical and Statistical Sciences ISSN 2320–6047
Vol. **3(7),** 9-14, July **(2015)** *Res. J. Mathematical and Statistical Sci.*

Projection of some Fertility Indicators of India using MCMC Technique in Bayesian Procedure

Abhishek Pandey and Om Prakash Singh
Department of Statistics Udai Pratap Autonomous College, Varanasi, Uttar Pradesh, INDIA

Available online at: www.isca.in, www.isca.me
Received 30th May 2015, revised 25th June 2015, accepted 10th July 2015

Abstract

In the present paper we have projected the Age Specific Fertility and Total Fertility Rate of Uttar Pradesh using Gompertz model, assuming lower and upper asymptotes from the past estimates of Age Specific Fertility rate of Uttar Pradesh. 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: Gompertz Model, Bayesian methodology, Age specific fertility rates, Total Fertility rate. Bayesian Methodology, Non-informative prior.

Introduction

Human development and improvement in quality of life are the ultimate objectives of all Planning. This is to be achieved through policies and programs aimed at promotion of both equity and excellence. Population growth has long been a concern of the government, and India has a lengthy history of explicit population policy.

The idea about future population is obtained by the help of projections. Projections are conditional statements about the future. They refer mostly to the exercises of extrapolation of the past trends into the future; and they do not take into account changes in the policy parameters. For example, a projection of the future population growth may not be taking into account changes in the government health policies, selected to family welfare programs, etc. Projections are based on the assumption that the past trends will continue to operate in the future. The reliability and usefulness of projections depend on the assumptions and their closeness to reality. In the long run, the policy parameters are to be incorporated in the projections.

The likely effects of policy changes are to be judged and projections are to be made accordingly. Thus, when an element of judgment is added to the projections, it becomes a forecast. Forecasts enjoy the advantage of being based upon a set of assumptions which are likely to be realized in the near future and can yield a relatively more realistic picture of the future. 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.

The total fertility rate (TFR) is one of the main components in population projections. It is the average number of children a woman would bear if she survived through the end of the reproductive age span, experiencing at each age the age-specific fertility rates of that period. Uttar Pradesh is the largest populated state among the so called BIMARU states, Uttar Pradesh ranks 15 among the states of India based on the report Inequality-adjusted Human Development Index for India's states published by UNDP India.

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, 2006[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 Age Specific Fertility Rate and Total Fertility Rate of Uttar Pradesh. Time series estimates of Age Specific Fertility Rates (ASFR) and Total Fertility Rates (TFR) for the Uttar Pradesh has collected from various SRS Statistical Reports of the year ranging from 1971 to 2010.Our objective is to project Age Specific Fertility Rates for each age group separately using the Gompertz model in the Bayesian frame work and from these estimates to compute Total Fertility Rate.

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:–

$$P(\theta / x) = \frac{P(\theta)\, x\, P(x/\theta)}{P(x)}$$

$$= \frac{prior\ x\ likelihood}{M\ \arg inal}$$

$$= \frac{P(\theta)\, x\, P(x/\theta)}{\int P(\theta)\, x\, P(x/\theta) d\theta} \qquad (1)$$

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 $A_{1[i,j]}$, denotes Age Specific fertility rate of i[th] age group at j[th] year Uttar Pradesh in the year ti (i = 1, 2,…, 81) where i starts from 1971 and lasts up to 2051.where ti takes values 1971, 1972,1973,…2010, 2011, 2016, 2021, 2026, 2031, 2036, 2041, 2046, 2051. The data are given in the Table 1. We have used the Gompertz model for the projection of ASFR for any of the age group, stated as under -

$$\frac{ASFRt-ll}{ul-ll} = a^{b^t} \text{ where, } 0<a<1 \text{ and } b>1 \quad \textbf{[2]}$$

Where time is represented by t, ll is the lower asymptote and ul is the upper asymptote of ASFR for a specific age group a is called lag parameter indicating the value of the left hand side at the time t = 0 and b is the parameter which measures the speed of fall ASFRs (Age Specific Fertility Rates) for each age group were separately projected using this model in the Bayesian framework And from these estimates Total Fertility Rates are computed. While using this model we have provided some prior information about the lower and upper asymptotes of ASFRs in the model. In all the projections we have assumed upper asymptote as maximum of ASFRs in the past for each corresponding age group and lower asymptotes to vary between 82% and 121% of the lowest value of corresponding age group observed in Kerala in the year 2010.

We have assumed non informative prior for each unknown parameter of the model using normal distribution with the large variance. Thus we have non-linear regression model as $ASFR_t = eta_t + \varepsilon_t$ where eta_t is the deterministic part of the model as ll +(ul – ll) a^{b^t} depending upon time t and ε_t is the disturbance part with $\varepsilon_t \sim N(0,\tau)$. Naturally $ASFR_t \sim N(eta_t, \tau)$ where τ is taken as common precision (=1/variance) for all time. As stated above, 'ul' is assigned assumed value equal to maximum of Age Specific Fertility Rates in the past in the corresponding age group of Uttar Pradesh. 'll','a', and 'b' all are given prior N(0, 0.0000001) and τ is given prior Gamma (0.001, 0.001) under the condition stated above. 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 WinBUGS manual by Spiegelhalter, 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. WinBUGS (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[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[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 Uttar Pradesh as reported by SRS (Sample Registration System, Registrar General of India). It includes the data continuous 40 SRS reported data starting from 1971 to 2010.

Table-1
ASFR of Uttar Pradesh from 1971 to 2010

Age group	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
15-19	98.8	107.1	98.9	103.6	101.9	105.1	103.7	99.6	111.9	94
20-24	280.1	273.4	268.3	269.3	274.3	274.5	275.5	279.3	267.6	267.6
25-29	289.6	285.6	279.9	272.7	292.6	264.3	303.1	284.1	284.4	280.9
30-34	246.6	248	227.9	322.7	247.4	239.5	240	254.3	220.4	239.5
35-39	187.9	168.9	171.7	158.2	165.6	173	160.8	151.9	164.8	156.3
40-44	93.9	91	91.8	89.9	89.6	102.4	97.4	87.5	66.9	91.3
45-49	43.4	47.2	42.3	41.5	43.3	24.9	38.3	35.4	42.2	42.3

Age group	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
15-19	91.5	87.5	91.1	89.5	93.9	89.2	87.8	88.5	79.8	74.8
20-24	277.6	263.4	275.4	280	282.9	276.1	287.6	281	268.8	265.2
25-29	281.9	271.8	285.7	290.9	276.4	272.9	270.8	273.1	258.9	254.4
30-34	230.7	230.4	226.7	228.1	219.8	211.2	206.3	201.8	202.8	202.4
35-39	159.7	153.8	154.5	156.2	133.5	135.9	139.6	134.1	136.5	141.7
40-44	74.4	83.9	87.3	83.7	70.8	70.1	69.1	64.9	74.4	64.7
45-49	40.5	45.9	48.6	47.7	34.6	24.5	31.6	30.7	27.5	32.8

Age group	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
15-19	72.2	72.9	72.3	51	42.1	40.1	39.5	40.7	43	40.5
20-24	270.9	272.4	278.6	268.9	254.5	261.1	248.2	245	239.5	260.2
25-29	255	266.5	266.9	275.6	274.1	255.1	270.6	251.7	262.6	258.1
30-34	196.9	200.4	203.4	200.7	208.6	202.1	198.3	188	192.5	187.4
35-39	126.9	138.6	130.5	132.2	121.1	122.2	115.1	115.6	129.4	125.6
40-44	66.9	70.3	66.6	68.2	70	68.2	62.3	59.7	58.8	54.4
45-49	25	28.7	27.6	31	28.7	29.9	22.8	23	22.2	20.9

Age group	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
15-19	37.1	30.7	29.1	48.7	39.7	40.2	33.4	32.6	31.9	32
20-24	243	243	241.7	270.1	282.6	251.3	269.7	233.6	239	226.4
25-29	244.8	252.2	256.1	258.2	231	261.5	234.1	243.2	231.1	233.9
30-34	188.9	178.1	175.5	153	161.6	144.8	137.3	131.7	138	111.7
35-39	117.1	113.9	104.8	85	77.3	82.8	72.2	75	63.9	67.3
40-44	54.7	51	51.2	42.7	39.3	36.8	32.4	30.3	29.9	22.1
45-49	21.7	20.1	19.2	20.2	14.9	15.3	10.1	15.3	12.9	10.7

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 Bayesian R^2 for the goodness of the fit for seven age groups, which is found to be good fit for the all the age groups stated above. Figure-1 shows the density function of the estimates which is found bell shaped for the most of the estimates. Figure-2 shows time series plot of the estimates. Figure-3 shows the mixing of the chains. Figure-4 shows bgr-trace.

Table-2
Projected values of the Age Specific Fertility Rate and Total Fertility Rates

		2011			**2016**			**2021**		
	Age	**2.50%**	**Mean**	**97.50%**	**2.50%**	**Mean**	**97.50%**	**2.50%**	**Mean**	**97.50%**
ASFR	15-19	12.78	31.01	49.32	8.521	26.4	44.47	5.914	23.35	41.3
	20-24	**206.6**	233.4	259.3	180.7	215.7	246.7	151.2	195.1	233
	25-29	201.5	224.7	247.2	164.3	196.4	226.7	109.7	159	201.4
	30-34	79.58	115.8	152.2	48.3	85.72	122.8	28.36	64.3	100.2
	35-39	48.59	65.88	83.06	30.69	48.08	65.59	16.52	33.72	51.51
	40-44	19.51	32.47	49.24	9.291	22.9	41.49	2.595	15	33.7
	45-49	7.649	17.86	28.74	5.186	15.35	26.17	3.244	13.03	24
TFR		**3.3**	**3.6**	**3.9**	**2.7**	**3.1**	**3.4**	**2.1**	**2.5**	**2.9**

		2026			**2031**			**2036**		
	Age	**2.50%**	**Mean**	**97.50%**	**2.50%**	**Mean**	**97.50%**	**2.50%**	**Mean**	**97.50%**
ASFR	15-19	4.829	21.58	39.19	3.801	20.34	38.1	3.843	19.8	37.34
	20-24	130.1	174.8	218.9	119.4	157.6	204.5	114.2	145.9	189.7
	25-29	54.6	115.8	172.3	22.31	75.53	139.7	10.5	46.45	105.5
	30-34	20.21	54.17	87.73	17.95	50.95	84.61	17.39	50.44	83.17
	35-39	7.58	24.1	41.19	4.072	19.19	35.48	2.796	17.31	33.06
	40-44	0.6864	10.02	27.44	0.3473	7.411	22.32	0.275	6.262	19.13
	45-49	1.899	11.13	21.97	1.058	9.43	19.98	0.721	8.173	18.44
TFR		**1.615**	**2.1**	**2.5**	**1.3**	**1.7**	**2.2**	**1.1**	**1.5**	**1.9**

		2041			**2046**			**2051**		
	Age	**2.50%**	**Mean**	**97.50%**	**2.50%**	**Mean**	**97.50%**	**2.50%**	**Mean**	**97.50%**
ASFR	15-19	3.423	19.56	37.2	3.335	19.41	37.12	3.35	19.3	37.14
	20-24	111.5	138.7	175	109.9	134.9	164.9	109.3	132.9	158.6
	25-29	6.129	30.72	73.85	4.183	24.09	51.1	3.777	21.73	42.34
	30-34	17.26	50.46	83.46	17.92	50.34	83.66	17.84	50.4	83.13
	35-39	**2.822**	16.88	32.69	2.666	16.69	32.45	2.635	16.72	32.46
	40-44	0.2402	5.783	16.96	0.229	5.546	16.03	0.2055	5.441	15.7
	45-49	0.4986	7.152	17.23	0.3734	6.411	16.16	0.2902	5.845	15.24
TFR		**1.1**	**1.4**	**1.7**	**1.0**	**1.3**	**1.6**	**1.0**	**1.3**	**1.5**

Table-3
Bayesian R^2 for the goodness of fit

Age	mean	2.50%	97.50%
15-19	0.9407	0.9131	0.9599
20-24	0.9467	0.8948	0.9705
25-29	0.9869	0.9759	0.9925
30-34	0.9586	0.9366	0.9735
35-39	0.9822	0.9736	0.9881
40-44	0.9674	0.9116	0.9831
45-49	0.8804	0.7842	0.9326

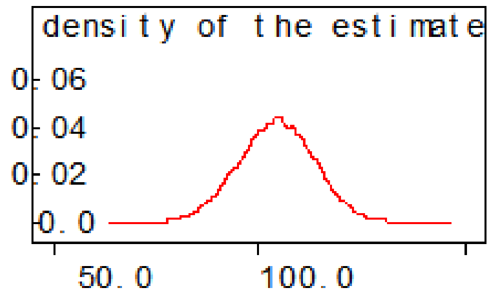

Figure-1
Karnel Density

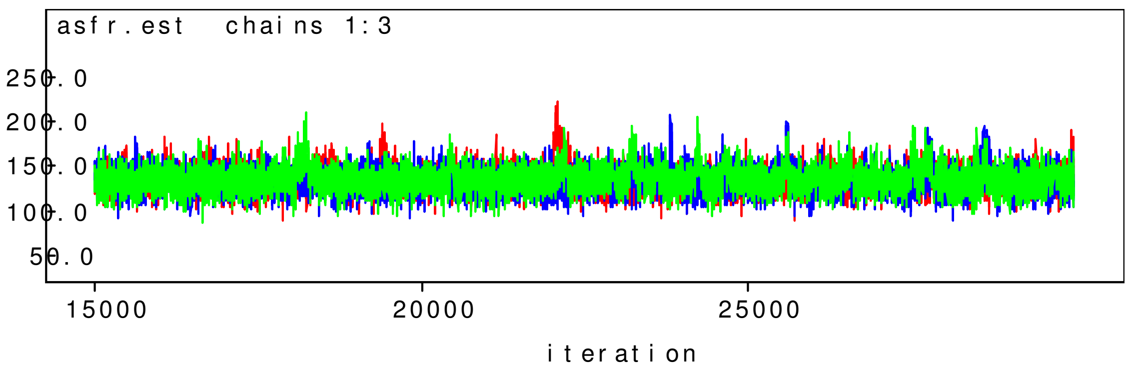

Figure-2
Time Series plot for all age groups (approximate)

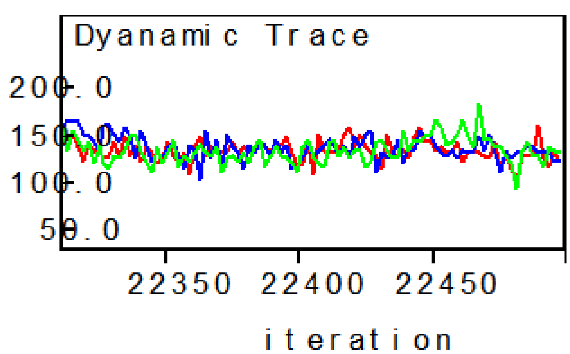

Figure-3
Dynamic Trace

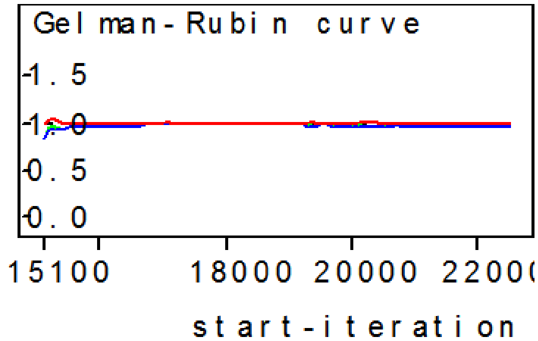

Figure-4
Bgr diagnostic

Conclusion

We used a Bayesian approach, implemented in WinBUGS, to check the suitability of Gompertz growth model on population 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-2. 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 TFR of UP will reach to its minimum of 1.262 i.e. below replacement level (2.1) this will seems to happen in 2051. TFR of Uttar Pradesh will reach below the replacement level in year 2031. Table Age Specific Fertility Rates will fall for each group. In all the Bgr-plots for all the age groups found to converge and few of the chains diverges from each other by some margin.

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 WinBUGS 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)**