Research Journal of Mathematical and Statistical Sciences ISSN 2320–6047
Vol. **3(9),** 1-6, September **(2015)** *Res. J. Mathematical and Statistical Sci.*

Projection of Brass's Alpha and Beta estimate for female in U.P., India using MCMC Technique in Bayesian Procedure

Pandey Abhishek
Department of Statistics, Udai Pratap Autonomous College, Varanasi, Uttar Pradesh, INDIA

Available online at: www.isca.in, www.isca.me
Received 20th July 2015, revised 21th August 2015, accepted 8th September 2015

Abstract

In the present paper we have projected Brass's Alpha and Beta for female population in Uttar Pradesh using Linear regression model. Alpha and Beta are estimated from past lx values for female population. 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, Brass logit formula, Bayesian Methodology, Non-informative prior.

Introduction

The Life table was devised initially by John Graunt, he published a rudimentary life –table in 1662 based on the analysis of "Bills of Mortality". It was based on mortality data only, so it was defective. After him William Farr gave a new systematic method for calculating complete abridged life-table. After him numerous attempt were made for calculating life-table. Chiang gave an improved method to calculate person years lived. A life-table is defined as the summary presentation of the death history of cohort. Cohort life-tables can be prepared for a population at a given for present and past time. Population Projection is the method of calculating future population by using available population data. Reliability of projected estimates is depend upon valid assumption. Projection are the probabilistic forecast so there is no requirement to check its accuracy. 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. There are difference between male and female population in Uttar Pradesh so we have projected the α and β for female population obtained from past estimates.

Brass's method is different among all the available methods, it relates two lx values of two life tables.
Brass formula is given as-

$$\log it(l - lx^*) = \alpha + \beta \log it(1 - lx)$$

$$\log it(l - lx) = 0.5 \log\left(\frac{1 - lx}{lx}\right)$$

Parameters α and β are supposed to be constant for all the lx values, it can be obtained by solving to simultaneous equations obtained by using successive lx values into two sections. We can project the values of α and β for future in 95% confidence interval. In this paper we have projected these parameters under Bayesian framework using MCMC technique.

Sample registration system has provided the data of Age specific death rate for years ranging from 1971 to 2010. From these data α and β values were calculated. The estimated values of α and β are given 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 theBayesian 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. Spiegel halter 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 α and β in Uttar Pradesh. Time series estimates Age specific death rates for the Uttar Pradesh has collected from various SRS Statistical Reports of the year ranging from 1971 to 2010. Our objective is to project α and β separately using the linear regression model in the Bayesian frame work.

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 –

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

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

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

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 $A_{3[i,j]}$ is matrix representing the α and β for the female population for years 1971 to 2010 in Uttar Pradesh Pradesh in the year ti (i = 1, 2,…, 49) where i starts from 1971and lasts up to 2051.where ti takes values 1971, 2000, 2001,…….2009, 2011, 2016, 2021, 2026, 2031, 2036, 2041, 2046, 2051. The data of α and β is given in the Table 2. We have used the linear regression model for the projection α and β in the 95% confidence interval in Bayesian framework. All the parameters are projected using this model in the Bayesian framework.
WinBUGS language program for projection of α and β:-

```
model{
for ( i in 1:2) {
for(j in 1:N){
A3 [i,j]~ dnorm(eta[i,j],tauC[i])I(,)
eta[i,j]<-phi[i,1]+phi[i,2]*(time[j]-mean(time[]))/sd(time[])
A3.rep [i,j]~dnorm(eta[i,j],tauC[i])I(,)
}
}
for(i in 1:2) {
tau C[i]~dgamma(0.001,0.001)
sigmaC[i]<-1/sqrt(tauC[i])
}
for( i in 1:2){
for(k in 1:2) {
phi[i,k]~dnorm(0,1.0E-6)I(,)
}
}
## projection of alpha and beta
For (j in 1:BASEANDSTEPS) {
Alpha. f[j]<-eta[1,40+j]
Beta. f[j]<-eta[2,40+j]
}
## Computation of bayesian R square
For (i in 1:2){
For (j in 1:40){
c. A3[i,j]<-A3[i,j]-mean(A3[i,])
}
s2[i]<-sigmaC[i]*sigmaC[i]
s2A3[i]<-inprod(c.A3[i,],c.A3[i,])/(39)
R2B[i]<-1-s2[i]/s2A3[i]
}
}
```

"α" and slope "β" are given 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 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.

Table-1
The estimated values of α and β

time	alpha	beta
1971	0.561552	0.740981
1972	0.758031	0.839865
1973	0.554898	0.764922
1974	0.557027	0.789456
1975	0.664744	0.806829
1976	0.686049	0.828256
1977	0.596992	0.840931
1978	0.638039	0.781473
1979	0.432281	0.800552
1980	0.416382	0.800178
1981	0.410964	0.78396
1982	0.389924	0.830873
1983	0.429443	0.900164
1984	0.497169	0.834793
1985	0.437014	0.860456
1986	0.369924	0.829898
1987	0.390219	0.849273
1988	0.304508	0.854233
1989	0.275276	0.875964
1990	0.274386	0.917145
1991	0.224217	0.912063
1992	0.322556	0.921629
1993	0.212869	0.926295
1994	0.181984	0.912674
1995	0.159109	0.912581
1996	0.182131	0.918504
1997	0.18676	0.921053
1998	0.245107	1.007812
1999	0.206154	1.002889
2000	0.147829	0.946539
2001	0.121035	0.922641
2002	0.135249	0.983762
2003	0.106373	0.986508
2004	0.076294	0.972547
2005	0.076326	0.961455
2006	0.055015	0.959835
2007	0.072947	1.032022
2008	0.081854	1.041047
2009	0.064512	1.090895
2010	0	1
2011	NA	NA
2016	NA	NA
2021	NA	NA
2026	NA	NA
2031	NA	NA
2036	NA	NA
2041	NA	NA
2046	NA	NA
2051	NA	NA

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[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 estimate of α and β for female population in Uttar Pradesh calculated by Brass's method taking 2010 as standard year. It includes the data continuous 40 years. SRS reported ASDR data starting from 1971 to 2010. Table-2 given below is the represents the estimated value of Brass's α and β. Table-3 represents other parameters. Table-4 shows Bayesian R^2 for testing goodness of fit.

Table-2
Projected values of α, β and other parameters

Year	Alpha.f.2.5%	Alpha.f.mean	Alpha.97.5%	Beta.f.2.5%	Beta.f.mean	Beta.f.97.5%
2011	-0.07124	-0.02965	0.01122	1.013	1.035	1.057
2016	-0.1627	-0.1132	-0.06446	1.043	1.069	1.094
2021	-0.181	-0.1299	-0.07962	1.049	1.076	1.102
2026	-0.1994	-0.1466	-0.09479	1.055	1.082	1.11
2031	-0.2178	-0.1634	-0.1099	1.061	1.089	1.117
2036	-0.2361	-0.1801	-0.125	1.067	1.096	1.125
2041	-0.2544	-0.1968	-0.1402	1.073	1.103	1.133
2046	-0.2728	-0.2135	-0.1553	1.079	1.109	1.14
2051	-0.2911	-0.2302	-0.1704	1.085	1.116	1.148

Table-3

	Other parameters		
Node	**2.50%**	**mean**	**97.50%**
phi[1,1]	0.1068	0.1345	0.1619
phi[1,2]	-0.3708	-0.3357	-0.3009
phi[2,1]	0.9538	0.9683	0.9828
phi[2,2]	0.1176	0.1361	0.1547
sigmaC[1]	0.05141	0.06409	0.08074
sigmaC[2]	0.02715	0.03387	0.04273

Table-4
Bayesian R^2 testing goodness of fit

S.N.	2.50%	mean	97.50%
R2B[1]	0.8686	0.9161	0.9468
R2B[2]	0.7852	0.8639	0.9141

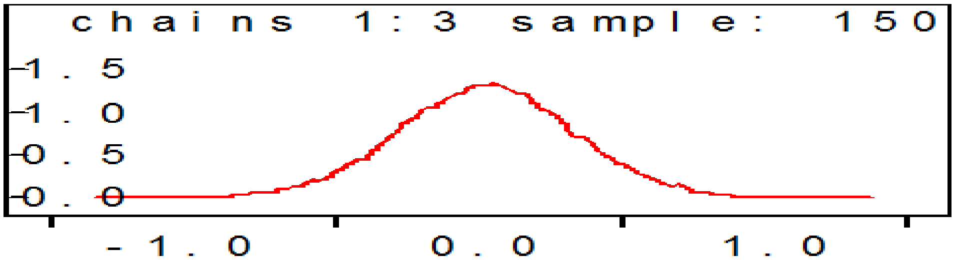

Figure-1
arnel density

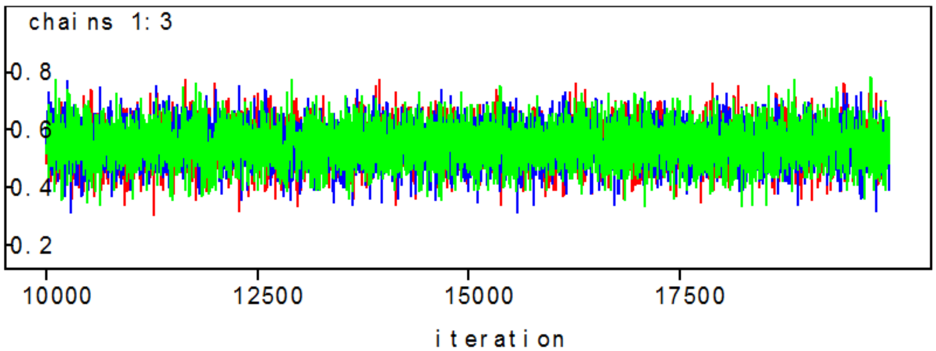

Figure-2
Time series plot

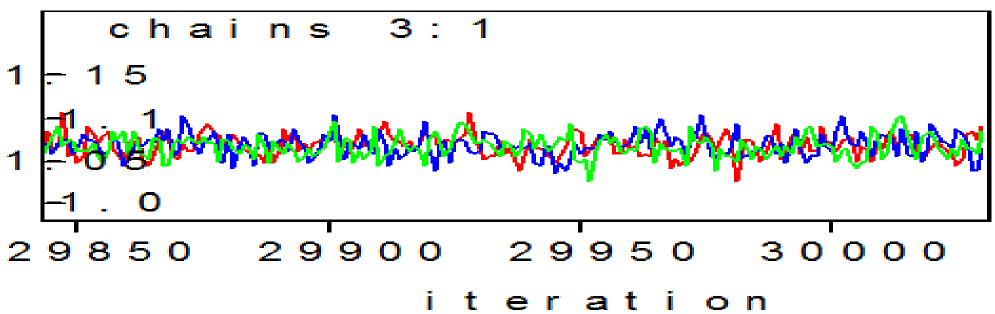

Figure-3
Dynamic Trace

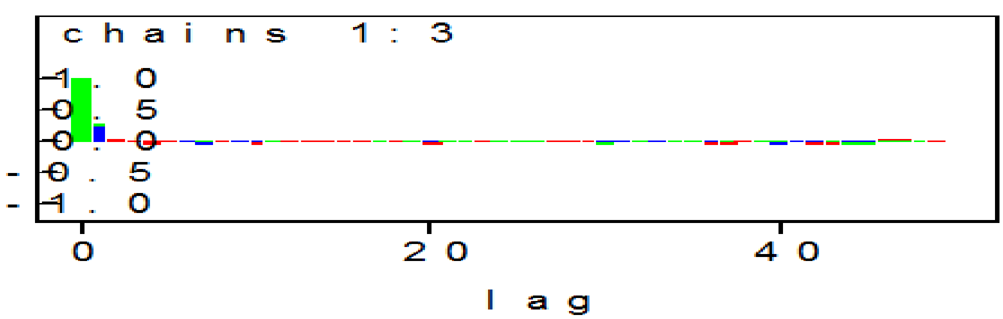

Figure -4
Auto correlation Plot

Figure-1 shows that Karnel Density for all the Parameters is found bell shaped (normal). It shows that α and β will be distributed normally in the future for female population. Figure-2 and 3 shows the mixing of chains. Figure-4 shows that autocorrelation will decrease with lag for all parameters.

Conclusion

We used a Bayesian approach, implemented in WinBUGS, to check the suitability of Linear regression model for the projected estimate of α and β. 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 parameters of the model for different years. The estimated values of α and β will be used to project the female population in the given confidence interval. Bayesian R^2 is found good fit for the model.

References

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

2 Dyson T., India’s Population-The Future, 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),** (2009), 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)**