RESEARCH PAPER | Statistics | Volume : 4 | Issue : 2 | Feb 2014 | ISSN - 2249-555X

Bayesian Frailty Model for Time to Event Breast Cancer Data

KEYWORDS: Survival data, Frailty Model, Bayesian approach, Gibbs sampler, WinBUGS

Leo Alexander T	* Pari Dayal L	Valarmathi S
Loyola College, Chennai	Head, Department of Statistics, DRBCCC Hindu College, Pattabiram, Chennai-600072 * correspondent author	The Tamilnadu DrMGR Medical University, Chennai

Ponnuraja C	Venkatesan P
NIRT(ICMR), Chennai	NIRT(ICMR), Chennai

ABSTRACT *Survival analysis has become standard tools for modeling cancer trial data when the event of interest is the "time to event". In survival analysis, the proportional hazard model was introduced by Cox (1972) in order to estimate the effects of different covariates influencing the time-to-event data. This model has been used extensively in time to event of cancer trial data. It is known that the Bayesian analysis has the advantage in dealing with small sample of censored data over frequentist methods. Frailty models in survival analysis deal with the unobserved heterogeneity among subjects. The objective of this article is to present a Bayesian analysis for survival models with frailty and is being compared with a frequentist method of proportional hazards model. Gibbs sampling technique is used to assess the posterior quantities of interest and to avoid the complexity in calculations. The posterior is arrived using WinBUGS package. An illustrative analysis is done within the context of survival time to death of breast cancer data.*

INTRODUCTION

Survival analysis technique to cancer trial data is valid when the endpoint of interest is "the time to the occurrence of a particular event". With modern computing technology, the analysis of "time to-event" data has become inexpensive in terms of time. The proportional hazards model, which simply regresses the logarithm of the survival time over the covariates, has been utilized in the analysis of censored survival data in cancer trial. In many cancer trials, the applications of proportional hazards model is often a more realistic model than the other survival models in the analysis of time to event data.

To describe the distribution of survival time has assumed that the hazard function is completely specified given the baseline hazard function and the values of the covariates. In cancer studies, there may be factors other than the measured covariates that significantly affect the distribution of survival time. This condition is often referred to as heterogeneity of the subjects. Among the early papers of Vaupel, Manton and Stallard (1979) who used the concept of frailty to describe the differences in survival time apparently among similar individuals. Hougaard (1995) presented an overview of the models, proposed for the use in time to event data. Aalen (1994) also provides a relatively non-technical summary with a focus on fully parametric models. Klein and Moeschberger (1997) presented methods based on incorporating frailty in proportional hazards models and its technical details.

The basic idea of a frailty model is to incorporate an unmeasured "random" effect in the hazard function to account for heterogeneity in the subjects. This random effect for the i^{th} cluster v_i is incorporated conditionally into the proportional hazard function previously examined:

$$h(t/v_i) = v_i h_0(t) exp(\beta x_j) \quad [1]$$

which may be re-expressed as

$$h(t/v_i) = h_0(t) exp(\beta x_j + \eta_i), \quad [2]$$

showing v_i, actually behaves as an unknown covariate for the i^{th} cluster in the model. Using the relationship between the survival and hazard function, it has the conditional survival function as

$$S(t/v_i) = Exp[v_i \Lambda_0(t) exp(\beta x_j)] \quad [3]$$

and the conditional likelihood as

$$L(\gamma, \beta \mid v_i) = \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(h(t_j \mid v_i)^{\delta^i} S(t_j \mid v_i) \right) \quad [4]$$

where there are i clusters, i^{th} one being of size η_i and γ and β represent baseline hazard and regression parameters, respectively. On substitution it gives

$$L(\gamma, \beta \mid v_i) = \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left([h_0(t) v_i exp(\beta X_j)]^{\delta_j} exp[-v_i \Lambda_0(t) exp(\beta X_j)] \right)$$
$$= \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(\frac{\phi}{\Phi} \right)^{\delta_j} \prod exp(-\Phi)(\Phi)^{\delta_i} \quad [5]$$

where

$$\phi = v_i exp(\beta' X_i) exp(\alpha' W_j) \gamma t_{ij}^{r-1}$$

$$\Phi = v_i exp(\beta' X_i) \sum_{j=1}^{n_j} exp(\alpha' W_j) \gamma t_j^r = v_i exp(\beta' X_i) e_i$$

The marginal (i.e. independent of n_i) likelihood, L(g,b), is obtained through integration of the random effect distribution. A common assumption for the random effect to follow a Gamma distribution is mean "1" and variance "t".

$$f(v_i) = \frac{v_i^{1/\tau - 1} exp(-v_i/\tau)}{\Gamma(1/\tau) \tau^{1/\tau}}$$

Volume : 4 | Issue : 2 | Feb 2014 | ISSN - 2249-555X

The marginal likelihood is then obtained as

$$L(\gamma,\beta,\alpha,\tau)=\prod_{i=1}^{I}\prod_{j=1}^{n_i}\int_0^{\infty}L(\gamma,\beta,\alpha\,|\,v_i)dG(v_i)$$

$$=\prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{\exp(\alpha v_{ij})\tau_{ij}^{r-1}}{e_i}\right)^{\delta_{ij}}\prod_{i=1}^{I}(\exp(\beta x_i)e_i\tau)^{\delta_i}\frac{\Gamma(\delta_i+1/\tau)}{\Gamma(1/\tau)}\left(\frac{1}{1+\exp(\beta x_i)e_i\tau}\right)^{\delta_i+1/\tau}$$

Since the hazard cannot be negative, distributions must have positive values. These technical issues have led to the use of the gamma distribution. In particular, the most frequently used model assumes that the frailties represent a sample from a gamma distribution with mean equal to one and variance parameter "τ". If the value of the frailty is less than one, the subject is less frail than an average subject. Aalen (1994) pointed out that there are advantages in using a fully parametric model, such as the Weibull regression model, with a frailty. The frailty model has been extended by Yashin et al. (1995). The parameter vector "τ" is then extended to include the random effects $v_i,(i=1,2,...,n)$ for each class from which the data is given and overall variance of the logarithm of random effects v_i. The data is in the form of $D_j\,(j=1,2,...,m;i=1,2,...,n)$ where m is the number of observations in clusters *i*. The variance of the frailty estimates is drawn at an each iteration using a Gibbs step assuming with full conditional for the dispersion, since it is assumed that $E(\log(v_i))=0$. In the Bayesian point of view, the random frailty models estimate the unknown frailties under the assumption of a priori exchangeability. In addition, the frailty variance "τ", would be expressed as a hyperparameter and prior knowledge concerning its value which will be summarized in a hyper prior distribution.

MCMC ALGORITHM

The most commonly used algorithm in MCMC applications are of two types and they are Metropolis Algorithm and Gibbs sampler. Geman and Geman (1984) presented the Gibbs sampler in context of spatial processes involving large number of variables like image reconstructions. They consider under which situations the conditional distributions given neighbourhood subsets of the variables, which they uniquely determines the joint distribution. The basic contribution in framework of iterative Monte Carlo algorithms were performed by Tanner and Wang(1987). Further developments in the fields are listed here the data augmentation by Gelfand and Smith (1990). The Metropolis-Hastings algorithm was developed by Metropolis, et al., (1953) and consequently generalized by Hastings (1970). A broad theoretical description of Metropolis-Hastings was given by (Tierney, 1994; Chib and Greenberg, 1995) provide outstanding discussions.

Using Metropolis Algorithm to construct a Markov chain with equilibrium distribution $\pi(x)$ for discrete case, Let $Q=\{q_{ij}\}$ be specified symmetric transition matrix and draw state s_j from i^{th} of row of Q. With known probability α_{ij} move from the state s_i to the state s_j, otherwise, remain at step s_i. The construction defines a Markov chain with transition matrix $p_{ij}=\alpha_j q_j \quad i\neq j$ and $p_{ij}=1-\sum_{j\neq i}p_j$. Metropolis et al (1953)

provided as given below,

$$\alpha_{ij}=\begin{cases}1, & if \quad \pi_j/\pi_i\geq 1\\ \pi_j/\pi_i, & if \quad \pi_j/\pi_i<1\end{cases} \qquad [6]$$

GIBBS SAMPLER ALGORITHM

The Gibbs sampler technique is one of the best known MCMC sampling algorithms in the Bayesian computational methods. The Gibbs sampler by Grenander (1983), the prescribed term is introduced by Geman and Geman (1984). Gibbs sampling is the landmark in problem of Bayesian inference (Gelfand and Smith, 1991). The Gibbs sampler tutorial is provided by Casella and George (1992).

Let $\theta=(\theta_1,\theta_2,...,\theta_p)'$ be a p-dimensional vector of parameters and let $\pi(\theta\,|\,D)$ be its posterior distribution given the data D. Then, the fundamental format of the Gibbs sampler is given as

Step 1. Select an arbitrary starting point

$$\theta_0=(\theta_{1,0},\theta_{2,0},...,\theta_{p,0})' \; and \quad set \quad i=0$$

Step 2. Generate $\theta_{i+1}=(\theta_{1,i+1},\theta_{2,i+1},...,\theta_{p,i+1})'$

Generate $\theta_{1,i+1}\sim\pi(\theta_1\,|\,\theta_{2,i},...,\theta_{p,i},D)$;

Generate $\theta_{2,i+1}\sim\pi(\theta_2\,|\,\theta_{1,i+1},\theta_{3,i},...,\theta_{p,i},D)$;

… … …

Generate $\theta_{p,i+1}\sim\pi(\theta_p\,|\,\theta_{1,i+1},\theta_{2,i+1},...,\theta_{p-1,i+1},D)$;

Step 3. Set $i=i+1$, and go to step 2

Each component of θ is in the natural order and a cycle in this scheme requires generation of p random variates. Gelfand and Smith (1990) show that under certain regularity conditions, the vector sequence $\{\theta_i,\ i=1,2,...\}$ has a stationary distribution $\pi(\theta\,|\,D)$. The performance of a Metropolis-Hastings algorithm depends on the choice of a proposal density q. The Metropolis-Hastings algorithm can be used within the Gibbs sampler when direct sampling from the full conditional posterior is difficult.

PRIOR

Prior elicitation perhaps plays the most crucial role in Bayesian inference. Survival analysis with covariates, the most popular choice of informative prior for β is the normal prior, and the most common choice of non informative prior for β is the uniform prior. The non-informative and improper priors may be useful and easier to specify for certain problems, but they cannot be used in all applications, such as model selection or model comparison, as it is well known that proper priors are required to compute Bayes factors and posterior model probabilities (Ibrahim, et al., 2004). Also non informative priors may cause instability in the posterior estimates and lead to convergence problems for the Gibbs sampler. Moreover, non informative prior do not make use of real prior information that one may use for a specific application.

APPLICATION

We consider the database consisting of 368 breast cancer women patients diagnosed at Cancer Institute (WIA), Chennai, India and follow-up period up to 180 months. The event of interest was time to death. Overall 187(51%) cases have experienced the event and 63% of 130 are of stage 3B cases.

The demographic and disease characteristics of the patients are given in table 1

Table 1: Classification of death according to Stages and Age group

	Stages			Age groups	
Status	Stage2B N (%)	Stage3A N (%)	Stage3B N (%)	Age <50 years N (%)	Age ³ 50 years N (%)
Alive	61 (55)	72 (56)	48 (37)	115 (53)	66 (44)
Dead	49 (45)	56 (44)	82 (63)	103 (47)	84 (56)
Total	110	128	130	218	150

From the table1, we see that death increases with the severity of stages and age. The event experienced cases among age group in more than 50 years is higher than the less than 50 years (Pari Dayal et al., 2013). The linear predictor is set equal to the intercept in the reference group (stage = 3); this defines the baseline hazard. The corresponding distribution of survival time is Gamma distribution (Cox and Oakes, 1984).

The Cox model with a frailty parameter for each individual is

Volume : 4 | Issue : 2 | Feb 2014 | ISSN - 2249-555X

using for identifying the risk variables for breast cancer patients. Here, the age and stages are considered as risk variables. We analyzed the data assuming a Weibull distribution for the survivor function, and including random effect (b_i) for each patient. The hazard model is as follows

$$t_i \sim Weibull(r, \lambda_i) \quad i = 1,...,368$$

$$\log\mu_i = \alpha + \beta_{age}\mathrm{AGE}_i + \beta_{stage1}\mathrm{STAGE}_{i1} + \beta_{stage2}\mathrm{STAGE}_{i2} + \beta_{stage3}\mathrm{STAGE}_{i3} + b_i$$

$$b_i \sim Normal(0, \tau)$$

where AGEi is a continuous covariate, and stage is a 3-level categorical covariate (stage2B = 1(as reference), stage 3A = 2 and stage 3B = 3) STAGE_{ik} $(k = 1,2,3)$ are dummy variables representing the 3-level factor for underlying stage. The regression coefficients and the precision of the random effects τ are given as ``non-informative'' priors, namely

$$b_i \sim \mathrm{Normal}(0, 0.0001)$$

$$\tau \sim \mathrm{Gamma}(0.0001, 0.0001)$$

The prior for the shape parameter of the survival distribution r is given as

$$r \sim \mathrm{Gamma}(1, 0.0001)$$

which is slowly decreasing on the positive real line.

In this analysis, we have used the BUGS program Spiegelhalter et al., (2003). This program performs based on the assumptions of Gibbs sampler by simulating from the full conditional distributions. The Bayesian estimators were obtained through the implementation of the Gibbs sampling scheme. It was executed 50,000 iterations of the algorithm and described the first 1000 iterations as a burn-in. The chains are used to check its convergence of the Gibbs sampler as recommended by the Spiegelhalter et al., (2004). Hence, convergence has been achieved, every 10,000 observations and are taken from each chain after the burn-in period. The summary (Table2) is showing posterior mean, median and standard deviation with a 95% posterior credible interval along with MC error, as well as the number of iterations as sample at the final after the burn-in period

The posterior distribution is provided using the density option in the Sample Monitor Tool which draws a kernel density estimate of the posterior distribution for a chosen parameter, as in Figure 1. There are various additional options for displaying the posterior distribution. They are quantiles, trace and history etc., like the survival curves.

Table 2: WinBUGS output for the Breast Cancer data: Posterior Statistics

Node	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
Alpha	-6.61800	0.36360	0.02555	-7.34900	-6.61300	-5.87600	1001	10000
beta.age	0.00608	0.00562	0.00030	-0.00524	0.00617	0.01675	1001	10000
beta.stage[2]	-0.06237	0.13370	0.00281	-0.32160	-0.06157	0.20020	1001	10000
beta.stage[3]	0.29290	0.13400	0.00308	0.03794	0.29110	0.56030	1001	10000
r	1.41400	0.05834	0.00381	1.29800	1.41400	1.52400	1001	10000
sigma	0.09608	0.04432	0.00408	0.02968	0.08937	0.19030	1001	10000
alpha	-6.62100	0.39690	0.02247	-7.43000	-6.61400	-5.86000	1001	20000
beta.age	0.00599	0.00573	0.00025	-0.00526	0.00595	0.01719	1001	20000
beta.stage[2]	-0.06236	0.13330	0.00175	-0.32130	-0.06207	0.20050	1001	20000
beta.stage[3]	0.29160	0.13380	0.00196	0.03577	0.29090	0.55700	1001	20000
r	1.41600	0.06222	0.00306	1.29600	1.41400	1.54300	1001	20000
sigma	0.08726	0.04589	0.00348	0.02510	0.07726	0.19410	1001	20000
alpha	-6.60900	0.40460	0.01605	-7.41200	-6.61000	-5.81500	1001	50000
beta.age	0.00578	0.00588	0.00018	-0.00584	0.00586	0.01717	1001	50000
beta.stage[2]	-0.06144	0.13270	0.00112	-0.31910	-0.06175	0.20030	1001	50000
beta.stage[3]	0.29180	0.13310	0.00129	0.03320	0.29190	0.55410	1001	50000
r	1.41500	0.06098	0.00208	1.29800	1.41300	1.53700	1001	50000
sigma	0.08788	0.04727	0.00268	0.02509	0.07813	0.20210	1001	50000

Table2 presents the posterior statistics for 50000 iterations in three spell of every 10000 with different nodes. The risk of the stage2 (3A) and stage3 (3B) are compared with the stage1 (2A) like ***exp(-0.06144) = 0.9404*** and ***exp(0.29180)=1.3388*** respectively. WinBUGS also implements the ***Deviance Information Criterion*** (DIC) (Spiegelhalter et al., 2002 & 2003) for model comparison criterion. This is a convenient information criterion measure that trades off goodness-of-fit against the complexity of a model. The DIC is computed as $DIC = \overline{D} + P_D = \hat{D} + 2P_D$. The Lowest value of the criterion indicates the better fitting models. $\overline{D}$ (Dbar) is the posterior mean of -2LL (log likelihood); $\hat{D}$ (Dhat) is the -2LogLikelihood at posterior mean of stochastic nodes.

Volume : 4 | Issue : 2 | Feb 2014 | ISSN - 2249-555X

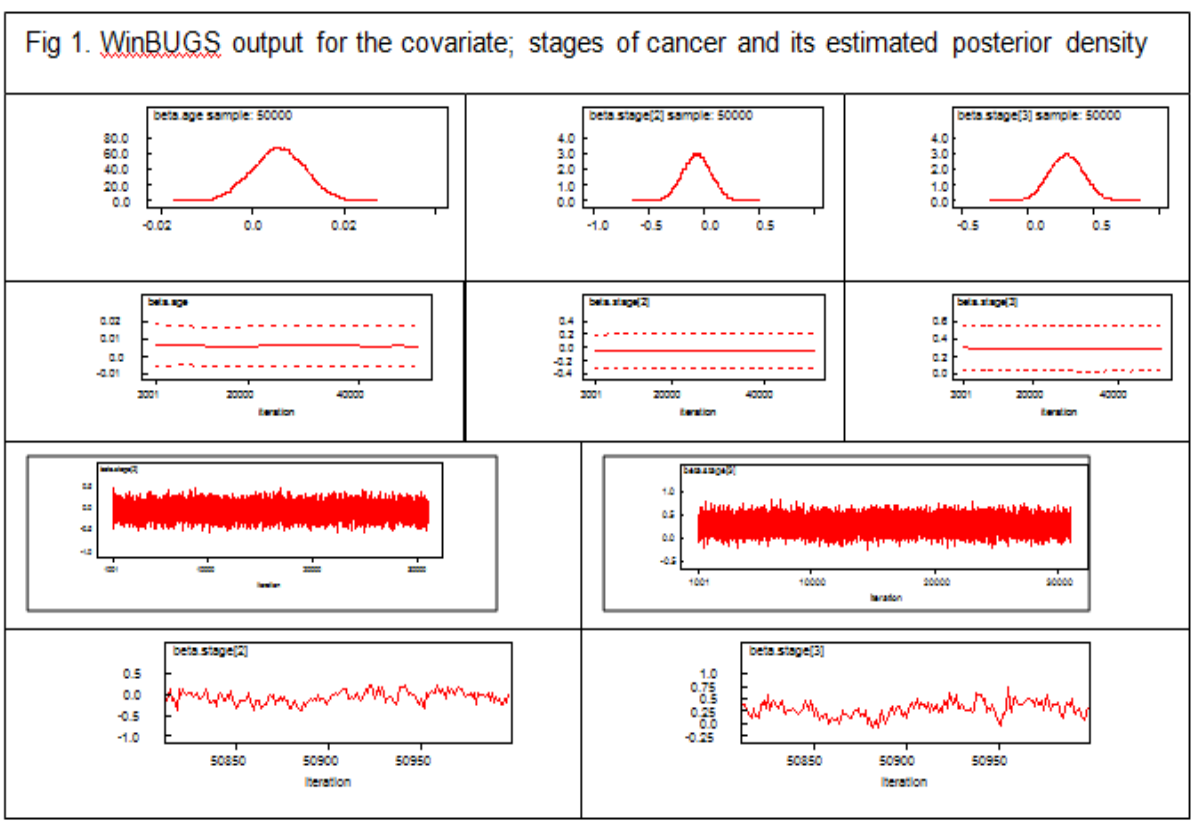
Table3: WinBUGS output for the Breast Cancer data Deviance Information Criterion (DIC)

	Dbar	Dhat	pD	DIC
Sample of 10,000 Iterations				
beta.stage	11.04800	11.04800	0.00000	11.04800
t	871.42000	862.09000	9.2700	880.75000
total	882.47000	873.14000	9.2700	891.80000
Sample of 20,000 Iterations				
beta.stage	11.04800	11.04800	0.00000	11.04800
t	871.81000	863.11000	8.70200	880.51000
total	882.86000	874.15000	8.70200	891.56000
Sample of 50,000 Iterations				
beta.stage	11.04800	11.04800	0.00000	11.04800
t	871.91000	863.6800	8.22400	880.10000
total	882.96000	874.7000	8.22400	891.18000

The DIC values for stage are illustrated in Table3 with different stages of iterations like 10000, 20000 and 50000 respectively. There are marginal changes in each stage of iterations. The lesser the DIC value will be considered as the better model. Since we have a simple model for this non-informative censored data, it is not required for model comparison. However, there is no reasonable change in the DIC values after 50000 iterations and in fact, it is increasing marginally.

There are some visual approximate estimates as confirmative measures such as posterior density or probability function, "trace" plots, posterior percentiles, quantiles etc. The figure1 demonstrates all types of visual approximate estimates. The first stage of the graphs is kernel density. The evolution for the median and the 2.5% and 97.5%percentiles for each iteration of the algorithm are obtained by using this quantiles plot, is in the second stage of the graph. The "trace" and "history" plots provide an on-line plot of the generated value as in the third and fourth stages of the figure1. The trace plot shows the full history of the samples for any parameter for which we have previously set a samples monitor and carried out the updates: The "trace" and history are related in several aspects. These plots are called "trace of beetles". In figure1, the chains for which convergence looks reasonable and the chains which have clearly reached convergence.

Fig 1. WinBUGS output for the covariate; stages of cancer and its estimated posterior density



Discussion

Bayesian frailty model proposed to fit flexible survival models for non-informative censored breast cancer data. Using WinBUGS software, we presented the comparable results as compared with the results of the seminal paper (Pari Dayal et al. 2013) with using the DIC value and the other supportive measures. The results which are presented in this paper followed the same trend and in fact it showed the reality. The prior for the nonparametric part in the Cox form of frailty model is taken place immensely. This method is easy to implement and allows a flexible class of survival models and it is also being seen as a simple alternative to the maximum likelihood method. WinBUGS is a tool for analyzing survival data in a Bayesian framework using MCMC. The results of the statistical analysis in all the tables and all visual approximate estimates which are presented in this paper are consistent. The DIC is an appropriate tool in the BUGS software package.

However, an enormous statistical knowledge is required for it to be used correctly. This approach provides an alternative validation that could be used to confirm results of 'frequentist' approach. Bayesian inference has a number of advantages over the frequentist approaches, mostly in the flexibility of model-building for time to event survival data. In addition, for many models, 'frequentist' inference can be obtained as a special case of Bayesian inference with the use of non-informative priors (Ibrahim et al., 2001). The Bayesian

Volume : 4 | Issue : 2 | Feb 2014 | ISSN - 2249-555X

approach enables us to formulate accurate inference based on the posterior distribution for any sample size, whereas the 'frequentist' approach relies heavily on the large sample approximation. The most important concern is that there is a risk involved in the erroneous usage of the Bayesian methods which could lead to improper data analysis

REFERENCE

1. Aalen O. O. (1994). Effects of frailty in Survival Analysis. Statistical Methods in Medical Research, 3, 227-43. | 2. Casella, G., and George, E.I. (1992). Explaining the Gibbs sampler. The American Statistician, 46, 167-74. | 3. Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," American Statistician, 49, 327–335. | 4. Clayton. D.(1991). A Monte Carlo for Bayesian inference in frailty models. Biometrics, 47, 467– 485. | 5. Cox, D.R. (1972). Regression model and life tables (with discussion). Journal of the Royal Statistical Society(B), 34, 187–220. | 6. Cox, D.R and Oakes, D. (1984), Analysis of Survival Data. London Chapman and Hall. | 7. Gelfand, A.E., and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. Journal of the Americal Statistical Association, 85, 398-409. | 8. Gelfand, A.E., and Smith, A.F.M. (1991). Gibbs sampling for marginal posterior expectations, Communications in Statistics, A, 20, 1747-66. | 9. Geman, S and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on pattern analysis and Matching Intelligence, 6, 721-41. | 10. Grenander, U. (1983). Tutorial in pattern theorey. Technical Report. Providence, R.I: Division of Applied Mathematics, Brown University. | 11. Hastings, W.K. (1970). Monte Carlo methods using Markov Chains and their applications. Biometrika, 57, 97-109. | 12. Ibrahim, G. J., Chen, M-H., and Sinha, D. (2001). Bayesian Survival Analysis. New York, Springer. | 13. Ibrahim, G. J., Chen, M-H., and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. Statistica Sinica, 14, 863-83. | 14. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. Journal of Chemical Physics, 21, 1087-91. | 15. Pari Dayal L, Leo Alexander T, Ponnuraja C, and Venkatesan P. Modelling of breast cancer survival data: A frailty model approach. Indian Journal of Applied Research, 2013, 3(10), 22-24 | 16. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," Journal of the Royal Statistical Society, Series B, 64(4), 583–616, with discussion. | 17. Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). Wiley. com. | 18. Spiegelhalter, D.J, Thomas A, Best N, and Lunn D, (2003), WinBUGS User Manual, Version 1.4, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK, 2003, available at: http://www.mrc-bsu.cam.ac.uk/bugs | 19. Tanner, M. A., and Wang, W . H (1987). The calculations of the posterior distributions by data augmentation, Journal of the American Statistical Association, 82, 528-40. | 20. Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," Annals of Statistics, 22, 1701–1762. | 21. Yashin, A. I., Vaupel, J. W. and Iachine, I.A. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data, Mathematical Population Studeis, 5, 145-159. | 22. Yin, G. and Ibrahim, G.J. (2005). A class of Bayesian shared gamma frailty models with multivariate failure time data. Biometrics, 61,208-16. |