Proportional Hazards Regression Model for Time to Event Breast Cancer Data: A Bayesian Approach

T. Leo Alexander
Associate Professor
Department of Statistics
Loyola College, Chennai-600034, India

Abstract

The paper deals with a Bayesian based Cox regression model to consider strategies for performing survivability of patients with breast cancer through Bayesian aspects. The proportional hazards model (PHM) in the context of survival data analysis, is same as Cox model and was introduced by Cox (1972) in order to estimate the effects of different covariates influencing the times-to-event data. It's well known that Bayesian analysis has the advantage in dealing with censored data and small sample over frequentist methods. Therefore, in this paper it deliberately explores the PHM for right-censored death times from Bayesian perspective, and compute the Bayesian estimator based on the Markov Chain Monte Carlo (MCMC) method. In particular it focuses on the approaches based on Gibbs sampler. Such approaches may be implemented using the publically available software BUGS. It aims to compare and apply Bayesian models of survivability for prediction of patients with breast cancer using outcome as explanatory variables and to produce better descriptions to survival of patients with breast cancer and of subgroups of patients with different survival characteristics.

Keywords: Proportional Hazards Model, Posterior Distributions, Markov Chain Monte Carlo, Gibbs Sampler, Winbugs

I. INTRODUCTION

Survival analysis aims to estimate the three survival (survivorship, density, and hazard) functions, denoted by S(t), f(t) and h(t), respectively (Collet, 1994). There exist parametric as well as non-parametric methods for this purpose (Kleinbaum and Klein, 2005). The survival function S(t) gives the probability of surviving beyond time t, and is the complement of the cumulative distribution function, F(t). The hazard function h(t) gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t (Kleinbaum and Klein, 2005).In survival analysis, one must consider censored data. This is a key issue for the analysis of survival data and one of the reasons why survival analysis is a special topic in statistics. In essence, censoring occurs when there is some information about individual survival time, but the survival time is unknown exactly. According to Miller (1998) and Hougaard (2000) data are said to be censored if the observation time censored survival is only partial, not until the failure event. There are many reasons for censoring. For examples, the patients can be lost to follow-up, patients still alive at the end of the study or patients drop out of the study. There are also several types of censoring, including right censoring, left censoring, interval censoring, random censoring, Type I censoring and Type II censoring (Collet, 1994, Hosmer and Lemeshow, 1999, Kalbfleisch and Prentice, 2002, Kleinbaum and Klein, 2005).Survival time data have two important special characteristics (Kleinbaum and Klein, 2005) as follows: Survival times are non-negative, and consequently are usually positively skewed. However, one can adopt a more satisfactory approach as an alternative distributional model for the original data. Typically, some subjects (as mentioned above) have censored survival times. There are nonparametric and parametric approaches to modelling survival data.

A Cox proportional hazards (PH) model is a popular mathematical model used for modelling survival. This model was proposed by Cox and Oakes (1972) and has also known as the Cox regression model. The reason why the Cox PH model is so popular because it is a semiparametric and a "robust" model. The results from this model will closely approximate the results of the correct parametric model (Kleinbaum and Klein, 2005).

A survival analysis typically examines the relationship of the survival distribution to covariates. If the risk of failure at a given time depends on the value of $x_1, x_2, \ldots, x$ of p predictor variables $X_1, X_2, \ldots, X_p$, then the value of these variables are assumed to have the time origin. If $h_0(t)$ is the hazard function for each object with the value of all predictor variable X is zero, then the function of $h_0(t)$ is the baseline hazard function (Collet, 1994). The Cox PH model is usually written in terms of the hazard model as follows:

$$h(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p Xp) \quad (1)$$

There are two basic assumptions of this model; the effect of X is linear and β is constant over time. The latter being the assumption of proportional hazards. It is also assumed that individuals are independent and homogeneous (Andersen, 1991).

II. BAYESIAN MODELLING

Modern Bayesian analysis began with a posthumous publication in 1763 by Reverend Thomas Bayes, set the theoretical framework and after a status of around 200 years, publications by Geman and Geman (1984) and Bernando and Smith (1994). The idea of Bayesian statistics within the context of life data analysis is to integrate prior knowledge, along with a given set of current observations, in order to make statistical inferences. The prior information could come from operational or observational data, from previous comparable experiments or from engineering knowledge (Gelman et al., 2004). This type of analysis can be mainly useful when there is limited test data for a given design. By integrating prior information about the parameters, a posterior distribution for the parameters can be obtained and inferences on the model parameters and their functions can be made. Suppose θ is some quantity that is unknown and let p(θ) denote the prior distribution of θ. Next, let y be some observed data, whose probability of occurrence is assumed to depend on θ. This dependence is formalized by p(y / θ), the conditional probability of y for each possible value of θ, and when considered as a function of θ is known as the likelihood (Spiegelhalter et al., 2004). The probability for different values of θ, taking account of y is denoted by p(θ / y).

Bayes' theorem applied to a general quantity says that:

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{p(y)}$$

Where $p(y) = \int p(\theta \mid y) d\theta$ and is considered as normalizing factor to ensure that $\int p(\theta \mid y) d\theta = 1$. So that, $p(\theta \mid y) \propto \int p(y \mid \theta)\, p(\theta)$.

In classical inference the sample data y are taken as random while population parameter θ, of dimension p, are taken as fixed. Often, a prior amounts to a form of modeling assumption or hypothesis about the nature of parameters, for example, random effect models. In many situations, existing knowledge may be difficult to summarize in the form of an 'informative prior' and to reflect such essentially prior ignorance, resort is made to non-informative priors are uniformly distributed between - ∞ and + ∞ and Jeffreys prior, $\pi(\theta) \propto \det\{I(\theta)\}^{0.5}$,

Where I(θ) is the expected information matrix. It is possible that a prior is improper. Such priors may add to identifiability problems (Gefland and Sahu, 1999), and so many studies prefer to adopt minimally informative priors, which are 'just proper'. The strategy is considered in terms of possible prior densities to adopt for the variance. The gamma process can be described, Let G(α,λ) denote the gamma distribution with shape parameter α>0 and scale parameter λ>0, where the density is given by

$$f(x \mid \alpha, \lambda) = \begin{cases} \{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}\} / \Gamma(\alpha) & , x > 0 \\ 0 & otherwise \end{cases}.$$

To reflect prior ignorance while avoiding impropriety, Spiegelhalter et al (1996) suggesting a prior standard deviation at least an order of magnitude greater than the posterior standard deviation. Prior distributions play a very important role in Bayesian statistics. There are two different types of prior distributions; informative and non-informative (Gelman et al., 2004, Marin and Robert, 2007, Spiegelhalter et al., 2004). Non-informative prior distributions (vague, flat, and diffuse) play a minimal role in posterior inference, and posterior can be sensitive to prior (Gelman et al., 2004). The non-informative prior distributions can be used to make inferences that are not greatly affected by external information or when external information is not available Gelman et al., 2004, Marin and Robert, 2007, Spiegelhalter et al., 2004).According to the Bayesian rule, one can express posterior probability of certain event H given some data with the formula

$$P(H \mid data) = \frac{P(data \mid H)}{P(data)}.$$

The probability of H given the data is called the posterior probability of H. The posterior equals to the likelihood time the prior divided by marginal probability of data.

In classical approaches such as maximum likelihood, inference is based on the likelihood of the data alone. Lindley (1968) used noninformative priors and thus, for the linear case, the posterior means are the MLEs. The posterior mean derived above for the proportional hazards model is a weighted sum of the maximum partial likelihood estimators and the prior means. In Bayesian models, the likelihood of the observed data y given parameters θ, denoted f(y|θ) is used to modify the priors beliefs $\pi(\theta)$, with the updated knowledge summarized in a posterior density, $\pi(\theta \mid y)$. Thus $f(y,\theta) = f(y \mid \theta)\pi(\theta) = \pi(\theta \mid y) m(y)$ and therefore the posterior density can be written $\pi(\theta \mid y) = f(y \mid \theta)\pi(\theta) / m(y)$, where m(y) is known as the marginal likelihood of the data. This quantity plays a central role in some approaches to Bayesian model choice, but for the present purpose can be seen as a proportionality factor, such as $\pi(\theta \mid y) \propto f(y \mid \theta)\pi(\theta)$.

The Gibbs sampler is a Monte Carlo method for approximating joint and marginal distributions by sampling from conditional distributions. This method is well discussed by *Casella and George (1992), Gelfand and Smith (1990) and Geman and Geman (1984),* among others. The Gibbs sampler uses only full conditional densities in approximating joint and marginal densities. The Gibbs sampler is a way to generate empirical distributions of two variables from a model. Gibbs sampling is a special case of the Metropolis-Hastings algorithm, and thus an example of a Markov chain Monte Carlo.

The Gibbs sampler introduced by Geman and Geman (1984) on MCMC algorithm is known as Gibbs sampling and involves successive sampling from a complete conditional densities which conditions on both the data and the other parameters. (Applied Bayesian Modelling P.Congdon Ch-1).

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

1) *Step 1: Select an arbitrary starting point*

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

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

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

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

$$.....$$

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

3) *Step 3: Set i=+1, and go to step 2*

Such successive samples may involve simple sampling from standard densities (gamma, Normal…). An alternative schemes based on the Metropolis-Hastings algorithm, may be used for non-standard densities (*Mergan, 2000*). The program WINBUGS may be applied with some or all parameters sampled from formally coded conditional densities; provided with prior and likelihood WINBUGS will infer the correct conditional densities.

III. COMPUTATION OF BAYESIAN APPROACH

Bayesian computation generally exploits modern computer power to carry out simulations (Spiegelhalter et al., 2004) based on Markov chains and is known as Markov chain Monte Carlo (MCMC). Monte Carlo methods are techniques that have the aim of evaluating integrals rather than exact or approximate algebraic analysis (Spiegelhalter et al., 2004). Several MCMC algorithms that are commonly used are Gibbs sampling, Metropolis-Hastings, reversible jump, slice sampling, particle filters, perfect sampling and adaptive rejection sampling (Marin and Robert, 2007, Spiegelhalter et al., 2004).

A. Gibbs Sampler

Survival model can be conveniently inspected with the help of hazard function. A common approach to handling the prior probability for the baseline hazard function in PHM is a Gamma process prior. However, this can lead to biased and misleading results (Spiegelhalter et al., 1996). The Gibbs sampling is proposed to simulate the Markov chain of parameters' posterior distribution dynamically, which avoids the calculation of complex integrals of the posterior using WinBUGS package.

Gibbs variable selection was defined by Dellaportas et al. (2000, 2002) using the BUGS software (Spiegelhalter et al., 1996a,b,c). The specification of the likelihood, prior and pseudo-prior distributions of the parameters as well as the prior term and model probabilities are described. The prior is

$$f(\beta_i \mid \gamma_i) = \gamma_i N(0, \Sigma_i c_i) + (1 - \gamma_i) N(\bar{\mu}_i, S_i) \cdot$$

The prior where $f(\beta_i \mid \gamma) = f(\beta_i \mid \gamma_i)$ potentially makes the method less efficient.

The full conditional posterior distribution is

$$f(\beta_i \mid \gamma, \beta_{,i}, y) \propto \begin{cases} f(y / \gamma, \beta) N(0, \Sigma_i c_i), \gamma_i = 1 \\ V(\bar{\mu}_i, S_i) \quad , \gamma_i = 0 \end{cases}$$

Bayesian inference has been discussed by several authors for censored survival data where the integrated baseline hazard function is to be estimated non-parametrically Kalbfleisch (1978),Kalbfleisch and Prentice (1980), Clayton (1991), Clayton (1994).Clayton (1994) formulates the Cox model using the counting process notation introduced by Andersen and Gill (1982) and discusses estimation of the baseline hazard and regression parameters using MCMC and Gibbs sampler methods. Although these approaches may appear somewhat fixed, it forms the basis for extensions to random effect (frailty) models, time-dependent covariates, smoothed hazards, multiple events and so on. The Cox model in BUGS formulation is implemented as below.

For subjects i = 1,...,n, we observe processes $N_i(t)$ which count the number of failures which have occurred up to time t. The corresponding intensity process $I_i(t)$ is given by $l_i(t)dt = E(dN_i(t) \mid F_{t-})$,

Where $dN_i(t)$ is the increment of N_i over the small time interval [t, t+dt), and F_{t-} represents the available data just before time t. If subject i is observed to fail during this time interval, $dN_i(t)$ will take the value 1; otherwise $dN_i(t) = 0$. Hence $E(dN_i(t) \mid F_{t-})$ corresponds to the probability of subject i failing in the interval [t, t+dt). As dt tends to zero (assuming time to be continuous) then this probability becomes the instantaneous hazard at time t for subject i. This is assumed to have the proportional hazards form

$$l_i(t) = Y_i(t) \lambda_0(t) \exp(\beta Z_i),$$

Where $Y_i(t)$ is an observed process taking the value 1 or 0 according to whether or not subject i is observed at time t and $\lambda_0(t) \exp(\beta z_i)$ is the familiar Cox regression model. Thus we have observed data $D = N_i(t), Y_i(t), z_i$; i = 1,..n and unknown parameters

β and $\Lambda_0(t)$ = Integral($\lambda_0(u)$, u, t, 0), the latter to be estimated non-parametrically. The joint posterior distribution for the above model is defined by

$$P(\beta, \Lambda_0{}^0 \mid D) \sim P(D \mid \beta, \Lambda_0()) P(\beta) P(\Lambda_0()).$$

For BUGS, we need to specify the form of the likelihood $P(D|\boldsymbol{\beta},\Lambda_0())$ and prior distributions for **β** and $\Lambda_0()$. Under non-informative censoring, the likelihood of the data is proportional to

$$\prod_{i=1}^{n}[\prod_{t>=0} Ii(t)dNi(t)] \exp(-\int Ii(t)dt)$$

This is essentially as if the counting process increments $dN_i(t)$ in the time interval [t, t+dt) are independent random variables with means $I_i(t)d$:

$$dN_i(t) \sim \text{Poisson}(l_i(t)dt) \cdot$$

We may write

$$l_i(t)dt = Y_i(t)\exp(\beta Z_i)d\Lambda_o(t),$$

Where $d\Lambda_0(t) = \Lambda_0(t)dt$ is the increment or jump in the integrated baseline hazard function occurring during the time interval [t, t+dt). Since the conjugate prior for the Poisson mean is the gamma distribution, it would be convenient if $\Lambda_0()$ were a process in which the increments $d\Lambda_0(t)$ are distributed according to gamma distributions. We assume the conjugate independent increments prior suggested by Kalbfleisch (1978), namely

$$dL_0(t) \sim \text{Gamma}(cd\Lambda^0{}_0(t), c).$$

Here, $dL_0{}^0(t)$ can be thought of as a prior guess at the unknown hazard function, with c representing the degree of confidence in this guess. Small values of c correspond to weak prior beliefs. In the example below, we set $d\Lambda^0{}_0(t) = r\,dt$ where r is a guess at the failure rate per unit time, and dt is the size of the time interval.

IV. APPLICATION TO BREAST CANCER DATA

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 event experienced cases among age group in more than 50 years is higher than the less than 50 years (Pari Dayal et al., 2013, Leo Alexander et al., 2014). The linear predictor is set ,equal to the intercept in the reference group (stage = 3)[from the database], this defines the baseline hazard.

In this analysis, BUGS program have been used, 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 for every 10,000 observations and is taken from each chain after the burn-in period. The summary (Table1) 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 - 1

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*

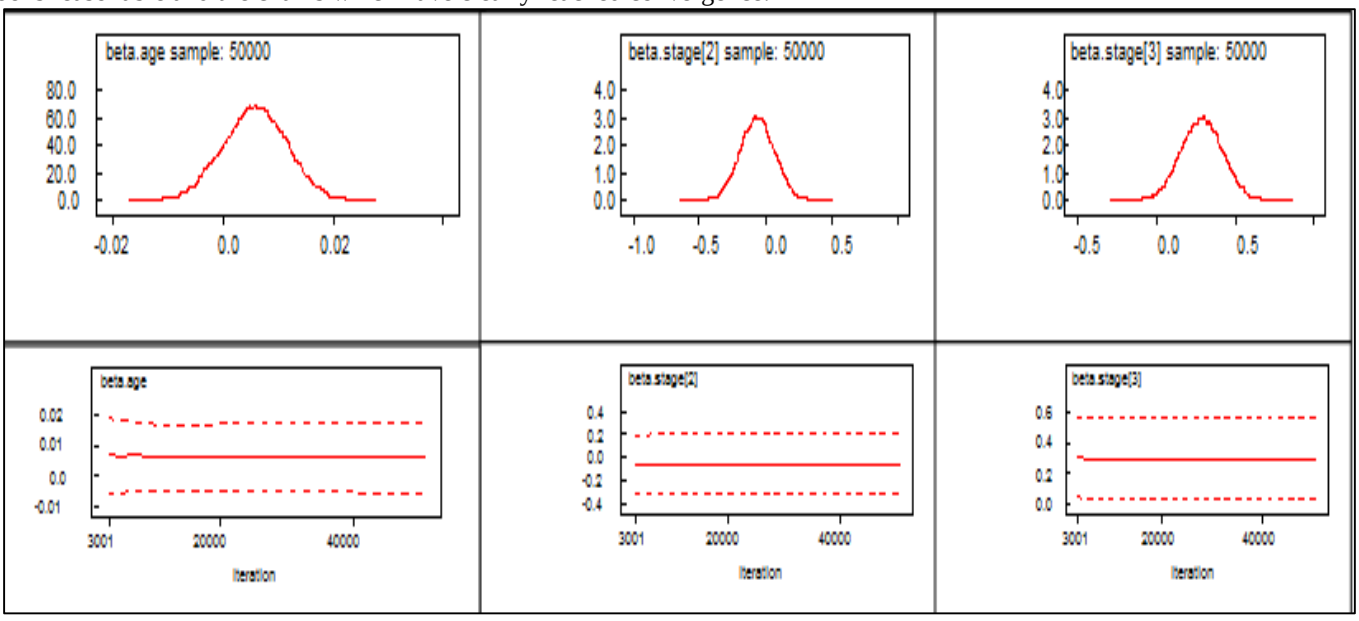
Table1 presents the posterior statistics for 50000 iterations in three spell of every 10000 with different nodes. The risk of the stage2 (3A) and stage3 (3B) are compared with the stage1 (2A) like exp(-0.06144) = 0.9404 and exp(0.29180)=1.3388 respectively. WinBUGS as well implements the *Deviance Information 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.

Table - 2
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 Table2 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.



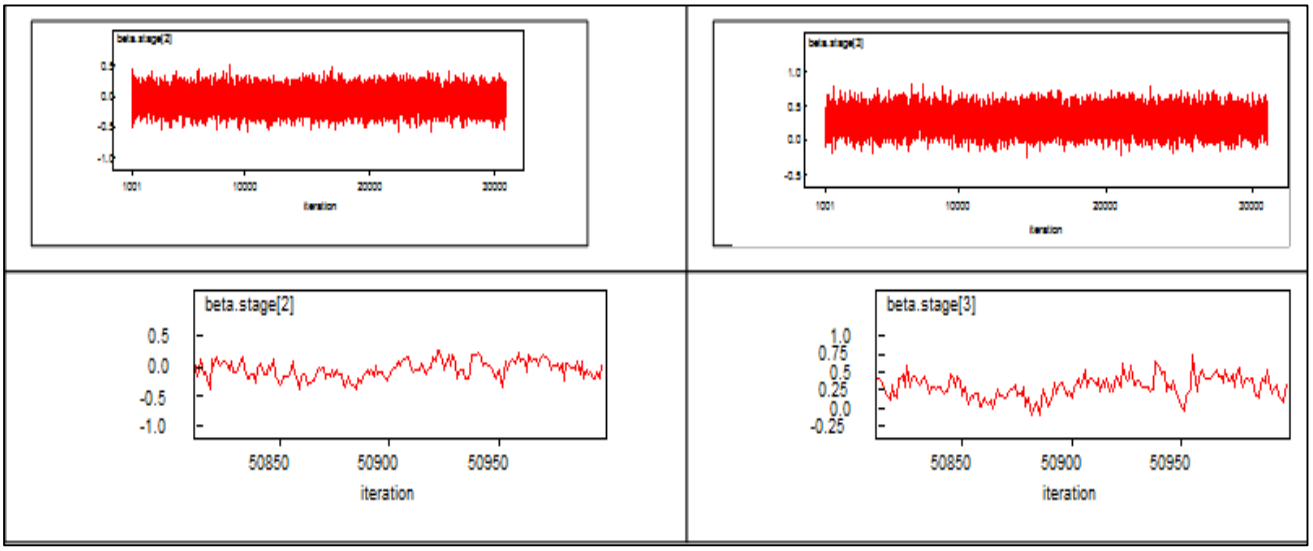


Fig. 1: Chains for which convergence looks reasonable and the chains which have clearly reached convergence

The proposed Bayesian method was used to fit the survival models for non-informative right censored breast cancer data. The results which are presented in this paper followed the trend and in fact it showed the reality. The statistical analysis of hazard function is assumed for the time to death. Using WinBUGS software (Lunn et al. 2000, 2009), the computational problem become easily and we proved that with baseline hazard, the DIC value is smaller than the Gamma process prior for the nonparametric part in Cox model. WinBUGS is a tool for analyzing survival data in a Bayesian framework using Markov Chain Monte Carlo (MCMC) with Gibbs sampler. Using DIC (deviance information criterion); it is suited to compare Bayesian models whose posterior distributions that have been obtained using MCMC. DIC has been implemented as a tool in the BUGS software package. However, much technical statistical knowledge is required for it to be used correctly. These programs provide an alternative platform that could be used to confirm results of 'frequentist' software. Moreover, 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 approach enables us to make exact inference based on the posterior distribution for any sample size especially when sample size is too small, whereas the 'frequentist' approach relies heavily on the large sample approximation, and there is always the issue of whether the sample size is large enough for the approximation to be valid (Ibrahim et al., 2001). There is a danger that the additional complexity of Bayesian methods could lead to improper data analysis if it is not used correctly or choosing inappropriate prior. Although, this article proposed the prior elicitation for the baseline hazard in the Cox model, it is straight forward to extend to additive hazard model with simple modification as suggested by Aalen et al, (2009).

REFERENCES

[1] Andersen P.K. Survival analysis 1982-1991: The second decade of the proportional hazards regression model. Statistics in Medicine, 10:1931–1941,1991.
[2] Andersen P.K, and R.D. Gill,.Cox's regression model for counting processes: A large sample study. Annals of Statistics, 10 (1982) , 1100–1120
[3] Aalen.O.O, Andersen P.K., Borgan, R.D.Gill, N. Keiding,. History of applications of martingales in survival analysis. Electronic Journal for History of Probability and Statist. 5, 1 (2009), 1–28.
[4] Besag.J, Green. E, Higdon D., and Mengersen K. Bayesian computation and stochastic systems. Statistical Science, 10(1):3–41, 1995.
[5] Chen M. H., Ibrahim J. G. and Sinha D. A new Bayesian model for survival data with a surviving fraction. Journal of American Statistical Association, 94:909–919, 1999.
[6] Chen W. C., Hill B. M., Greenhouse J. B. and Fayos J. V. Bayesian analysis of survival curves for cancer patients following treatment. In: J. M. Bernardo, M. H. Degroot, D. V. Lindley and A. F. M. Smith. Bayesian Statistics 2, pages 299–328, 1985.
[7] Chib S. and Greenberg E. Understanding the Metropolis Hastings algorithm. American Statistician, 4:327–335, 1995.
[8] Collet D. Modelling Survival Data in Medical Research. Chapman and Hall, 1st edition, 1994.
[9] Clayton D. A Monte Carlo for Bayesian inference in frailty models. Biometrics, 47, (1991), 467– 485.
[10] Clayton D., Bayesian analysis of frailty models. Technical Report, Medical Research Council Biostatistics Unit, Cambridge, 1994.
[11] Cox D.R., Regression models and life tables (with discussion), Journal of the Royal Statistical Society, Series B, 34 (1972), 187-220.
[12] Cox, D.R. (1975). Partial likelihood. Biometrika, 62, 269–276
[13] Congdon P. Applied Bayesian Modeling. John Wiley & Sons, 2003.
[14] Congdon P. Bayesian Statistical Modelling. JohnWiley & Sons, 2nd edition, 2006.
[15] Gelman, A. and Rubin D. B., Inference from iterative simulation using multiple sequences, Statistical Science, 7 (1992), 457–511.
[16] Geman S., and Geman D., Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6 (1984), 721–741
[17] Hastings W. K.. Monte carlo sampling methods using Markov Chains and their applications. Biometrika, 1:97–109, 1970.
[18] Hosmer D. W. and S. Lemeshow. Applied Survival Analysis: Regression Modeling of Time to Event Data. John Wiley and Sons, 1999.
[19] Hougaard P.. Frailty models for survival data. Lifetime Data Analysis, 1(3):255–273, 1995.
[20] Hougaard P.. Analysis of Multivariate Survival Data. Springer-Verlag, 2000
[21] Ibrahim J. G., Chen M., and Sinha D., Bayesian Survival Analysis. Springer –Verlag, New York, 2001.
[22] Ibrahim J. G., Chen M. H., and Sinha D. Bayesian semiparametric models for survival data with a cure fraction. Biometrics, 57:383–388, 2001a.

[23] Ibrahim J. G., Chen M. H., and Sinha D.. Bayesian Survival Analysis. Springer, 2001
[24] Kalbfleisch J. D. and Prentice R.L., The statistical analysis of failure time data. Wiley, New York, 1980.
[25] Kalbfleisch J. D., Nonparametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society, Series B 40 (1978) , 214 – 221
[26] Kalbfleisch J. D. and Prentice R. L.. The Statistical Analysis of Failure Time Data. Wiley & Sons, New Jersey, 2nd edition, 2002.
[27] Kaminskiy M. P. and Krivtsov V. V. A simple procedure for Bayesian estimation of the Weibull distribution. IEEE Transactions On Reliability, 54(4):612–616, 2005.
[28] Kaplan E. L. and Meier P. Nonparametric estimation from incomplete observations. Journal of American Statistical Association, 53:457–481, 1958.
[29] Kenney J. F. and Keeping E. S.. Root mean square. In J.F. Kenney and E.S. Keeping (eds.). Mathematics of Statistics (3rd ed., pp. 59-60), 1962.
[30] Kim S., Chen M. H, Dey D. K., and Gamerman. Bayesian dynamic models for survival data with a cure fraction. Lifetime Data Analysis, 13:17–35, 2007.
[31] Kim S. and Ibrahim J. G.. On Bayesian inference for proportional hazards models using non-informative priors. Life Time Data Analysis, 6(4):331–341, 2000.
[32] Kleinbaum D. G. and Klein M.. Survival analysis: A Self-learning Text. Springer, 2005.
[33] Lawless J. F.. Statistical Models and Methods for Lifetime Data. Wiley & Sons, New Jersey, 2nd edition, 2002.
[34] Leo Alexander T, Pari Dayal L, Ponnuraja C, Venkatesan P. Bayesian Cox Model with Categorical Predictors for Time to Event Breast Cancer Data. IJAR, 2014:4(8), 497-501
[35] Leo Alexander T. A Study on Madras Metropolitan Tumour Registry Breast Cancer patients in Chennai, International Journal of Innovative Research in Technology, Volume:03, Issuse:02, July 2016, ISSN : 2349-6002.
[36] Liu X.. Survival Analysis Models and Applications. Wiley & Sons, New Jersey, 2nd edition, 2012.
[37] Lunn D.J., Thomas A., Best N., Spiegelhalter D., WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10 (2000) , 325 – 337.
[38] Lunn D. J., Spiegelhalter D., Thomas A., and Best N. The BUGS project: Evolution, critique and future directions. Statistics in Medicine, 28 (2009), 3049 – 3067
[39] 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.
[40] Silva G. L., and Amaral-Turkman M. A., Bayesian Analysis of an Additive Survival Model with Frailty. Communications in Statistics – Theory and Methods, 33 (10) (2004), 2517–2533
[41] Spiegelhalter D., Thomas A., Best N., and Gilks W., BUGS 0.5: Examples Volume 1, MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, 1996.
[42] 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.
[43] Spiegelhalter D., 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
[44] Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). Wiley. com.
[45] Tsiatis A. A., A large sample study of Cox's regression model. Annals of Statistics, 9 (1981). 93 – 108.