# Malaysian Journal of Fundamental & Applied Sciences

### available online at http://mjfas.ibnusina.utm.my

# Estimating Proportional Hazards Model Using Frequentist and Bayesian Approaches

Noraslinda Mohamed Ismail*, Zarina Mohd Khalid and Norhaiza Ahmad

*Department of Mathematical Sciences, Faculty of Science, UTM, 81310 UTM Skudai, Johor, Malaysia*

**ABSTRACT**

In statistics, the proportional hazards model (PHM) is one of a class of survival models. This model estimates the effects of different covariates influencing the time-to-event data in which the hazard function has been assumed to be the product of the baseline hazard function and a non-negative function of covariates. In this study, we investigate the hazard function, also known as the risk function or intensity function, which is employed in modelling the survival data and waiting times. The model parameters can be estimated via frequentist or Bayesian approach. However, the Bayesian approach is well known to have the advantages over frequentist methods when the data are small in size and involve censored individuals. In this paper, the PHM for right-censored data from Bayesian perspective will be discussed and the Markov Chain Monte Carlo (MCMC) method will be used to estimate the posterior distributions of the model parameters using Leukemia data.

| Proportional hazards model | Hazard function | Bayesian | MCMC |

## 1.    INTRODUCTION

Modelling and survival data analysis is one of the oldest fields in statistics. Generally, survival data deals with lifetimes from some initial event at time zero to some terminal event of interest and basically the data would be an independent non-negative random variable, say *T*. A major advancement in survival analysis took place in the era of 1950's where Kaplan and Meier (1958) proposed their famous estimator of the survival curve [1]. Later in 1972, David Cox [2] introduced the proportional hazards model which incorporates covariates.

An important issue that always arises when studying time to event data is that some individuals are still alive or survive at the end of the study. This implies that the event of interest has not occurred and therefore results in the emergence of right censored data. The censoring process is called non-informative if the risk set at any time point for those who are still alive and uncensored, should be representative of the entire population within the same cohort. Non-informative censoring automatically occurs if the censoring process is independent of the survival time.

General classes of semi-parametric hazards regression model for survival data which include Cox proportional hazards model, the accelerated failure time model and the accelerated hazards model have been proposed by Chen and Jewell [3]. Their new models are flexible and may yield more accurate prediction of an individual's survival process and a covariate's effect can beidentified by separating two components namely a time

scale change on hazard progression and a relative hazards ratio.

Bayesian approach might be a convenient approximation and simplification to bypass the prior distribution of the nuisance parameters since the maximum likelihood method as a general technique might fail when there is a presence of many nuisance parameters.

Recently, semi-parametric Bayesian analyses of proportional hazards model has become computationally feasible due to modern technology and advancement in computing techniques such as the Gibbs sampler and other MCMC methods. Arjas and Gasbarra [4] considered simple right censored data with a common unknown hazard rate in which the hazard rate is modelled nonparametrically. In their study, the sample paths of the hazard rate were generated from the posterior distribution using Gibbs Sampler.

Sinha and Dey [5] provided an overview of Bayesian semiparametric methods for the Cox model. One of the advantages of using Bayesian method is that we can join the baseline hazard and the regression coefficients that can be used accurately to compute the target posterior quantities using MCMC simulation techniques. They investigated the potential of Bayes methods for the analysis of survival data using semiparametric models based on either the hazard or the intensity function. The nonparametric part of every model is assumed to be a realization of a stochastic process. The parametric part, which may include a regression parameter or a parameter quantifying the heterogeneity of a population, is assumed to have a prior distribution with possibly unknown hyperparameters.

*Corresponding author at:
E-mail addresses: noraslinda@utm.my (Noraslinda Mohamed Ismail)

Ibrahim and Sinha [6] developed a class of semi-parametric informative prior distributions for the Cox model. They specified a non-parametric prior for the baseline hazard rate and a parametric prior for the regression coefficients via the development of novel Markov chain Monte Carlo (MCMC) techniques for sampling from the posterior distribution of the parameters. Their approach seems to be a useful approach for this problem since it is difficult to specify meaningful prior distributions for the parameters in each model task and requiring contextual interpretations of a large number of parameters.

In this article, we have used the proportional hazards model that has been used extensively since 1972. In a fully parametric model, the lifetime distribution has been assumed to belong to a family of parametric distributions and reducing the regression problem to estimating the parameters from the data.

Proportional hazards model can be modelled from classical perspective by obtaining the partial likelihood approach in estimating the unknown parameters. Recently, the model has been the most common from a Bayesian perspective. It has been widely used in survival analysis for such realistic models. Fully Bayesian computations of multi-level or hierarchical model are now possible using simulation techniques. Gibbs Sampling is one of the new numerical algorithms which allow obtaining samples from posterior of interest and this new development has motivated the use of Bayesian methods in survival analysis. Gibbs Sampler is one of the best known Markov Chain Monte Carlo (MCMC) sampling in computational literature. Many authors appeared explaining in their papers the MCMC-based Bayesian analysis as this task would typically require significantly scientific programming ability and facility with such random number generators. Specialized software packages called BUGS [7] are created for implementing MCMC-based analyses of full probability models. These packages will treat all unknowns as random variables. This paper describes the use of freely available software for the analysis of complex statistical models using MCMC techniques, called WinBUGS [8] and R programming, in the context of PHM.

The proposed baseline hazards function in Cox model is slightly different compared to the original Cox model which has been proposed by Kalbfleisch [9]. He proposed gamma process prior as the baseline hazard function in Cox model and he fixed the value of $c$ (a specification of the weight attached to the guess) and $r$ (guess at the failure rate per unit time). In this paper, we proposed both $c$ and $r$ to have their own distributions and to make the proposed model more flexible, we assume $c$ and $r$ to have gamma and uniform distributions, respectively.

## 2. PROPORTIONAL HAZARDS MODEL

### 2.1 The Weibull Proportional Hazards Model

A parametric survival model is one in which the survival time is assumed to follow a known distribution. Examples of distributions that are commonly used for survival time are the Weibull distribution, the Exponential distribution, the Log-logistic distribution, the Lognormal distribution and the generalized gamma distribution.

In parametric survival analysis, all parts of the model are specified for both the hazard function and the effect of any covariates. The estimation is easier and estimated survival curves are smoother as they draw information from the whole data. In addition, it is possible to do more sophisticated analyses with parametric models, such as introducing random effects or using Bayesian methodology to pool sources of information.

In parametric survival model, time $T$ is assumed to follow some distribution whose probability density function, $f(t)$ can be expressed in terms of unknown parameters. The corresponding survival and hazard function can be determined once a probability density function is specified for survival time.

The Weibull model is the most widely used parametric survival model. Survival time $T$ has a Weibull distribution of $W(t|\lambda, \alpha)$, with its survival function is given by

$$S(t|\lambda, \alpha) = \exp(-\lambda t^\alpha) \quad (2.1)$$

and the probability density function is

$$f(t|\lambda, \alpha) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha) \quad (2.2)$$

where $\lambda$ is the scale parameter, $\alpha$ is the shape parameter, and both $\lambda$ and $\alpha$ are positive ($\lambda > 0, \alpha > 0$). The important feature of Weibull distribution is the failure rate. The failure rate of a subject is decreasing when $\alpha < 1$, it is increasing when $\alpha > 1$ and it is constant when $\alpha = 1$, which indicates an exponential distribution. Therefore the hazard rate is defined as

$$h(t|\lambda, \alpha) = \lambda \alpha t^{\alpha-1} \quad (2.3)$$

The cumulative distribution is

$$F(t|\lambda, \alpha) = 1 - \exp(-\lambda t^\alpha) \quad (2.4)$$

The likelihood function in survival analysis is more complicated. The likelihood function is only related to the density function if there is no censored data; else the likelihood depends on the density function and the survival function. The likelihood function for Weibull distributed data, $t_i, i = 1,2,\ldots,n$ with random censoring $\delta_i, i = 1,2,\ldots,n$ is

$$L(t|\lambda, \alpha) = \prod_{i=1}^{n} [\lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)]^{\delta_i} [\exp(-\lambda t^\alpha)]^{1-\delta_i} \quad (2.5)$$

and the log-likelihood function is

$$l(t|\lambda, \alpha) = \sum_{i=1}^{n} \big[\delta_i \log(\lambda \alpha t^{\alpha-1}\exp(-\lambda t^\alpha)) + (1-\delta_i)\log(\exp(-\lambda t^\alpha))\big] \quad (2.6)$$

The Weibull hazard function is defined as $h(t|\lambda, \alpha) = \lambda \alpha t^{\alpha-1}$ while a Weibull PH model is then be defined by reparameterizing $\lambda$ as

$$\exp\left(\sum_{i=1}^{p} \beta_i X_i\right) = \exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right)$$

## 2.2 The Cox Proportional Hazards Model

The Cox proportional hazards (PH) model is known as a semi-parametric model because the form of the baseline hazard is not specified, only the form of the effect of covariates. The Cox proportional hazards (PH) model can be written in terms of the hazard model formula shown below:

$$h(t, \mathbf{X}) = h_0(t)\exp(\boldsymbol{\beta}^T \mathbf{X}) \qquad (2.7)$$

This model gives an expression for the hazard at time *t* for an individual with a given specification of a set of explanatory variables denoted by **X** where $\mathbf{X} = (X_1, X_2, \ldots, X_p)$.

An important feature of this formula is that the baseline hazard is a function of *t*, which concerns the proportional hazards (PH) assumption and does not involve the $X's$. It is shown that the exponential expression in contrast involves the $X's$ but does not involve *t*. The $X's$ here are called time-independent variables and are defined to be any variable whose value for a given individual does not change over time.

The Cox PH model has the property that if all of the $X's$ are zero then the formula will reduce to the baseline hazard function. This property of the Cox PH model is the reason why $h_0(t)$ is called the baseline function. In other words, the Cox PH model will reduce to the baseline hazard function when there are no $X's$ in the model.

The survival function, the likelihood function and the log likelihood function for the Cox PH model are as follows:

$$S(t, \mathbf{X}) = S_0(t, \mathbf{X})^{\exp(\boldsymbol{\beta}^T X)} \qquad (2.8)$$

$$L(t|\boldsymbol{\beta}, \mathbf{X}) = \prod_{i=1}^{n} h_0(t_i)^{\delta_i} \exp(\boldsymbol{\beta}^T X)^{\delta_i} S_0(t_i)^{\exp(\boldsymbol{\beta}^T X)}$$

$$(2.9)$$

$$l(t|\boldsymbol{\beta}, \mathbf{X}) = \sum_{i=1}^{n} (\delta_i \log\{h_0(t_i)\} + \delta_i(\boldsymbol{\beta}^T X)$$
$$+ \exp(\boldsymbol{\beta}^T X)\log\{S_0(t_i)\})$$

$$(2.10)$$

## 3. ESTIMATING THE PARAMETERS

## 3.1 Maximum Likelihood Estimation (MLE)

The idea behind maximum likelihood parameter estimation is to determine the parameters that maximize the probability (likelihood) of the sample data. From a statistical point of view, the method of maximum likelihood is considered to be more robust (with some exceptions) and yields estimators with good statistical properties. In other words, MLE methods are more versatile and can be applied to most models and to different types of data. In addition, they provide efficient methods for quantifying uncertainty

through confidence bounds. Although the methodology for maximum likelihood estimation is simple, the implementation is mathematically intense. Using today's computer power, however, mathematical complexity is not a big obstacle. The MLE methodology is presented next.

The maximum likelihood estimator (MLE) for the proportional hazards model with partly interval-censored data was proposed by Kim [11]. The MLEs of the regression parameter and the cumulative hazard function shows consistency and asymptotically normal under an appropriate regularity conditions.

The Newton Raphson method is one of the popular methods in maximizing the likelihood function and estimating the unknown parameters. The Newton-Raphson method is deterministic because there is no element of randomness in optimisation. It is also an iterative procedure because it consists of a series of iterations, with an improved estimation in each of the iteration.

Generally, if we have a parameter vector $\boldsymbol{\beta}$ of dimension *p* and wish to estimate $\hat{\boldsymbol{\beta}}$ which maximize the likelihood function $l(\boldsymbol{\beta})$, the algorithm is as follows:

a) Let $k = 0$
b) Arbitrarily choose $\boldsymbol{\beta}^{(k)}$
c) Solve

$$I(\boldsymbol{\beta}^{(k)})(\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}) = U(\boldsymbol{\beta}^{(k)})$$

for $\boldsymbol{\beta}^{(k+1)}$
d) Increment *k* by 1
e) Go back to step (c) and repeat until converge.

Some notations have been used are:

- $\boldsymbol{\beta}^{(k)}$ is the value of the parameters at the $k^{th}$ iteration of the routine
- the score function is

$$U(\boldsymbol{\beta}) = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1}, \quad \cdots, \quad \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p}\right)$$

- the information matrix is

$$I(\boldsymbol{\beta}) = -\begin{pmatrix} \dfrac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_1^2} & \cdots & \dfrac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_p} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_p \partial \beta_1} & \cdots & \dfrac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_p^2} \end{pmatrix}$$

- and $\beta_q$ is the $q^{th}$ element of $\boldsymbol{\beta}$

The partial likelihood is useful especially when it is appreciable simpler than the full likelihood, for example when it involves only the parameter of interest and not nuisance parameters [10]. Later, Cox (1972) [2] gave the equation below

$$\prod_j \frac{\exp(\boldsymbol{\beta}^T X_j)}{\sum_{k \in R_j} \exp(\boldsymbol{\beta}^T X_k)}$$

as a likelihood for inference about $\boldsymbol{\beta}$.

## 3.2 A Bayesian MCMC Approach

The analysis of counting process in survival data is usually based on the modelling of the intensity function. The random variable *T* denote the survival time of an

individual and hence, the survival curve is given by $S(t) = P(T > t)$.

The baseline hazard rate $\lambda_0(t)$ is defined as the following limit, common to all subjects,

$$\lambda_0(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt)}{dt} \tag{3.1}$$

by assuming that $T$ is continuous and considering the calculation of the probability of an event happening over some finite time interval $[t, t + dt)$. The baseline cumulative hazard rate can be defined as

$$\Lambda_0(t) = \int_0^t \lambda_0(u) du$$

Cox model can describe the hazard function (also known as the risk function or intensity function) for subject $i$ with covariates $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ defined as

$$\lambda(t|X) = \lambda(t|X = 0)\exp(\beta^T \mathbf{X}) = \lambda_0(t)\exp(\beta^T \mathbf{X}) \tag{3.2}$$

The above equation shows the product of the unknown baseline hazard rate, $\lambda_0(t)$ and the exponential function of the unknown regression coefficient, $\beta^T \mathbf{X} = \sum_j \beta_j X_{ij}$.

The processes $N_i(t)$ can be observed for subjects $i = 1, 2, \dots, n$, that count the number of failures which have occurred up to time $t$. The counting process is constant and takes the value zero between failures and jumps one unit at each failure time. The new failure rate is then seen to be $I_i(t) = Y_i(t)\lambda(t|X_i)$. The intensity function is characterized as the probability of an event of interest occurs in the small time interval, given that it has not happened before.

The corresponding intensity process $I_i(t)$ is formulated as follows:

$$I_i(t)dt = P(dN_i(t) = 1 | F_{t-})$$
$$= P([N_i(t + dt) - N_i(t)] = 1 | F_{t-}) \tag{3.3}$$

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 $t$. Thus, the above equation can be written as

$$I_i(t)dt = E(dN_i(t) | F_{t-}) \tag{3.4}$$

If subject $i$ is observed to fail during this time interval, $dN_i(t)$ will take the value 1, otherwise $dN_i(t)$ is equal to zero. Hence $E(dN_i(t) | F_{t-})$ corresponds to the probability of subject $i$ failing in the time interval $[t, t + dt)$. As $dt$ approaches to zero (assume that time is continuous) this probability becomes the instantaneous hazard at time $t$ for subject $i$. This is assumed to have the proportional hazards form as follows:

$$I_i(t) = Y_i(t)\lambda_0(t)\exp(\beta X_i) \tag{3.5}$$

where $Y_i(t)$ is an observed process and take the value 1 or 0 according to whether or not subject $i$ is observed at time $t$. As we can see that $\lambda_0(t)\exp(\beta X_i)$ is the familiar Cox regression model. Thus we have observed data $D = N_i(t)$, $Y_i(t)$, $X_i$; $i = 1, 2, \dots, n$ and unknown parameters $\beta$ and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$, the latter to be estimated non-parametrically. The joint posterior distribution for the above model can be defined as

$$P(\beta, \Lambda_0() | D) \sim P(D | \beta, \Lambda_0())P(\beta)P(\Lambda_0()) \tag{3.6}$$

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

$$\prod_{i=1}^n \left[ \prod_{t \geq 0} I_i(t)^{dN_i(t)} \right] \exp(-I_i(t)dt) \tag{3.7}$$

The counting process increments $dN_i(t)$ in the time interval $[t, t + dt)$ are assumed to be independent Poisson random variables with means, $I_i(t)dt$, where

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

$I_i(t)dt$ can be written as,

$$I_i(t)dt = Y_i(t)\exp(\beta X_i)d\Lambda_0(t) \tag{3.8}$$

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 $[t, t + dt)$.

Kalbfleisch [9] proposed gamma process prior for the baseline hazard function by assuming $d\Lambda_0(t)$ to have gamma distribution with mean $d\Lambda_0^*(t)$ and variance $d\Lambda_0^*(t)/c$, which can be written as

$$d\Lambda_0(t) \sim \text{Gamma}(cd\Lambda_0^*(t), c)$$

Later, Ayman and Anis [12] proposed the baseline hazard function as the non-parametric part of the model to be a non-negative polygonal function with the vertices located at times $a_0 = 0 < a_1 < \cdots < a_T < a_{T_{max}+1}$, where the polygonal take the values $\tau_0 = 0 < \tau_1 < \cdots < \tau_T < \tau_{T_{max}+1}$, respectively and becomes constant after time $a_{T_{max}}$.

In this paper, the proposed baseline hazards function for Cox model is slightly different compared to the original Cox model which used gamma process prior as the baseline hazards function with both $c$ and $r$ are fixed values. 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 assumed to be the conjugate independent increments prior.

In this paper, we modified the increments to be the product of $d\Lambda_{01}^*(t)$ and $d\Lambda_{02}^*(t)$ in which we proposed the increments as

$$d\Lambda_0(t) = d\Lambda_{01}^*(t)d\Lambda_{02}^*(t)$$

where the increments $d\Lambda_0(t)$ is a product between $d\Lambda_{01}^*(t)$ and $d\Lambda_{02}^*(t)$. $d\Lambda_{01}^*(t)$ is assume to have gamma distribution with hyperparameters mu and $c$ as the shape and scale parameters, respectively and can be written as

$$d\Lambda_{01}^*(t) \sim \text{Gamma}(mu, c)$$

Here we suggested that shape, mu and scale, $c$ to have its own distribution in which mu has gamma distribution with mean 10 and variance 0.1 while $c$ also has gamma distribution with mean 1 and variance 10. We set $d\Lambda_{02}^*(t) = rdt$ where $r$ is a guess at the failure rate per unit time and has a uniform distribution, while $dt$ is the size of the time interval.

This paper also proposed both $c$ and $r$ to have its own distributions rather than assuming the values to be arbitrary constant values. In order to make the proposed model more flexible, we assume $c$ and $r$ to have gamma and

uniform distributions, respectively. The proposed full BUGS model is attached in the appendix.

## 4. RESULTS & DISCUSSION

We consider leukemia data where the effect of 6-MP (6-Mercaptopurine) therapy on the duration of remissions induced by adrenal corticosteroids has been studied as a model for testing of new agents [13]. Randomly, patients in remission were assigned to maintenance therapy with either 6-MP or placebo. The median duration of 6-MP-maintained complete remissions was 33 weeks and for placebo, 9 weeks. While the study was in progress, a sequential experimental design was used in analyzing remission times. This resulted in the study being stopped after analysis of the remission times of 21 pairs of patients (42 patients).

### 4.1 An analysis of Maximum Likelihood Estimation

We use the leukemia data to estimate the unknown parameters in Weibull proportional hazards model using maximum likelihood estimation approach. Table 1 shows the MLE estimates of the hazard function, cumulative hazard function, density function and survivor function for Weibull proportional hazards model.

Newton Raphson method is the popular method in maximizing the likelihood function and estimating the unknown parameters. It takes six iterations to converge in the analysis and the results are tabulated in Table 1. The unknown parameter beta of Weibull proportional hazard model is 1.731 and the standard error is 0.413. The likelihood ratio test is 19.6, while the log likelihood is -106.58. Scale and shape parameters for Weibull

proportional hazards model are 17.850 and 1.366, respectively. As can be seen, the p-value is very small and we can conclude that this is a significant value.

With the same data, we proceed to estimate the parameter for Cox regression model. The analysis involved the use of maximum likelihood estimation and maximum penalized partial likelihood estimation. Using Newton Raphson method as the maximum likelihood estimation, we obtain the estimated parameters and were tabulated in Table 2. The Newton Raphson iteration converges in four iterations and gives the estimated beta 1.659 with standard error 0.423 and log likelihood -93.599. With the same number of iterations using maximum penalized partial likelihood estimation, we obtained the estimated beta 1.650 with standard error 0.422 and log likelihood -86.869 and the results are also tabulated in Table 2.

As can be seen, the larger value of the log likelihood of maximum penalized likelihood estimation indicate better fitting models compared to the value of the log likelihood of maximum likelihood estimation using Newton Raphson method. This is because the partial likelihood is simpler than the full likelihood and it involves only the parameters of interest and not nuisance parameters.

### 4.2 An analysis of Bayesian modelling process

Using the same data, we continue our analysis with Bayesian approach using statistical packages WinBUGS and R. The choice of hyperparameters and initial values are not sensitive to the estimation of the parameters. Five chains with the same starting values were carried out simultaneously. 10,000 iterations is performed for each chain after 1000 iterations for burn-in, and one out of every 100th values is used.

**Table 1** Summaries of parameter estimations for Weibull Proportional hazards model using Maximum Likelihood Estimation (Newton Raphson Method)

| Parameter | Estimate | Std Error | $t$-value | $\Pr(> t)$ | Gradient |
|---|---|---|---|---|---|
| Alpha | 1.36576 | 0.20100 | 6.7949 | 1.084e-11 | 4.547474e-07 |
| Beta0 | -3.93614 | 0.60165 | -6.5422 | 6.061e-11 | 5.258016e-07 |
| Beta1 | 1.73087 | 0.41331 | 4.1878 | 2.817e-05 | -4.263256e-08 |

**Table2** Summaries of parameter estimation for Cox Regression parameter estimation using maximum likelihood estimation (Newton Raphson Method) & Maximum Penalized Partial Likelihood

| Methods | Log likelihood | beta | Rel Risk | Std Error | Hessian | LR Tset Stats |
|---|---|---|---|---|---|---|
| Maximum Likelihood Estimation (Newton Raphson) | -93.599 | 1.659 | 5.252 | 0.423 | 5.5997 | 17.7 |
| Maximum Penalized Partial Likelihood Estimation | -86.869 | 1.650 | 5.206 | 0.422 | 5.6121 | 17.5 |

**Table 3** Posterior statistics mean for Weibull Proportional Hazards Regression Model with different number of chain (10,000 from each of generated samples).

(a)   Summaries of posterior statistics

| No. of chain | node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|---|
| | b | -4.044 | 0.6512 | **0.05761** | -5.3 | -4.043 | -2.831 | 1001 | 10000 |
| 1 | beta | 1.772 | 0.4412 | **0.01967** | 0.9208 | 1.757 | 2.703 | 1001 | 10000 |
| | shape | 1.385 | 0.2134 | **0.01819** | 0.9862 | 1.384 | 1.803 | 1001 | 10000 |
| | b | -3.969 | 0.6088 | **0.03671** | -5.252 | -3.925 | -2.935 | 1001 | 20000 |
| 2 | beta | 1.745 | 0.4194 | **0.01253** | 0.9444 | 1.732 | 2.62 | 1001 | 20000 |
| | shape | 1.361 | 0.2026 | **0.0117** | 1.011 | 1.344 | 1.787 | 1001 | 20000 |
| | b | -4.149 | 0.6182 | **0.03154** | -5.474 | -4.082 | -3.099 | 1001 | 30000 |
| 3 | beta | 1.813 | 0.4212 | **0.01019** | 1.008 | 1.801 | 2.651 | 1001 | 30000 |
| | shape | 1.416 | 0.2058 | **0.01003** | 1.053 | 1.4 | 1.854 | 1001 | 30000 |
| | b | -3.873 | 0.5722 | **0.02411** | -5.051 | -3.859 | -2.821 | 1001 | 40000 |
| 4 | beta | 1.733 | 0.4106 | **0.008096** | 0.9384 | 1.728 | 2.557 | 1001 | 40000 |
| | shape | 1.331 | 0.1921 | **0.007686** | 0.9697 | 1.328 | 1.716 | 1001 | 40000 |
| | b | -3.952 | 0.5664 | **0.0226** | -5.041 | -3.952 | -2.888 | 1001 | 50000 |
| 5 | beta | 1.747 | 0.4214 | **0.0076** | 0.9439 | 1.736 | 2.601 | 1001 | 50000 |
| | shape | 1.356 | 0.1877 | **0.007578** | 1.004 | 1.351 | 1.725 | 1001 | 50000 |

(b)   Summaries of DIC and log likelihood

| No. of chain | $\bar{D}$ | $\hat{D}$ | DIC | $p_D$ | - Log likelihood |
|---|---|---|---|---|---|
| **1** | 216.4 | 213.2 | **219.5** | 3.135 | 108.2 |
| **2** | 216.2 | 213.2 | **219.2** | 2.970 | 108.1 |
| **3** | 216.3 | 213.3 | **219.2** | 2.925 | 108.15 |
| **4** | 216.1 | 213.2 | **219.1** | 2.919 | 108.05 |
| **5** | 217.2 | 213.2 | **221.1** | 3.949 | 108.6 |

different number of chains. All the value shows the estimation of the parameters with the credible intervals. $\bar{D}$ is the posterior mean of the deviance and is defined as $\bar{D} = -2 \times \log$ likelihood. $\hat{D}$ is a plug-in estimate of the latter based on the posterior mean of the parameters. The deviance information criterion, DIC is defined and computed as

$$\text{DIC} = \bar{D} + p_D = \hat{D} + 2p_D$$

and the lower values of the criterion indicate better fitting models. Table 3 shows different values of posterior means for parameter estimation and different DIC depends on the number of chain. We know that $\bar{D} = -2 \times \log$ likelihood thus the $\log$ likelihood $= \frac{\bar{D}}{-2}$. Log likelihood

for each chain has also tabulated in the same table. As we can see that the standard deviation and MCMC error reduced as we increased the number of chain. The smaller the standard deviation and MCMC error, the best results we get for the parameter estimation. The posterior distribution of the final sample after generating MCMC and an acceptable convergence to the stationary distribution for all hyperparameters of the model is shown in Figure 1. The generated observations of the trace plots are more convincing in terms of convergence, with all generated values within a parallel zone and no obvious tendencies or periodicities. The densities provide a graphical representation of the posterior densities estimate for each node. Figure 1 shows trace plots and posterior densities for

50,000 from each of five generated samples (10,000 for each chain).
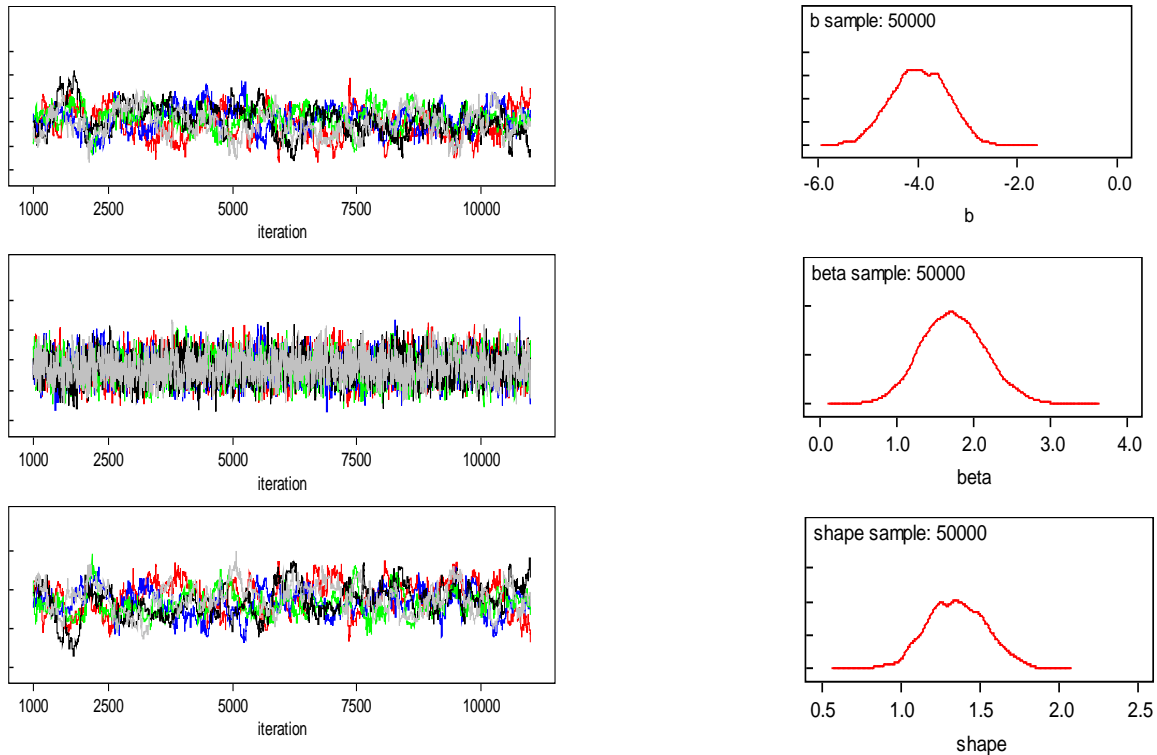


**(a) Trace plots for b, beta and shape parameter**

**(b) Posterior densities for b, beta and shape parameter**

**Fig. 1** MCMC diagnostic plots for Weibull Proportional Hazards Regression model using Gibbs Sampling when generating five chains with 10,000 iterations in each chain.

We analyze the data using WinBUGS and R programming to Cox regression model with gamma process prior as the baseline hazards function. Five chains with the same starting values were carried out simultaneously. After 1000 iterations for burn-in, 10,000 iterations is performed for each chain and one out of every 100th values is used. The choice of hyperparameters and initial values are not sensitive to the estimation of the parameters.

The posterior mean, standard deviation, MCMC error, and median with a 95% posterior credible interval and the deviance information criterion, DIC for original Cox regression model (Gamma process is used as a prior belief for the baseline hazard) and proposed baseline hazards function for Cox regression model for different number of chain are shown in Table 4. As can be seen in the Table 4, the proposed baseline hazards function for Cox regression model gives better fit compared to the original baseline hazard function using gamma process prior of Cox regression model. The deviance information criterions, DIC for original gamma process prior of Cox regression model in each chain are higher compared to the proposed baseline hazards function of Cox regression model. The lower values of the criterion indicate better fitting models and this value shows that the proposed baseline hazards function of Cox regression model gives better fit compared to the original gamma process prior of Cox regression model. We also compare the log likelihood for both Cox regression models and found that the log likelihood for original Cox regression model is higher compared to log likelihood for proposed Cox regression model in each chain. Figure 2 shows trace plots and posterior densities for 50,000 from each of five generated samples (10,000 for each chain). The densities provide a graphical representation of the posterior densities estimate for each node.

**Table 4** Posterior statistics mean with different number of chain (10,000 from each of generated samples).

    (a)    Summaries of posterior statistic

|  | No. of chain | node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|---|---|
| **Cox Model** | **3** | **Beta** | 1.538 | 0.4175 | **0.003017** | 0.7507 | 1.53 | 2.38 | 1001 | 30000 |
|  | **4** |  | 1.541 | 0.4205 | **0.002701** | 0.7485 | 1.53 | 2.392 | 1001 | 40000 |
|  | **5** |  | 1.541 | 0.4164 | **0.00224** | 0.7545 | 1.531 | 2.391 | 1001 | 50000 |
| **Proposed Cox Model** | **3** | **Beta** | 1.532 | 0.4072 | **0.003623** | 0.7629 | 1.519 | 2.367 | 1001 | 30000 |
|  | **4** |  | 1.542 | 0.4123 | **0.003378** | 0.7655 | 1.531 | 2.386 | 1001 | 40000 |
|  | **5** |  | 1.539 | 0.4115 | **0.002845** | 0.7595 | 1.528 | 2.38 | 1001 | 50000 |

    (b)    Summaries of DIC and log likelihood

|  | No. of chain | $\bar{D}$ | $\hat{D}$ | $p_D$ | DIC | -Log Likelihood |
|---|---|---|---|---|---|---|
| **Cox Model** | **3** | 212.677 | 192.907 | 19.771 | **232.448** | **106.3385** |
|  | **4** | 212.736 | 192.908 | 19.828 | **232.564** | **106.368** |
|  | **5** | 212.719 | 192.907 | 19.811 | **232.530** | **106.3595** |
| **Proposed Cox Model** | **3** | 204.734 | 200.336 | 4.397 | **209.131** | **102.367** |
|  | **4** | 204.789 | 200.366 | 4.423 | **209.211** | **102.3945** |
|  | **5** | 204.791 | 200.375 | 4.416 | **209.208** | **102.397** |
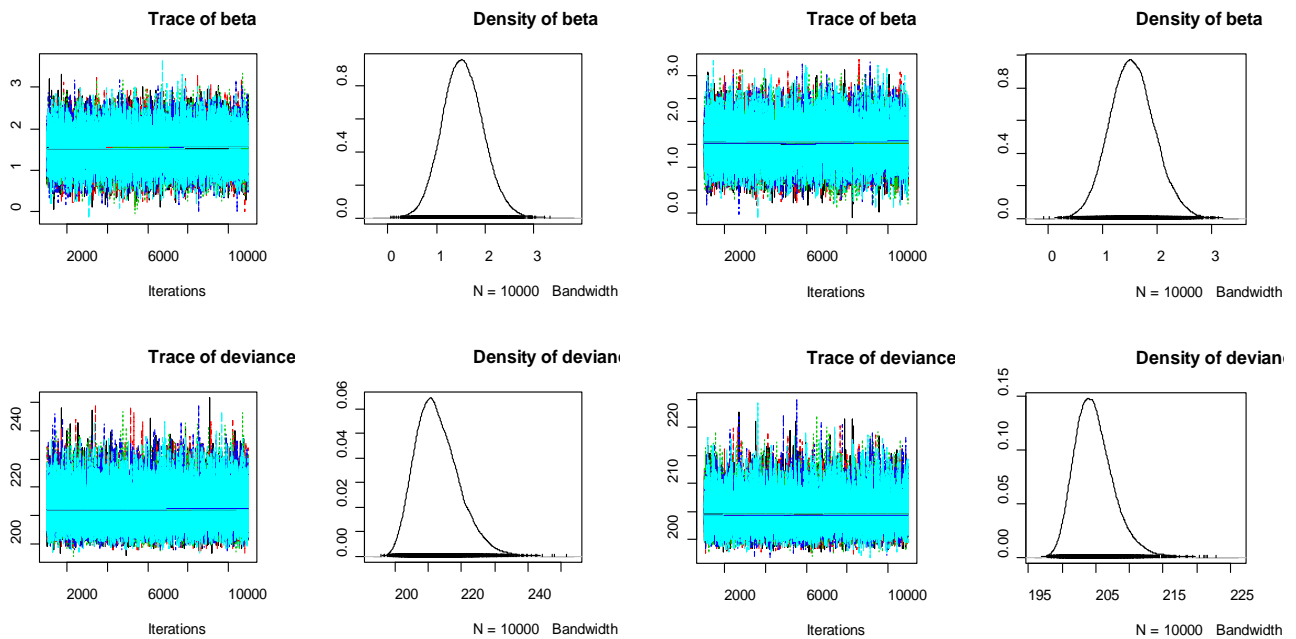


**(a)** **Trace plots and posterior densities of the parameters in the gamma process prior as the baseline hazard function in Cox Regression model.**

**(b)** **Trace plots and posterior densities of the parameters for proposed Cox Regression model.**

**Fig. 2** MCMC diagnostic plots for Cox Regression model (gamma process prior baseline hazard function and proposed baseline hazard function) using Gibbs Sampling when generating five chains with 10,000 iterations for each chain.

## 5.     CONCLUSION

We proposed a Bayesian method to fit more flexible survival models for non-informative censored data. Using WinBUGS and R software, we proved that by using the proposed Cox regression model the DIC value is smaller than the Gamma process prior for the nonparametric part in the original Cox regression model. This proposed model is simple and alternative to the maximum likelihood method or the Gamma process prior.

WinBUGS is a tool for analyzing survival data in a Bayesian framework using Markov Chain Monte Carlo (MCMC). While in R, we use R2WinBUGS package to make it possible to call a BUGS model, summarize inferences and convergence in a table and graph, and save the simulations in arrays for easy access in R.

The results of the statistical analysis in the example are consistent with those obtained from previous analysis. Bayesian models can be compared using the deviance information criterion (DIC), which its posterior distributions have been obtained using MCMC. DIC has been implemented as a tool in the BUGS software package. Bayesian inference has several advantages particularly in the flexibility of model-building for complex data over the frequentist approaches. Bayesian approach enables us to make exact inference for any sample size based on the posterior distribution.

## REFERENCES

[1]     E. L. Kaplan and P. Meier, Nonparametric Estimation from Incomplete Observations, Journal of the American Statistical Association, 53 (1958), 457– 481.

[2]     D.R. Cox, Regression Models and Life Tables, Journal of the Royal Statistical Society, 34 (1972), 187-220.

[3]     Y. Q. Chen and N. P Jewell, On a General Class of Semiparametric Hazards Regression Models, Biometrika, 88, (2001), 687 – 702.

[4]     E. Arjas and D. Gasbarra, Nonparametric Bayesian Inference From Right Censored Survival Data Using Gibbs Sampler, Statistica Sinica, 4, (1994), 505 – 524.

[5]     D. Sinha, and D. K. Dey, Semiparametric Bayesian analysis of survival data, Journal of the American Statistical Association, 92 (1997), 1195-1212.

[6]     J. G. Ibrahim and M. H. Chen, Prior Distributions and Bayesian Computation for Proportional Hazards Models, The Indian Journal of Statistics, 60, (1998), 48-64.

[7]     D. Spiegelhalter, A. Thomas, N. Best, and W. Gilks, BUGS 0.5: Examples Volume 1, MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, 1996.

[8]     D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, (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.

[9]     J. D. Kalbfleisch, Nonparametric Bayesian Analysis of Survival Time Data. Journal of the Royal Statistical Society, 40 (1978), 214 – 221.

[10]    D. R. Cox, Partly Likelihood, Biometrika, 62, (1975), 269 – 276.

[11]    J. S. Kim, Maximum Likelihood Estimation for the Proportional Hazards Model with Partly Interval-Censored Data, Journal of the Royal Statistical Society, 65 (2003), 489-502.

[12]    A. M. Ayman and B. G. Anis, Using WINBUGS to Cox Model with Changing from the Baseline Hazard Function, Journal Applied Mathematical Sciences, 5 (2011), 2217 – 2240.

[13]    E.O. Freireich, et al. The effect of 6-mercaptopmine on the Duration of Steroid Induced Remission in Acute Leukemia. Blood, 21 (1963), 699 – 716.

**Appendix 1**

```
# Model in WinBUGS
Model{
# Start model
# Set up data
        for(i in 1:N) {
        # Begin loop over observations within subjects T = 17, which indicates the number of unique failure times
        # Assume that obs.t[i] is the follow up time for patient i (i = 1,2,…,42) with fail [i]  indicating whether this
        # corresponds to failure or a censored observation.
                for(j in 1:T) {
                # Y[i,j] = Y[i](t[j]) is at risk set, Y[N,T] = 1 if observed time is at risk;  i.e., obs.t >= t and zero otherwise.
                        Y[i,j] <- step(obs.t[i] - t[j] + eps)
.               # eps = 0.000001 is used to guard against numerical imprecision in step function
                # counting process jump = 1 to indicates event time,
                        dN[i, j] <- Y[i, j] * step(t[j + 1] - obs.t[i] - eps) * fail[i]
                # obs.t in interval [ t[j], t[j+1]); i.e. if t[j] <= obs.t < t[j+1]
                # N[i,j] = Ni(tj)
                # fail[i] is an indicator to be failure = 1; censored = 0
                        } # End loop over observations within subjects
                } # End loop over subjects
        # Model
                for(j in 1:T) {
                # Begin loop over subjects
                        for(i in 1:N) {
                        # Begin loop over observations within subjects
                                dN[i, j] ~ dpois(Idt[i, j])
                        # Likelihood dL0[T], increment in unknown hazard function
                        # Intensity
                                Idt[i, j] <- Y[i, j] * exp(beta * Z[i]) * dL0[j]
                                } # End loop over subjects
                } # End loop over observations within subjects
                for (j in 1 : T) {
                # Begin loop over observations within subjects using proposed baseline hazard function
                                dL0[j] <- dL01.star[j] * dL02.star[j]
                                dL01.star[j] ~ dgamma(mu[j], c)
                # prior mean hazard
                                mu[j] ~ dgamma(1.0E3,1.0E2)
                                dL02.star[j] <- (1/r) * (t[j + 1] - t[j])
                }# End loop over observations within subjects
                # Priors for beta
                # The prior for beta parameter is mutually independent normal with mean 0 and variance 100000 (recall
                # that WinBUGS 1.4 specifies the precision is 0.000001 instead of the variance = 100000).
                        beta ~ dnorm(0.0,1.0E-6)
                # Flexibility for the proposed baseline hazard function for Cox model by assuming both c and r to have
                # gamma distribution and uniform distribution, respectively.
                        c ~ dgamma(1.0E-1, 1.0E-1)
                        r ~ dunif(1.0E4, 1.0E5)
} # End model
```