



**BAYESIAN ANALYSIS AND MODEL SELECTION FOR
INTERVAL-CENSORED SURVIVAL DATA**

by

Debajyoti Sinha, Ming-Hui Chen, and Sujit K. Ghosh

Institute of Statistics Mimeograph Series No. 2298

May 1997

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

The Library of the Department of Statistics
North Carolina State University

Mimeo Series	
No. 2298	
May 1997	
Bayesian Analysis and Model Selection for Interval- censored survival data	
BY: Sinha, Chen and Ghosh	
Name	Date

Department of Statistics Library

Bayesian Analysis and Model Selection for Interval-censored Survival Data

Debajyoti Sinha*, Ming-Hui Chen[†] and Sujit K. Ghosh[‡]

April 10, 1997

Abstract

Interval-censored data occur in survival analysis when the survival time of each patient is only known to be within an interval and these censoring intervals differ from patient to patient. This kind of data pose some challenges to the semiparametric analysis and model diagnostics. For such data, we present some Bayesian discretized semiparametric models, incorporating the proportional and non-proportional hazards structures, along with the associated statistical analyses and tools for model selection using sampling based methods. The scope of these methodologies is illustrated through a re-analysis of the historical data set from Finkelstein (1986).

Key Words: CPO, Gibbs sampler, Prior process.

1 Introduction

Many clinical trials and medical studies use periodic scheduled follow-ups of each patient to monitor the time to an event of interest or disease (i.e. survival time T of the patient) whose occurrence is not apparent from outside. The occurrence of such event can be detected only through some invasive procedure (such as testing blood or tissue samples etc.) performed during these clinic visits. Medical researchers often come across interval censoring in such studies when the patients miss some of the scheduled appointments for reasons not related to the survival times and the observed censoring intervals containing their survival times frequently overlap with each other. Interval-censored survival data have

*Department of Mathematics, University of New Hampshire, Durham, NH 03824-3591. Dr. Sinha's research was supported by the grant R29-CA69222-02 from NCI.

[†]Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609-2276

[‡]Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203

recently received much attention in biostatistical and statistical literature due to diseases such as AIDS and some forms of cancers. For recent reviews, see Satten (1996), and Frydman (1995). The data set in Table 3 of Finkelstein and Wolfe (1985) is a historical data set of interval-censored data. In this data set, 46 early breast cancer patients receiving only radiotherapy (covariate value $z = 0$) and 48 patients receiving radio-chemotherapy ($z = 1$) were monitored for cosmetic change during weekly clinic visits. But, some patients missed some of their weekly visits. So, the data on survival time are typically recorded as, for example, (7,18] (at the 7th week clinic-visit, patient had shown no change and then in the next clinic visit at the 18th week the patient's tissue showed that the change had already occurred). Since, the clinic visits of different patients occurred at different times, the censoring intervals in the data set are found to be often overlapping.

We are interested to see the effect of the covariate z associated with the patient, on the survival time T . A popular semiparametric approach to model survival time, in the presence of covariate effects is proposed in the Cox's (1972) proportional hazards model, given by, $\lambda(t|z) = \lambda_0(t)e^{\beta z}$. Here $\lambda(t|z) = -\frac{d}{dt} \log P(T > t|z)$ is the hazard function of T given z , β is the time-independent regression coefficient for the covariate z and $\lambda_0(t)$ is the baseline hazard function. Finkelstein (1986) and Satten (1996) analyzed interval-censored data under the assumption of Cox model. But, such an assumption of time-independent regression coefficient may not always be valid. The major contribution of the present paper is two fold. With the advancement of the sampling based computational tools, it is now feasible to consider more general models which incorporates time-varying coefficients. Secondly, while powerful computational tools enable us to fit remarkably complex models we should not lose sight of the need to make suitably parsimonious choices. So, we develop some Bayesian tools for model selection and model validation. So far, to our knowledge there is no formal statistical method to select among the models we propose or to check any modeling assumption such as time-independent coefficient for interval-censored data. In addition, Bayesian method enables us to obtain exact small sample inference on the parameter of interest (i.e. the regression coefficient), from the moderate sized data set even with a high-dimensional nuisance parameter (i.e., the baseline hazard).

In Section 2, we propose a Bayesian version of discretized Cox model and a model with time-varying coefficients. In Section 3, we describe model fitting using sampling based method. In Section 4, we present some Bayesian model selection and model checking methods. In Section 5, we illustrate the proposed methods by reanalyzing the breast cancer data of Finkelstein and Wolfe (1985). Section 6 concludes with some remarks.

2 Models

We take the hazard to be a piecewise constant function with $\lambda(t|z) = \lambda_k \theta_k^z$ for $t \in I_k$, where $\theta_k = e^{\beta_k}$, $I_k = (a_{k-1}, a_k]$ for $k = 1, 2, \dots, g$, $0 = a_0 < a_1 < \dots < a_g = \infty$, and g is the total number of grid intervals. The length of each grid can be taken to be sufficiently small to approximate any hazard function for all practical purposes. Now, we present two Bayesian semiparametric discretized models, viz. a discretized version of the Cox model (which we call \mathcal{M}_0) and a discretized hazard model with time-dependent regression coefficient (which we call \mathcal{M}_1). More precisely, these features are captured through their prior specifications as follows:

\mathcal{M}_0 : (1) $\lambda_k \overset{\text{indep}}{\sim} \text{Gamma}(\eta_k, \gamma_k)$ for $k = 1, \dots, g$; (2) $\beta \sim N(\beta_0, w_0^2)$ and β is apriori independent of $\lambda = (\lambda_1, \dots, \lambda_g)$.

\mathcal{M}_1 : (1) λ has same prior as in \mathcal{M}_0 ; (2) $\beta_{k+1} | \beta_1, \dots, \beta_k \sim N(\beta_k, w_k^2)$ for $k = 0, \dots, g-1$ and the β_k 's are apriori independent of λ .

In above, we assume that the hyperparameters of these models, viz., η_k 's, γ_k 's, w_k 's and β_0 are known in advance.

\mathcal{M}_0 is a discretized version of Cox model with a discretized version of the gamma process prior (Kalbfleisch 1978) for the baseline hazard $\lambda_0(\cdot)$ where η_k/γ_k is the prior mean and η_k/γ_k^2 is the prior variance of λ_k . When the grid intervals are sufficiently small, this discretized version will be indistinguishable from the actual time-continuous gamma process. The discretized autocorrelated prior process for β_k 's in \mathcal{M}_1 allows the covariate effect to change over time, but also incorporates the prior information that the values of the coefficient β in adjacent intervals are expected to be somewhat close and the dependence among the β 's decrease as the intervals become further apart. This assumption seems to be in complete accord with some studies where the covariate effect may change over time, but is not expected to change too wildly over time. The parameters w_k 's can be used as a tuning device to determine our prior opinion about the possible change in the magnitude of β over time. For example, apriori we expect the β_{k+1} to be within approximately $1.96w_k$ from the β_k with 95% confidence. The w_k 's should depend on the lengths of the I_k 's allowing the coefficient to change more for bigger grid intervals. It is possible to use an autocorrelated prior process for the baseline hazard also. For details on the use and properties of an autocorrelated process, see Sinha and Dey (1997), and Sargent (1996). Our major interest is to compare the Cox model (\mathcal{M}_0) with the time varying coefficient model (\mathcal{M}_1).

For the example of breast cancer data, we consider following values of the hyperparameters.

	η_k	γ_k	β_0	w_k	w_0
\mathcal{M}_0	0.2	0.4	0	-	2.0
\mathcal{M}_1	0.2	0.4	0	1.0	2.0

In practice, these hyperparameters should be chosen to represent a carefully developed prior opinion based on the results from previous studies related to the present experiment. Often in clinical trials setting, scientists conduct multiple studies to investigate single important medical or biological issue. One advantage of Bayesian method is that it allows us to incorporate the prior expertise into the experiment. For example, we can use data from previous studies to elicit the prior mean (η_k/λ_k) and variance (η_k/λ_k^2) of λ_k . But in our case, we have chosen the hyperparameters to represent prior opinions which are nearly non-informative (flat) in the subset of the parameter space supported by the likelihood. For example, the common prior mean of λ_k 's is taken as 0.5 and the common prior variance is taken as 1.25 ($\lambda_k > 0$) which gives very non-informative prior opinion in the the range (0.1) and it is clear that for this example we expect λ_k 's to be less than 1. We have also made some minor simplifications, for example, taking all the λ_k 's to be identically distributed.

3 Model Fitting

Let us denote the observed interval-censored data from n patients, by $\mathbf{Y} = \{(a_{l_i}, a_{r_i}]; z_i : i = 1, 2, \dots, n\}$, where survival time T_i for the i th patient is known to be within $(a_{l_i}, a_{r_i}]$ and $a_{l_i} < a_{r_i}$ are two of the grid points (a_1, a_2, \dots, a_g) but not necessarily consecutive ones. And, z_i is the covariate value for the i th patient. Let $\mathbf{X} = \{T_i, z_i : i = 1, 2, \dots, n\}$ be the unobserved complete (augmented) data. For notational simplicity, we denote the set of parameters (λ 's and β 's) under any model by Θ , (however the dimension of Θ depends on the model under consideration). The distribution of T_i given z_i is piecewise exponential and thus the complete-augmented-data likelihood is

$$L_c(\Theta | \mathbf{X}) \propto \prod_{k=1}^g \left[\lambda_k^{d_k} \exp \left\{ -\lambda_k \sum_{j \in R_k} \theta_k^{z_j} \Delta_{jk} \right\} \theta_k^{\sum_{j \in D_k} z_j} \right], \quad (3.1)$$

where R_k is the set of patients at risk at a_{k-1} , $\Delta_{jk} = \min(T_j, a_k) - a_{k-1}$, D_k is the set of patients failing in $I_k = (a_{k-1}, a_k]$, and d_k is the number of patients in D_k . The observed data likelihood is complicated and the joint posterior distribution is analytically intractable even for \mathcal{M}_0 .

We employ sampling based methods, in particular, the Gibbs sampler, (see Tanner 1996 for a review) to sample from the joint posterior of Θ given \mathbf{Y} . Gibbs is an iterative algorithm which alternates between (1) generating the augmented survival times (T_i 's) from the conditional distribution of T_i 's given the

current values of the parameters (Θ) and observed censoring intervals (\mathbf{Y}) and then (2) generating new values of the parameters from the conditional distribution of Θ given the augmented data (the set of T_i 's). We use the conventional notations $[V|W]$ to denote the conditional distribution of V given W , and $A_{(a)}$ to denote the set A after removing the element a from it. For \mathcal{M}_0 and \mathcal{M}_1 , $[\lambda_k | \Theta_{(\lambda_k)}, \mathbf{X}]$ is obtained by multiplying (3.1) with the joint prior distribution of Θ and then by collecting the terms involving λ_k from that product. It might be noted that in order to sample (the parameters or augmented data), we just need to know the complete conditional densities up to their normalizing constants. Thus, we obtain the $[\lambda_k | \Theta_{(\lambda_k)}, \mathbf{X}]$ as $\text{Gamma}(\eta_k + d_k; \gamma_k + \sum_{j \in R_k} \theta_k^{z_j} \Delta_{jk})$. For \mathcal{M}_1 , the conditional posterior of β_k can be obtained by similar procedure but only up to the corresponding normalizing constants and

$$[\beta_k | \Theta_{(\beta_k)}, \mathbf{X}] \propto \phi \left(\beta_k \mid \frac{(\sum_{j \in D_k} z_j) w_k^2 w_{k-1}^2 + \beta_{k-1} w_k^2 + \beta_{k+1} w_{k-1}^2}{w_k^2 + w_{k-1}^2}, \frac{w_k^2 w_{k-1}^2}{w_k^2 + w_{k-1}^2} \right) \\ \times \exp \left(-\lambda_k \sum_{j \in R_k} \theta_k^{z_j} \Delta_{jk} \right), \quad (3.2)$$

for $k = 1, 2, \dots, g$, where $\beta_{g+1} = 0$ and $\phi(\cdot | \mu; \sigma^2)$ is the $N(\mu; \sigma^2)$ density function. For \mathcal{M}_0 , all the θ_k 's will be equal to $\theta = e^\beta$ and

$$[\beta | \Theta_{(\beta)}, \mathbf{X}] \propto \left(\prod_{k=1}^g \exp \left\{ -\lambda_k \sum_{j \in R_k} e^{\beta z_j} \Delta_{jk} \right\} \exp \left\{ \beta \sum_{j \in D_k} z_j \right\} \right) \times \phi(\beta | \beta_0; w_0^2). \quad (3.3)$$

The conditional distribution of $[T_i | \Theta, \mathbf{Y}]$ is truncated piecewise exponential with parameters $\theta_k \lambda_k$ for $l_i + 1 \leq k \leq r_i$ and support $(a_{l_i}, a_{r_i}]$. So, the density of $[T_i | \Theta, \mathbf{Y}]$ is

$$f(T_i | \Theta, \mathbf{Y}) = \frac{\lambda_k \theta_k^{z_i} \exp \left\{ -\sum_{l=l_i+1}^{k-1} \lambda_l \theta_l^{z_i} \bar{\Delta}_l - \lambda_k \theta_k^{z_i} (T_i - a_{k-1}) \right\}}{1 - \exp \left\{ -\sum_{l=l_i+1}^{r_i} \lambda_l \theta_l^{z_i} \bar{\Delta}_l \right\}}, \quad (3.4)$$

for $T_i \in I_k$, $l_i + 1 \leq k \leq r_i$, and $\bar{\Delta}_l = a_l - a_{l-1}$. Note that (3.4) is a product of multinomial and truncated exponential densities. Conditional distribution of $[T_i | \Theta, \mathbf{Y}]$ for \mathcal{M}_0 is similar to (3.4).

For both \mathcal{M}_0 and \mathcal{M}_1 , λ 's and T_i 's can be sampled straightforwardly using standard statistical subroutines, such as IMSL. However complete conditional distributions of β 's, mentioned above do not correspond to standard statistical distributions, so we use the adaptive rejection algorithm of Gilks and Wild (1992), an algorithm to sample from any univariate log-concave density, to simulate those parameters as all of these conditional distributions are log-concave. Therefore, implementation of the

iterative Gibbs algorithm to sample from the joint posterior $[\Theta|\mathbf{Y}]$ becomes relatively easy although the density of $[\Theta|\mathbf{Y}]$ has a fairly complicated form. It is noted that, in addition to ease of implementation, we no longer restrict our inference to a set of stringent regularity assumptions (as often needed to obtain asymptotic optimality).

4 Bayesian Model Selection

In this section, we propose two different model selection approaches using a generalization of the predictive loss of Laud and Ibrahim (1995) and the Conditional Predictive Ordinate (CPO) (Gelfand, Dey, and Chang 1992) for the interval-censored survival data. Both approaches are appropriate for model selection when the competing models are nested within each other (as \mathcal{M}_0 is a special case of \mathcal{M}_1) and even when they are not.

Our first model selection criterion combines both the predictive variability of the model and performance of the model at observed data points. Let \mathcal{M} denote a model, which can be one of \mathcal{M}_0 or \mathcal{M}_1 . Following Gelfand and Ghosh (1995), we begin with a quadratic loss function

$$\mathcal{L}_{\mathcal{M}}(\mathbf{y}_{rep}, \mathbf{a}, \mathbf{y}_{obs}) = (\mathbf{y}_{rep} - \mathbf{a})^T(\mathbf{y}_{rep} - \mathbf{a}) + \kappa(\mathbf{y}_{obs} - \mathbf{a})^T(\mathbf{y}_{obs} - \mathbf{a})$$

where \mathbf{y}_{rep} is an unobservable replication with the same distribution of the vector of observed uncensored log-survival times \mathbf{y}_{obs} and action \mathbf{a} , which depends on the underlying model \mathcal{M} , is a guess for \mathbf{y}_{rep} . The motivation behind the consideration of such a loss function relies on the standard utility ideas as in, e.g., Raiffa and Schlaifar (1961). The key idea is to replace the experiments (under standard decision-theoretic set-up) with models and then separately for each model, minimize the posterior expected loss with respect to \mathbf{a} to compute $\mathcal{L}_{\mathcal{M}}^*$ and then select the model with minimum $\mathcal{L}_{\mathcal{M}}^*$ value. The action \mathbf{a} can be viewed as a “compromise”, which we like to be close to \mathbf{y}_{obs} (a goodness-of-fit property) as well as to \mathbf{y}_{rep} (predictive property of \mathbf{a}) under the model \mathcal{M} . When the model is too restrictive, it is difficult to keep \mathbf{a} close to \mathbf{y}_{obs} and when the model is too general it is difficult to keep \mathbf{a} close to \mathbf{y}_{rep} which has higher variability for bigger model. Thus κ can be viewed as a predetermined constant to decide how much emphasis should be given on goodness-of-fit of \mathcal{M} compared to its predictive capacity. The use of squared loss (on the log-scale) is kind of arbitrary. However it may be noted that, under regularity condition as obtained in Gelfand and Ghosh (1995), squared loss appear as an approximation to more general classes of loss functions. For more details, see Gelfand and Ghosh (1995). Our aim is to minimize posterior expected loss $E(\mathcal{L}_{\mathcal{M}}|\mathbf{Y})$ w.r.t. \mathbf{a} where \mathbf{y}_{obs} restricted to $t_{i,obs} \in (a_{l_i}, a_{r_i}]$ for $i = 1, \dots, n$.

With basic calculations we can show that the minimum posterior expected loss for interval-censored data can be expressed as

$$L_{\mathcal{M}}^* = \sum_{i=1}^n \left[\text{Var}(\ln T_{rep} \mid z_i, \mathbf{Y}, \mathcal{M}) + \delta(\mu_i - \nu_{i,obs})^2 \right], \quad (4.5)$$

where $\mu_i = E(\ln T_{rep} \mid z_i, \mathbf{Y}, \mathcal{M})$: expected log-survival time of a new patient similar to the i th patient with covariate value z_i . More details on the derivations of $L_{\mathcal{M}}^*$ can be found in section 2.5 of Gelfand and Ghosh (1997). The expectation is taken under the predictive distribution of \mathbf{y}_{rep} given the observed interval-censored data \mathbf{Y} and model \mathcal{M} . Also, $\text{Var}(\ln T_{rep} \mid z_i, \mathbf{Y}, \mathcal{M})$ variance of the log-survival time of the new patient under the same predictive distribution, $\nu_{i,obs}$ is the point in $(a_{l_i}, a_{r_i}]$ closest to μ_i , that is

$$\nu_{i,obs} = \begin{cases} \mu_i & \text{if } \ln a_{l_i} \leq \mu_i \leq \ln a_{r_i} \\ \ln a_{l_i} & \text{if } \mu_i < \ln a_{l_i} \\ \ln a_{r_i} & \text{if } \mu_i > \ln a_{r_i}, \end{cases}$$

and $\delta = \frac{\kappa}{\kappa+1}$. A typical value of δ is 1/2 which corresponds to $\kappa = 1$ when we give equal importance to the goodness-of-fit and the variability in prediction for model selection purpose. To compare between models \mathcal{M}_0 and \mathcal{M}_1 we choose the model with lower $L_{\mathcal{M}}^*$ (minimum posterior expected loss under the model) value. Note that $L_{\mathcal{M}}^*$ can be also viewed as an extreme special case of the so-called $L_{\mathcal{M}}$ criterion of Laud and Ibrahim (1995), which is obtained by letting $\kappa \rightarrow \infty$ (or equivalently $\delta \rightarrow 1$). One attractive feature of this decision theoretic formulation of the model choice criteria (such as $L_{\mathcal{M}}^*$) is that it extends the usual testing procedures (whish are based on 0-1 type loss functions) to a fairly general class of loss functions. In fact, the $L_{\mathcal{M}}^*$ as proposed above can be used as an analogue for standard LRT statistic (if we replace the mle's in LRT by the corresponding posterior expected values with reasonably flat priors). For more details, see Gelfand and Ghosh (1997). As a computational remark, it may be noted that the $L_{\mathcal{M}}^*$ can be computed easily (as compared to LRT statistic for our models) from the output of the Gibbs sampler and using the fact that $[\mathbf{y}_{rep} \mid \mathbf{y}_{obs}] = \int [\mathbf{y}_{rep} \mid \Theta] \cdot [\Theta \mid \mathbf{y}_{obs}] d\Theta$. This is one of the reasons, we prefer to use $L_{\mathcal{M}}^*$ over LRT statistic to select the "best" model.

The scope of the Bayesian cross validation approach of CPO in model selection and model adequacy study is explored in Gelfand et al. (1992). For our problem, the CPO value for the i th observation is defined as $\text{CPO}_{\mathcal{M}}(i) = P(T_i \in (a_{l_i}, a_{r_i}] \mid z_i, \mathbf{Y}_{(i)}, \mathcal{M})$, where $\mathbf{Y}_{(i)}$ is the interval-censored data with the i th patient removed. $\text{CPO}_{\mathcal{M}}(i)$ is the posterior predictive probability of the observed data from the i th patient given the modified data $\mathbf{Y}_{(i)}$ and under the assumption that the true model is \mathcal{M} . The larger the value of $\text{CPO}_{\mathcal{M}}(i)$ is, the more the i th observation supports the model \mathcal{M} . Gelfand et al. showed

that the $\text{CPO}_{\mathcal{M}}(i)$ can be computed as

$$\begin{aligned} \text{CPO}_{\mathcal{M}}(i) &= \left(\mathbb{E} \left[\frac{1}{P(T_i \in (a_{l_i}, a_{r_i}] \mid \Theta, z_i)} \right] \right)^{-1} \\ &= \left(\mathbb{E} \left[\frac{1}{\exp\{-\sum_{j=1}^{l_i} \lambda_j \theta_j^{z_i} \tilde{\Delta}_j\} - \exp\{-\sum_{j=1}^{r_i} \lambda_j \theta_j^{z_i} \tilde{\Delta}_j\}} \right] \right)^{-1}, \end{aligned} \quad (4.6)$$

where the expectation is taken with respect to the joint posterior of $[\Theta \mid \mathbf{Y}]$ under model \mathcal{M} . From (4.6), $\text{CPO}_{\mathcal{M}}(i)$ can be easily computed by using Gibbs samples from the $[\Theta \mid \mathbf{Y}]$ under model \mathcal{M} . In terms of single summary measure based on the $\text{CPO}_{\mathcal{M}}(i)$'s, we define the Pseudo Marginal Likelihood (PML) for model \mathcal{M} on the natural log scale as $\ln[\text{PML}(\mathcal{M})] = \sum_{i=1}^n \ln[\text{CPO}_{\mathcal{M}}(i)]$. Therefore, based on the above single measure $\ln[\text{PML}(\mathcal{M})]$, to compare between models \mathcal{M}_0 and \mathcal{M}_1 , we choose the model with maximum $\ln[\text{PML}(\mathcal{M})]$ value. Since we have the $\text{CPO}_{\mathcal{M}_j}(i)$'s ($j = 0, 1$ and $i = 1, \dots, n$) for all the patients and under both models, we will also plot $\ln \left(\frac{\text{CPO}_{\mathcal{M}_0}(i)}{\text{CPO}_{\mathcal{M}_1}(i)} \right)$ versus i to see how many data points support model \mathcal{M}_0 over \mathcal{M}_1 .

5 An Illustrative Example: The Breast Cancer Data

In this section, we are reanalyzing the breast cancer data of Finkelstein and Wolfe (1985). We implement the Gibbs samplers for the two proposed models under consideration. The convergence of the Gibbs samplers are checked by using several diagnostic procedures as recommended by Cowles and Carlin (1996) and after convergence, we generate 10,000 Gibbs iterates for calculating 95% highest posterior density (HPD) intervals of parameters of interest, $L_{\mathcal{M}}^*$ criteria, the CPO's, and the PML's for the two proposed models.

The marginal posterior of β for the Cox model (\mathcal{M}_0) has mean 0.559492 and the 95% HPD interval of β is (0.036, 1.071). It shows that there is an evidence of slightly higher risk for the second group under Cox model. Under this assumption, this finding is very comparable to the highly significant p-value 0.004 obtained by Finkelstein (1986) with her score test for $H_o : \beta = 0$. Also, the marginal distribution of β is slightly skewed to the left. We also calculate 95% HPD intervals for all β_k 's for \mathcal{M}_1 and we have plotted them versus k in Figure 1 along with two horizontal lines marking the 95% HPD interval for β of \mathcal{M}_0 . From Figure 1, we observe that for \mathcal{M}_1 , all 95% HPD intervals of the β_k 's contain 0 and the HPD interval for β is much narrower compared to the HPD intervals for β_k 's. It suggests that the final conclusion may depend on the underlying modeling assumptions and the selection of model is very crucial for correct inference in such a problem.

To compare between models \mathcal{M}_0 and \mathcal{M}_1 , we compute the $L_{\mathcal{M}}^*$ values for $\delta = 1, 1/2$ and the $\ln[\text{PML}_{\mathcal{M}}]$ for both models. The results are given in Table 1. Based on the two $L_{\mathcal{M}}^*$ criteria and the $\ln[\text{PML}_{\mathcal{M}}]$ criterion, \mathcal{M}_0 is more preferable model than \mathcal{M}_1 and there is no need to use more general model \mathcal{M}_1 instead of \mathcal{M}_0 to make inference from this data.

Table 1: The $L_{\mathcal{M}}^*$ Values and $\ln[\text{PML}_{\mathcal{M}}]$ for Two Bayesian Models

Model	$\delta = 1/2$	$\delta = 1$	$\ln[\text{PML}_{\mathcal{M}}]$
\mathcal{M}_0	65.15	76.52	-157.61
\mathcal{M}_1	70.65	81.9	-188.33

Next, we plot the $\ln\left(\frac{\text{CPO}_{\mathcal{M}_0}(i)}{\text{CPO}_{\mathcal{M}_1}(i)}\right)$'s versus i in Figure 2. In Figure 2, we notice that 84% of log CPO-ratios are positive. Therefore, we conclude that the most of the data points support \mathcal{M}_0 over \mathcal{M}_1 , which is consistent with our conclusion from the single summary measures in Table 1.

6 Concluding Remarks

In this article, we have presented a method of Bayesian modeling of interval censored data allowing the regression coefficient of the covariate effect under Cox model to change over time. Cox model is a special case of this class of models. Our models can be used with discretized version of different types of prior processes for the baseline hazard, including gamma and other Levy processes and most of the non-Levy processes (e.g. autocorrelated process). The choices for modeling the prior information on the covariate effect are also very large. Bayesian analysis using MCMC is shown to be feasible for such models. We also present, for the first time in literature, two model selection criteria which can be applied even to other types of censoring (including case-1 censored data). We envision that these model selection criteria can be utilized to diverse field of applications including variable selections and comparing between two or more non-nested models, say, discretized Bayesian Cox model and discretized Bayesian proportional odds model. Here we haven't explored the optimal properties of our criteria from any theoretical aspect. However we feel comfortable to use these criteria, as they don't depend on their large-sample properties.

References

Sinha, D. and Dey, D.K. and (1997), Semiparametric Bayesian methods for survival data. To appear in the *Journal of the American Statistical Association*

- Cowles, M.K. and Carlin, B.P. (1996), Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *Journal of the American Statistical Association*, **90**, 883-904.
- Cox, D.R. (1972), Regression models and life tables, *Journal of the Royal Statistical Society, B*, **34**, 187-220.
- Finkelstein, D.M. (1986), A proportional hazards model for interval-censored failure time data, *Biometrics*, **42**, 845-854.
- Finkelstein, D.M. and Wolf, R.A. (1985), A semiparametric model for regression analysis of interval-censored failure time data, *Biometrics*, **41**, 933-945.
- Frydman, H. (1995), Nonparametric estimation of a Markov 'illness-death' process from interval-censored observations, with application to diabetes survival data, *Biometrika*, **82**, 773-790.
- Gelfand, A.E., Dey, D.K. and Chang, H. (1992), Model determination using predictive distributions with implementation via sampling-based methods, *Bayesian Statistics 4*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.), Oxford University Press, 147-167.
- Gelfand, A.E. and Ghosh S.K. (1995), Model Choice: A minimum posterior predictive loss approach, *Technical Report*, Department of Statistics, University of Connecticut.
- Gilks, W.R. and Wild, P. (1992), Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, **41**, 337-348.
- Kalbfleisch, J. D. (1978), Nonparametric Bayesian analysis of survival time data, *Journal of the Royal Statistical Society, B*, **40**, 214-221.
- Laud, P.W. and Ibrahim, J.G. (1995), Predictive Model Selection, *Journal of the Royal Statistical Society, B*, **57**, 247-262.
- Raiffa, H. and Schlaifer, R. (1961), *Applied Statistical Decision Theory*. Boston: Harvard University Press.
- Sargent, D.J. (1996), A flexible approach to time-varying coefficients in the Cox regression setting, *Technical Report*, Division of Biostatistics, School of Public Health, University of Minnesota.
- Satten, G.A. (1996), Rank based inference in the proportional hazards model for interval censored data, *Biometrika*, **83**, 355-370.
- Tanner, M.A. (1996), *Tools for Statistical Inference*, Third Edition, Springer Verlag, New York.

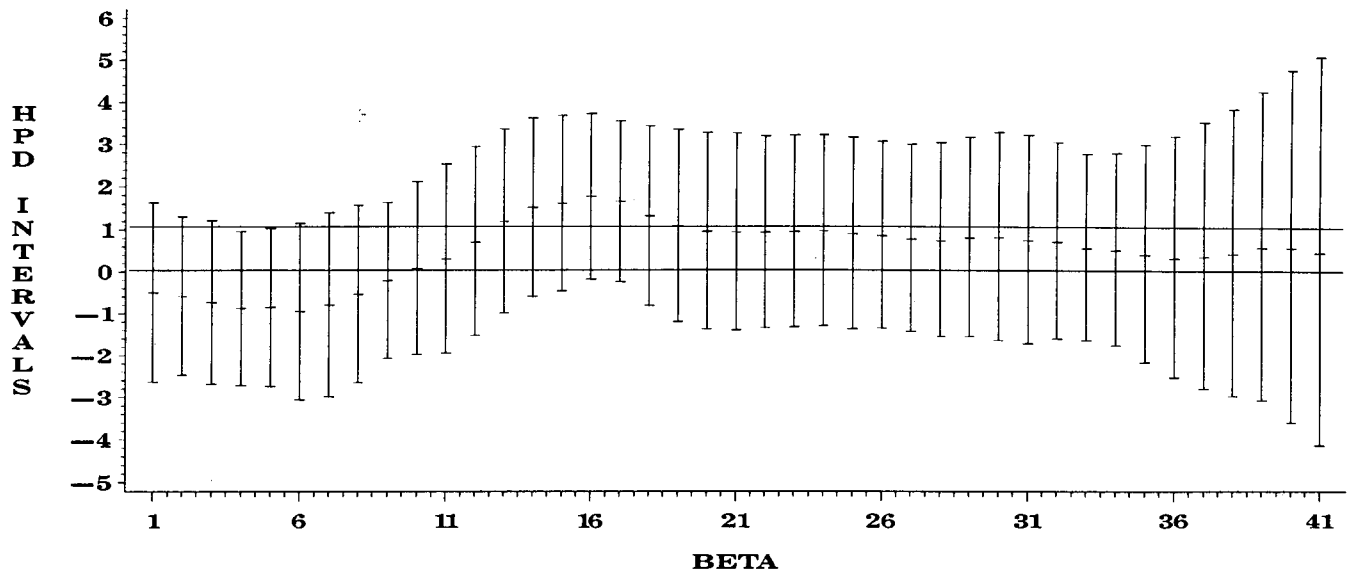


Figure 1: 95% HPD intervals of β_k 's versus k for \mathcal{M}_1 (41 vertical lines) and 95% HPD interval of β for \mathcal{M}_0 (two horizontal lines)

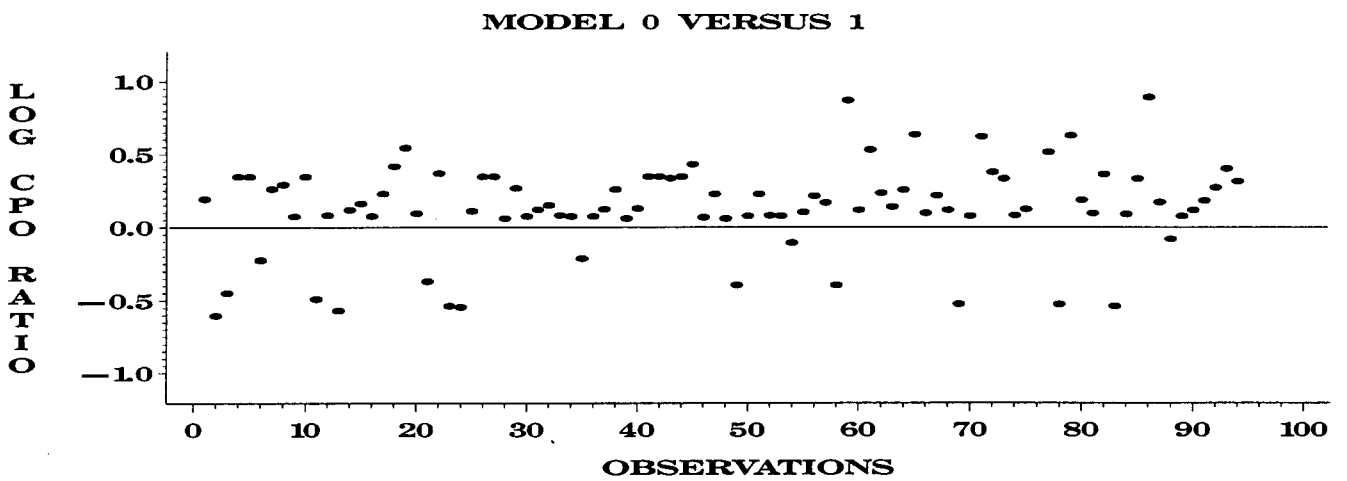


Figure 2: Log CPO ratios for Model 0 versus Model 1.