Bayesian Semiparametric Inference for the Accelerated Failure Time Model

Lynn Kuo and Bani Mallick
University of Connecticut and Imperial College

Abstract

Bayesian semi-parametric inference is considered for a log-linear model. This model consists of a parametric component for the regression coefficients and a nonparametric component for the unknown error distribution. Bayesian analysis is studied for the case of a parametric prior on the regression coefficients and a mixture-of-Dirichlet-processes prior on the unknown error distribution. A Markov chain Monte Carlo (MCMC) method is developed to compute the features of the posterior distribution. A model selection method for obtaining a more parsimonious set of predictors is studied. The method adds indicator variables to the regression equation. The set of indicator variables represents all the possible subsets to be considered. A MCMC method is developed to search stochastically for the best subset. These procedures are applied to two examples, one with censored data.

Acknowledgments

The authors wish to thank Michael Escobar for pointing out the MacEachern algorithm to them. They also wish to thank Thomas Ferguson and Alan Gelfand for constructive comments and helpful discussions on the paper.

Bayesian Semiparametric Inference for the Accelerated Failure Time Model

Abstract

Bayesian semi-parametric inference is considered for a log-linear model. This model consists of a parametric component for the regression coefficients and a nonparametric component for the unknown error distribution. Bayesian analysis is studied for the case of a parametric prior on the regression coefficients and a mixture-of-Dirichlet-processes prior on the unknown error distribution. A Markov chain Monte Carlo (MCMC) method is developed to compute the features of the posterior distribution. A model selection method for obtaining a more parsimonious set of predictors is studied. The method adds indicator variables to the regression equation. The set of indicator variables represents all the possible subsets to be considered. A MCMC method is developed to search stochastically for the best subset. These procedures are applied to two examples, one with censored data.

Key words and phrases: Censored data; Log linear model; Markov chain Monte Carlo algorithm; Metropolis algorithm; Mixtures of Dirichlet processes; Variable selection.
AMS 1991 subject classifications: 62F15, 62J99.
Abbreviated Title: Bayesian Semiparametric Inference

1. Introduction

In survival analysis, covariates often provide essential information. In this paper we pursue a semiparametric approach for an accelerated failure time model. Let $\mathbf{T} = (T_1, \ldots, T_n)$ be the survival times of n subjects. In the accelerated failure time model, the covariates act multiplicatively on the survival time. We assume for each i,

$$\log T_i = -\mathbf{x}_i \boldsymbol{\beta} + W_i, \tag{1.1}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ is a known vector of covariates for the ith subject and $\boldsymbol{\beta}$ is an unknown column vector of p regression coefficients. The error terms W_i are usually assumed to be independent and identically distributed from a parametric family, for example, the normal, logistic, gamma, Weibull, or the extreme value distribution. Classical analysis of this model can be found in Kalbfleisch & Prentice (1980). The parametric assumption, however, may be too restrictive in applications. Often, it is hard to justify a specific parametric family for the error distribution. Therefore, we take a nonparametric approach to the distribution, denoted by G, of the error terms. This makes the model in (1.1) the usual semiparametric model. The parameter $\boldsymbol{\beta}$ in (1.1) that explains the relationship between the log survival times and the covariates is our primary object of inference. The unspecified function G can be treated as an infinite-dimensional nuisance parameter. Cox's proportional hazard model (1972) is also an attempt to avoid restrictive parametric assumptions. As Oakes (1988) points out: "A practical motivation for consideration of semi-parametric models is to avoid restrictive assumptions about secondary aspects of a problem while preserving a tight formulation for the features of primary concern."

Large sample theory for classical (non-Bayesian) semiparametric inference is given in Bickel *et al.* (1993). Lin and Geyer (1992) develop computational methods using simulated annealing for rank regression procedures often used in semiparametric inference. This paper offers a Bayesian perspective for the semiparametric model. The method allows us to incorporate our prior belief about the regression coefficients as well as the unknown error distribution. The prior can come from past experience, control studies, or subjective judgement. Let π denote the prior for $\boldsymbol{\beta}$. It can be quite arbitrary, for example, a multivariate normal, a truncated multivariate normal, or an improper flat prior. We allow the possibility of flat priors in the absence of having a prior description or in an attempt to avoid being "subjective".

A prior on the family of unknown distributions G can be constructed as in the nonparametric Bayesian literature. We model our uncertainty about G by a stochastic process. The support of this process prior is a large family of distributions. This is a robust approach against misspecification of the error distribution, consequently of the distribution for the survival time T. Christensen and Johnson (1988) model $V_i = \exp(W_i),\ i = 1, \ldots, n$, as a sample from an unknown distribution G chosen from a Dirichlet process (Ferguson 1974) indexed by M and G_0, in symbols, $G \in \mathcal{D}(MG_0)$. The Dirichlet process has the advantage that the choices of the parameters M and G_0 have an easy interpretation, where G_0 represents our prior belief about the mean of the distribution of V_i (also denoted by V) and M indicates the the prior sample size or the degree of concentration of the distribution of V around G_0. The larger the M is, the more concentrated the G is around G_0. The computation of the posterior distribution was, however, considered intractable by Christensen and Johnson (1988) and Johnson and Christensen (1989).

Instead of Dirichlet, we consider a mixture of Dirichlet processes (MDP) as in Antoniak (1974) for the distribution of V. The MDP models a family of random densities. It smooths the Dirichlet process with a continuous known kernel with unknown mixing weights where prior belief can be incorporated. The MDP provides a more flexible model than a Dirichlet. Moreover, the smoothing in the MDP eliminates the difficulty encountered by Christensen and Johnson of having to consider the Bayes estimates in many different subcases according to the compatibility of the data to the prior of $\boldsymbol{\beta}$.

We consider two hierarchical models. The first model, called MDPV, generalizes the Dirichlet process in Christensen and Johnson to the MDP for the distribution of $V = \exp(W)$. The second model, called MDPW, perhaps more direct, assumes MDP for the distribution of W.

In MDPV, we assume the random variables V_i $(\exp(W_i)),\ i = 1, \ldots, n$, are independent and identically distributed given G with the following density:

$$f(v_i|G) = \int k(v_i|\theta_i)G(d\theta_i). \tag{1.2}$$

Moreover, the unknown G is chosen by a Dirichlet process prior with known parameters M and G_0. Lo (1984) discusses various choices of the kernel density $k(v_i|\theta_i)$ that include histogram models, a uniform density over $(0, \theta_i)$, (Brunner and Lo, 1989), an exponential density with parameter θ_i, and a normal density with mean μ_i and standard deviation σ_i. In the last case, we use θ_i to denote the two dimensional vector (μ_i, σ_i). In the no-sample problem, the Bayes estimate of $f(v|G)$ is $\int k(v|\theta)G_0(d\theta)$, denoted by $f(v|G_0)$. Therefore, we choose G_0 so $f(v|G_0)$ represents our prior belief about the mean of the class of the MDP densities. Primarily, we choose the kernel to be a scale parameter family and G_0 with support on R^+ because $v_i > 0$. We also allow the possibility of the kernel to be the location and scale family with mean μ_i and scale σ_i. Then $G_0(d\theta) = G_0(d\mu \times d\sigma)$ is a random measure on $R \times R^+$ or $R^+ \times R^+$ for positive locations. The choice of M reflects the degree of concentration of the random MDP family around $f(v|G_0)$. Given (1.1) and (1.2), we derive the density of t_i, where $t_i = v_i \exp(-\mathbf{x}_i \boldsymbol{\beta})$,

$$f(t_i|\boldsymbol{\beta}, G) = f(v_i|G) \left| \frac{dv_i}{dt_i} \right| = \int \exp(\mathbf{x}_i \boldsymbol{\beta}) k(t_i \exp(\mathbf{x}_i \boldsymbol{\beta})|\theta_i) G(d\theta_i). \tag{1.3}$$

A directed graph representation of this model and related comments are given in Kuo and Mallick (1995).

Now, we consider the second model, MDPW. The random variables $W_i, \quad i = 1, \ldots, n$, are independent given G with $f(w_i|G)$ given as in (1.2), where w_i replaces v_i. In this case we choose G_0 so that $\int k(w|\theta)G_0(d\theta)$ reflects our prior belief about the mean of the random densities of the error term W, where G_0 can be supported on the whole real line. Given that we are more familiar with the distribution of W than V, this model seems to be more natural than the previous one. The random variables T_i are independent given G and $\boldsymbol{\beta}$ with density

$$f(t_i|\boldsymbol{\beta}, G) = f(w_i|G) \left| \frac{dw_i}{dt_i} \right| = \int \frac{1}{t_i} k(\log t_i + \mathbf{x}_i \boldsymbol{\beta}|\theta_i) G(d\theta_i). \tag{1.4}$$

For both models, we assume the densities of the error term or its exponential transformation are unknown and are modeled by the family of MDP. Essentially, we model the link function in the generalized linear model as a random smooth function with monotonicity, i.e., chosen from the MDP. Newton, Czado, and Chapell (1992) model the random link function in binary regression with the Dirichlet processes prior. Mallick and Gelfand (1994), and

Gelfand and Mallick (1995) model the random link function with mixture of beta densities. Other related references include the following. Brunner and Lo (1989) consider mixtures of uniforms to obtain monotonicities without smoothness; Dykstra and Laud (1981) and Lo and Weng (1989) consider scale mixtures of gamma processes on the space of hazard functions.

Having defined our model for independent and identically distributed observations $\mathbf{t} = (t_1, \ldots, t_n)$ given the unknown parameters $\boldsymbol{\beta}$ and the unknown function G (a component of the random link function), we desire to obtain the posterior distribution of $\boldsymbol{\beta}$ and G given the data. Although the posterior distribution can be obtained explicitly using the machinery of Lo (1984) and then evaluated with the Monte Carlo method of Kuo (1986a), we propose using the Markov chain Monte Carlo (MCMC) procedure to simulate variates from the posterior distribution in this paper. This is because we think the iterative MCMC procedure has the potential to be more efficient than the method of Kuo that is an importance sampling noniterative procedure. Moreover, the marginal posterior distribution of $\boldsymbol{\beta}$, that is our primary object of inference, can be obtained directly from the MCMC method.

To apply the MCMC algorithm, we need to know the conditional distribution of G given $\boldsymbol{\beta}$ and $\mathbf{t}$ and the conditional distribution of $\boldsymbol{\beta}$ given G and $\mathbf{t}$. Theoretically, we desire to sample G and $\boldsymbol{\beta}$ iteratively between the two conditional distributions. Doss (1994) discusses a method for generating G that is based on the Sethuraman and Tiwari (1982) representation. Doss also provides an approximation to G that truncates the infinite sum representation to a finite sum. However, it remains a difficult problem to decide how many terms in the finite sum is sufficiently large. Instead of constructing G that is infinite dimensional, we consider the posterior measure with G integrated out. All the posterior inference on G can be obtained from this posterior measure. It is a mixture of Dirichlet processes and can be summarized with a transitional measure (a Dirichlet distribution) and an n dimensional mixing distribution (cf. Antoniak (1974), Lo (1984) and Kuo (1986b)). Therefore, instead of sampling G given $\boldsymbol{\beta}$ and $\mathbf{t}$, we sample the n dimensional variates $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ from the mixing distribution as in Escobar (1994). This simplifies the problem from potentially infinite dimensional to n dimensional. Our algorithm can be considered as a collapsed Gibbs sampler discussed in Liu (1994). Given $\boldsymbol{\theta}$ and $\mathbf{t}$, we can sample $\boldsymbol{\beta}$ from the conditional density $f(\boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{t})$ using the the Metropolis algorithm (1952) or the adaptive rejection Metropolis algorithm of Gilks, Best, and Tan (1994). Then we continue iteratively to sample new $\boldsymbol{\theta}$ given $\boldsymbol{\beta}$ and $\mathbf{t}$, etc. MacEachern (1994), and Bush and MacEachern (1996) propose a refinement of the Escobar algorithm, where the new $\boldsymbol{\theta}$'s are updated by using clusters of identical θ's and new values. We implemented the more efficient MacEachern method.

In addition to completely observed event times t_i, we consider censored observations, including right, left, or interval censored. Let $\mathbf{t}'$ denote the data set which includes n completely observed failure times and m censored failure times. The posterior evaluation is obtained by data augmentation, where latent variables are generated from the censored distribution. The MCMC algorithm consists of generating the latent variables denoted by $\mathbf{z}$ (m dimensional) given $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\mathbf{t}'$, and generate $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ given $\mathbf{t}', \mathbf{z}$, iteratively. The former is done using the predictive distribution of $\mathbf{z}$ and the rejection method. The latter has already been discussed because $(\mathbf{t}', \mathbf{z})$ can be considered as the complete data set without missing or censored values.

We can also address issues of variable selection where a more parsimonious model using a subset of $\mathbf{x}$ may be desired. The usual method of evaluating Bayes factors or pseudo-Bayes factors can be computational prohibitive when p is large. Therefore, we develop a stochastic search procedure that searches for the most probable subset of predictors. We need to modify the search procedures in Kuo and Mallick (1994) where usual regression models (W has normal distribution) and generalized linear models are considered. The basic idea is that we expand the regression equation (1.1) to an equation that includes all the 2^p models using indicator variables. The stochastic search procedure will automatically search for the most probable model. Let γ_j be a parameter that takes on two values 0 or 1. When $\gamma_j = 1$, it represents that the j^{th} predictor is included in model building; when $\gamma_j = 0$, otherwise. The expanded regression equation can be written as

$$\log T_i = W_i + \sum_j -x_{ij}\beta_j\gamma_j. \tag{1.5}$$

Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. We select the subset with the highest posterior probability $\Pr(\boldsymbol{\gamma}|\mathbf{t})$ that can be easily evaluated using a MCMC algorithm. As demonstrated in Kuo and Mallick (1994), the MCMC algorithm searches for the most probable configuration $\boldsymbol{\gamma}$ very efficiently.

Although the literature on Dirichlet processes and on MDP is extensive (cf. Ferguson, Phadia, and Tiwari (1992)), there are only a few papers on nonparametric Bayesian analysis with covariates. Kalbfleisch (1978) proposes gamma processes for the Cox (1972) proportional hazard model. Burridge (1981) extends the Kalbfleisch work by examining grouped data and arbitrary independent increment process priors. Christensen and Johnson (1988) and Johnson and Christensen (1989) study accelerated failure time models with Dirichlet process priors. Clayton (1991) develops a Monte Carlo method for frailty models. West, Müller, and Escobar (1994), Erkanli, Müller, and West (1992), and Erkanli, Stangl, and Müller (1993) apply MDP for curve fitting, Bayesian regression estimation, and analysis of ordinal data. Brunner (1995) considers regression problems with error terms that have symmetric unimodal densities. Mukhopadhyay and Gelfand (1997) extend the generalized linear models to Dirichlet process mixed generalized linear models to add flexibility in modeling.

Section 2 develops the Gibbs sampling algorithm. Section 3 studies prediction of the survival curve. Section 4 discusses the stochastic search procedure for variable selection. Two numerical examples, one with censored failure times, are given in Section 5.

2. Gibbs Sampling

2.1 Completely Observed Data

In this subsection, we consider the data where we observe the survival times exactly. Because the Escobar (1994) algorithm is easier to understand than the MacEachern algorithm (1994), we first derive the MCMC algorithm using the idea of Escobar (1994). Then we give an improved algorithm using the idea of MacEachern (1994). West, Müller, and Escobar (1994), MacEachern and Müller (1994) also provide discussions on the improved algorithm.

For the MDPV model, we have $t_1, \ldots, t_n$ are independent and identically distributed with density (1.3). The prior on G is the Dirichlet process with parameter M and G_0, and the prior density for $\boldsymbol{\beta}$ denoted by π can be quite arbitrary. Moreover, G and $\boldsymbol{\beta}$ are independent in the prior specification. In our numerical example, we assume the prior π is improper with $\pi(\boldsymbol{\beta}) = 1$ for $-\infty < \beta_j < \infty$ and all j. Then, the likelihood function of G and $\boldsymbol{\beta}$ given $\mathbf{t}$ is

$$L(G, \boldsymbol{\beta}|\mathbf{t}) = \prod_{i=1}^{n} \exp(\mathbf{x}_i\boldsymbol{\beta}) \int k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_i) G(d\theta_i).$$

Let δ_{θ_j} denote a measure with point mass concentrated at θ_j. It follows from Antoniak (1974) and Lo (1984),

$$G|\boldsymbol{\beta}, \mathbf{t} \in \int \mathcal{D}\left(MG_0 + \sum_{j=1}^{n} \delta_{\theta_j}\right) dH(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{t}), \tag{2.1}$$

where

$$dH(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{t}) \propto \prod_{i=1}^{n} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_i) \left(MG_0 + \sum_{j=1}^{i-1} \delta_{\theta_j}\right) (d\theta_i).$$

Note that t_i is related to v_i and $\boldsymbol{\beta}$ deterministically. Therefore, we can use $\boldsymbol{\beta}$ and $v_1, \ldots, v_n$ as the conditioning variables in the above expression instead of $\boldsymbol{\beta}$ and $\mathbf{t} = (t_1, \ldots, t_n)$. Let $\boldsymbol{\theta}^{(-i)} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$. Observe the conditional density of θ_i given $\boldsymbol{\theta}^{(-i)}, \boldsymbol{\beta}, \mathbf{t}$ is

$$f(\theta_i|\boldsymbol{\theta}^{(-i)}, \boldsymbol{\beta}, \mathbf{t}) = \frac{k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_i) MG_0(d\theta_i) + \sum_{j \neq i} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_j)\delta_{\theta_j}(d\theta_i)}{a_i(\boldsymbol{\beta}) + \sum_{j \neq i} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_j)}, \tag{2.2}$$

where

$$a_i(\boldsymbol{\beta}) = \int k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta) MG_0(d\theta).$$

Moreover, the conditional density of $\boldsymbol{\beta}$ given $\boldsymbol{\theta}, \mathbf{t}$ is

$$f(\boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{t}) \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^{n} \exp(\mathbf{x}_i\boldsymbol{\beta}) k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_i). \tag{2.3}$$

Therefore, we summarize the MCMC algorithm as follows:

1. Generate $\boldsymbol{\theta}$ according to $dH(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{t})$. This is done by starting with $i = 1$, selecting θ_i from (2.2) sequentially for $i = 1, \ldots, n$. Let $b_i(\boldsymbol{\beta}) = \sum_{j \neq i} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_j)$. Then

   - with probability $a_i(\boldsymbol{\beta})/(a_i(\boldsymbol{\beta}) + b_i(\boldsymbol{\beta}))$, generate θ_i from the density that is proportional to $k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta) MG_0(d\theta)$;

   - with probability $b_i(\boldsymbol{\beta})/(a_i(\boldsymbol{\beta}) + b_i(\boldsymbol{\beta}))$, generate θ_i from the set $\{\theta_j : j \neq i\}$ using probabilities proportional to $k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_j)$.

2. Generate $\boldsymbol{\beta}$ given $\boldsymbol{\theta}, \mathbf{t}$ from (2.3) by the Metropolis algorithm. This can be done sequentially for $\boldsymbol{\beta}_l$, $l = 1, \ldots, p$, from the density $f(\beta_l|\boldsymbol{\beta}^{(-l)}, \boldsymbol{\theta}, \mathbf{t})$, where $\boldsymbol{\beta}^{(-l)} = (\beta_1, \ldots, \beta_{l-1}, \beta_{l+1}, \ldots, \beta_p)$.
3. Cycle through steps 1 and 2 until convergence. Moreover, we can also construct multiple chains by replicating the iterations using independent random starting values.

A more efficient algorithm can be obtained by means of the idea of MacEachern (1994), where updating θ's is done in clusters. Let $\phi_1^{(-i)}, \dots, \phi_{K^{(-i)}}^{(-i)}$ denote the values of distinct θ's in $\boldsymbol{\theta}^{(-i)}$ with multiplicities $n_1^{(-i)}, \dots, n_{K^{(-i)}}^{(-i)}$. Then observe

$$f(\theta_i|\boldsymbol{\theta}^{(-i)}, \boldsymbol{\beta}, \mathbf{t}) = \frac{k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_i) M G_0(d\theta_i) + \sum_{j=1}^{K^{(-i)}} n_j^{(-i)} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\phi_j^{(-i)}) \delta_{\phi_j^{(-i)}}(d\theta_i)}{a_i(\boldsymbol{\beta}) + \sum_{j=1}^{K^{(-i)}} n_j^{(-i)} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\phi_j^{(-i)})}, \tag{2.4}$$

This suggests that we can define a configuration index $\mathbf{s} = (s_1, \dots, s_n)$, where $s_i = j$ if $\theta_i = \phi_j^{(-i)}$ for $j = 1, \dots, K^{(-i)}$, and $s_i = 0$ if θ_i is a new draw from the density proportional to $k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta) M G_0(d\theta)$. From (2.4), we can define

$$\Pr(s_i = j|t, \boldsymbol{\theta}^{(-i)}, \mathbf{s}^{(-i)}, K^{(-i)}) = q_{i,j}$$

where

$$q_{i,j} = \begin{cases} c a_i(\boldsymbol{\beta}) & \text{if } j = 0, \\ c n_j^{(-i)} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\phi_j^{(-i)}) & \text{if } j > 0, \end{cases} \tag{2.5}$$

with c being the normalizing constant. Given the old $\boldsymbol{\theta}$ ($\boldsymbol{\phi}$), we can update the $\mathbf{s}$ using (2.5) for each i, $i = 1, \dots, n$. The new $\mathbf{s}$ not only determines the structure of the clusters and also determines a new K that is the number of distinct values in $\boldsymbol{\phi}$. Let $I_j = \{i : s_i = j\}$, the set of indices of observation in group j, where $j \geq 1$. Then we have

$$f(\boldsymbol{\phi}|\mathbf{t}, \mathbf{s}, K) \propto \prod_{i \in I_j} k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\phi) G_0(d\phi). \tag{2.6}$$

Therefore, the improved algorithm consists of replacing step 1 with 1a and 1b:

- 1a. Generate new $\mathbf{s}$ given old $\boldsymbol{\theta}$ ($\boldsymbol{\phi}$) using (2.5), for each $i = 1, \dots, n$. The conditioning variables are held fixed for all the steps 1 through n in 1a. That is, the new $\mathbf{s}$ is updated from the old $\boldsymbol{\phi}$. It also determines a new K, the total number of distinct groups.

- 1b. Given $\mathbf{s}$ and K, generate a new set of $\boldsymbol{\phi}$ as follows. If $s_i = 0$, then draw a new ϕ with the density that is proportional to $k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\phi) M G_0(d\phi)$; if $s_i = j$, then we draw a new ϕ, denoted by ϕ_j, from the density that is proportional to (2.6). We only need to draw one ϕ for all the indices in group I_j. Having exhausted all the indices, then reconstruct $\boldsymbol{\theta}$ using the updated $\boldsymbol{\phi}$ and $\mathbf{s}$.

For the MDPW model, we follow similar steps 1-3, where $k(t_i \exp(\mathbf{x}_i\boldsymbol{\beta})|\theta_i)$ is replaced by $k(\log t_i + \mathbf{x}_i\boldsymbol{\beta}|\theta_i)$ and (2.3) is replaced by

$$f(\boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{t}) \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^{n} k(\log t_i + \mathbf{x}_i\boldsymbol{\beta}|\theta_i). \tag{2.7}$$

2.2 Censored Data

Censored data occurs frequently in the analysis of failure time data. We use right censored data for illustration. The same principle applies to the left or interval censored data. In addition to n survival times described previously, we observe m right censored survival times denoted by $(t_{n+1}, \ldots, t_{n+m})$. That is, we know the survival time for the ith subject is bigger than t_i, for $i = n+1, \ldots, n+m$. We use $\mathbf{t}' = (t_1, \ldots, t_n, t_{n+1}^+, \ldots, t_{n+m}^+)$ to denote the data set. In the derivation of the MCMC algorithm, we use data augmentation, where we generate the latent variable $\mathbf{z} = (z_{n+1}, \ldots, z_{n+m})$ by the rejection method from the truncated density of $f(T_{n+1}, \ldots, T_{n+m}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{t}')$, where $T_i > t_i$ for $i = n+1, \ldots, n+m$, and $\boldsymbol{\theta}$ is a $n+m$ dimensional row vector. Then the complete data set is $\mathbf{t} = (t_1, \ldots, t_n, z_{n+1}, \ldots, z_{n+m})$. We apply the previous algorithm with size $n+m$ to generate $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n+m})$ given $\boldsymbol{\beta}, \mathbf{t}$ and to generate $\boldsymbol{\beta}$ given $\boldsymbol{\theta}, \mathbf{t}$.

Let us summarize the Gibbs algorithm:

0. Generate the complete data set $\mathbf{t} = (t_1, \ldots, t_n, z_{n+1}, \ldots, z_{n+m})$ given $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and and $\mathbf{t}'$. This is done by fixing the first n components to be the completely observed failure times and generating the latent variable $\mathbf{z} = (z_{n+1}, \ldots, z_{n+m})$ given $\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{t}'$. For each $i = n+1, \ldots, n+m$, generate z_i given $\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{t}'$. First generate θ_i' from the predictive density of new θ_i given old $\boldsymbol{\theta}$. This is done by generating a new θ from G_0 with probability $M/(M+n+m)$ and uniformly from the set of $\{\theta_1, \ldots, \theta_{n+m}\}$ with probability $(n+m)/(M+n+m)$. Then generate v_i given θ_i' according to the $k(v|\theta_i')$ and compute $d_i = v_i \exp(-\mathbf{x}_i\boldsymbol{\beta})$. Let $z_i = d_i$, if $d_i > t_i$; repeat this process otherwise, until $d_i > t_i$.

1. Given $\mathbf{t} = (t_1, \ldots, t_n,\ z_{n+1}, \ldots, z_{n+m})$, generated in step 0 and $\boldsymbol{\beta}$, apply step 1 of the previous subsection with size $n+m$ for $i = 1, \ldots, n+m$ to generate $\boldsymbol{\theta}$ of size $n+m$.

2. Given $\boldsymbol{\theta}$ and $\mathbf{t}$, apply step 2 of the previous subsection to generate $\boldsymbol{\beta}$, where the product in (2.3) has $n+m$ terms with z_i replacing t_i for $i = n+1, \ldots, n+m$. Cycle through steps

0, 1, and 2 until convergence.

For the MDPW model, we follow similar steps 0-3, with the kernel change as discussed at the end of previous subsection and in step 0, change v_i to w_i and $d_i = \exp(-\mathbf{x}_i\boldsymbol{\beta} + w_i)$.

3. Prediction of the Survival Curve

Suppose we wish to predict the survival curve for a new subject with covariate $\mathbf{x}_0$, a $p \times 1$ dimensional vector. We propose two methods, one by sampling variates from the predictive density. The other uses the functional form of the kernel for evaluation. Let us consider the MDPV model with completely observed data.

Method 1: After step 2 in the MCMC algorithm, we add step $2'$: generate θ_0 from G_0 with probability $M/(M+n)$, and uniformly from the set $\theta_1, \ldots, \theta_n$ with probability $n/(M+n)$. Then generate v_0 given θ_0 from the kernel $k(v|\theta_0)$. Let $t_0 = v_0 \exp(-\mathbf{x}_0\boldsymbol{\beta})$. Then the survival curve is constructed from the samples of t_0 generated in the MCMC algorithm.

Method 2: For this method, we only illustrate it with the normal kernel with mean μ and standard deviation σ. The same idea applies to other kernels. The parameter θ in the previous discussion is actually a two dimensional vector (μ, σ) for this example. The survival function for the new subject with covariate $\mathbf{x}_0$ is

$$\begin{aligned} S_{G,\boldsymbol{\beta}}(s) &= \int_s^\infty \int \exp(\mathbf{x}_0\boldsymbol{\beta}) k(s\exp(\mathbf{x}_0\boldsymbol{\beta})|\mu,\sigma) G(d\mu \times d\sigma) ds \\ &= 1 - \int \Phi(s\exp(\mathbf{x}_0\boldsymbol{\beta})/\sigma - \mu/\sigma) G(d\mu \times d\sigma), \end{aligned} \tag{3.1}$$

where Φ denotes the standard normal cumulative distribution function. The second equality in (3.1) is obtained by changing the order of integration and evaluate the inner integral. Let $\boldsymbol{\beta}^{(r,q)}$ and $\theta_j^{(r,q)}$ denotes the samples generated in the rth replication and qth iteration of the MCMC. We obtain an estimate of the survival function evaluated at s from a draw of the MCMC using (3.1) and (2.1)

$$\widehat{S}^{(r,q)}(s) = 1 - \int \frac{M}{M+n} \Phi\left(\frac{s\exp(\mathbf{x}_0\boldsymbol{\beta}^{(r,q)})}{\sigma} - \frac{\mu}{\sigma}\right) G_0(d\mu \times d\sigma) \tag{3.2}$$

$$-\frac{1}{M+n} \sum_{j=1}^n \Phi\left(\frac{s\exp(\mathbf{x}_0\boldsymbol{\beta}^{(r,q)})}{\sigma_j^{(r,q)}} - \frac{\mu_j^{(r,q)}}{\sigma_j^{(r,q)}}\right). \tag{3.3}$$

Then we can obtain a final estimate of the survival function evaluated at s by averaging $\widehat{S}^{(r,q)}(s)$ according to the Gelman-Rubin (1992) method. We can also check the convergence of the Gibbs sample by the Gelman-Rubin method. To plot the survival curve, we evaluate the survival curve at grids $s_1, \ldots, s_k$ for k sufficiently large.

Now let us consider the MDPW model. For method 1, the fully sampling-based approach, in the step $2'$, we change v_0 to w_0 and let $t_0 = \exp(w_0 - \mathbf{x}_0\boldsymbol{\beta})$. For method 2, we change (3.1) to

$$\begin{aligned} S_{G,\boldsymbol{\beta}}(s) &= \int_s^\infty \int \frac{1}{s} k(\log s + \mathbf{x}_0\boldsymbol{\beta}|\mu,\sigma) G(d\mu \times d\sigma) du \\ &= 1 - \int \Phi\left(\frac{\log s + \mathbf{x}_0\boldsymbol{\beta} - \mu}{\sigma}\right) G(d\mu \times d\sigma). \end{aligned} \tag{3.4}$$

Then we evaluate this survival function similarly to (3.3) by replacing $s\exp(\mathbf{x}_0\boldsymbol{\beta}^{(r,q)})$ with $\log s + \mathbf{x}_o\boldsymbol{\beta}^{(r,q)}$.

For the censored data problem, we need to change n to $n+m$ for both methods.

4. Subset Selection

In this section, we develop a stochastically search procedure that selects the best subset of predictors. Variable selection is an important issue because we would like to explain the relationship between t and $\mathbf{x}$ as simply as possible and we desire to identify important predictors for t. Consequently, we may improve the accuracy in prediction and reduce the cost in collecting future data.

Consider the expanded regression equation that was discussed briefly in the introduction

$$\log T_i = W_i + \sum_j -x_{ij}\beta_j\gamma_j, \tag{4.1}$$

where γ_j is a parameter that takes on two values 1 or 0, indicating whether or not the j^{th} predictor is included in the model. Note that the expanded equation (4.1) includes all the 2^p possible submodels. Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. We select the subset with the highest posterior probability $\Pr(\boldsymbol{\gamma}|t)$ that can be easily evaluated using the following MCMC algorithm. The highest posterior probability model is considered to be the best model, because it is the Bayes solution to a 0-1 loss function with 0 loss for the correct model and 1 otherwise.

Other criteria for choosing the best model such as minimizing mean squared errors can easily be incorporated by extending our MCMC algorithm. In the prior specification of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and G, we assume they are independent. The parameter $\gamma_j = 1$ with probability p_j (known) and 0 with probability $1 - p_j$; moreover, γ_j, $j = 1, \ldots, p$ are independent. The prior for $\boldsymbol{\beta}$ and G are unchanged from the previous sections. Other choices of priors including the normal distribution for $\boldsymbol{\beta}$, $\boldsymbol{\beta}$ depending on $\boldsymbol{\gamma}$, can also be considered as in Kuo and Mallick (1994).

In describing the stochastic search algorithm, we need to specify the conditional densities $f(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{t})$, $f(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \mathbf{t})$, and $f(\boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{t})$, where $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are exactly the same variables as in the previous section. We describe the MCMC algorithm for the MDPV model with no censored observations. Straightforward revisions apply to the MDPW model and data with censored observations. Let $\boldsymbol{\rho} = (\beta_1\gamma_1, \ldots, \beta_p\gamma_p)^T$. The MCMC algorithm consists of:

1. Generate $\boldsymbol{\theta}$ from $f(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{t})$: follow Step 1 (or 1a and 1b) in Subsection 2.1, where $\boldsymbol{\beta}$ is replaced by $\boldsymbol{\rho}$.

2. Generate $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \mathbf{t})$ by the Metropolis algorithm, where

$$f(\boldsymbol{\beta}|\boldsymbol{\theta},\ \boldsymbol{\gamma},\ \mathbf{t}) \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^{n} \exp(\mathbf{x}_i\boldsymbol{\rho}) k(t_i \exp(\mathbf{x}_i\boldsymbol{\rho})|\theta_i).$$

3. Let $\boldsymbol{\rho}_{j1} = (\beta_1\gamma_1, \ldots, \beta_{j-1}\gamma_{j-1}, \beta_j, \beta_{j+1}\gamma_{j+1}, \ldots, \beta_p\gamma_p)^T$ and $\boldsymbol{\rho}_{j0} = (\beta_1\gamma_1, \ldots, \beta_{j-1}\gamma_{j-1}, 0, \beta_{j+1}\gamma_{j+1}, \ldots, \beta_p\gamma_p)^T$. Then generate γ_j for $j = 1, \ldots, n$ in random order from the following density

$$\gamma_j = \begin{cases} 1 \text{ with } \Pr \propto p_j \prod_{i=1}^{n} \exp(\mathbf{x}_i\boldsymbol{\rho}_{j1}) k(t_i \exp(\mathbf{x}_i\boldsymbol{\rho}_{j1})|\theta_i). \\ 0 \text{ with } \Pr \propto (1 - p_j) \prod_{i=1}^{n} \exp(\mathbf{x}_i\boldsymbol{\rho}_{j0}) k(t_i \exp(\mathbf{x}_i\boldsymbol{\rho}_{j0})|\theta_i). \end{cases} \tag{4.2}$$

5. Numerical Examples

To illustrate the modeling in previous sections we study a data set in Feigl and Zelen (1965). Aitkin *et al.* (1989) also list the data (p. 35) and present an analysis of it using generalized linear models. The data are concerned with 33 patients suffering from leukemia. For each patient, the white blood cell (WBC) count in thousands at the time of diagnosis and the time (in weeks) to death from initial diagnosis were recorded, together with the test result on the AG-factor at the time of diagnosis (positive or negative). So there are two covariates, the WBC count and the factor AG-group with two levels. The response variable is survival time in weeks.

We model V by the MDPV model. For the kernel density our choice is $Normal(\mu, \sigma^2)$. Here we should note that we can't take high σ values as then for lower values of t the kernel can give some mass towards negative values of t, so the survival probability at the point 0 may not be 1. After some experiments we found $\sigma = .1$ is a good choice. We take the base line distribution G_0 as exponential with parameter 1 for μ and degenerate with point mass at .1 for σ. We vary the values of prior precision M from low (.05) to high (5000) to show the sensitivity of the analysis to different precision parameters.

We ran all the sampling based methods described in sections 2 and 3 to obtain the posterior estimates of the parameters. Here we have three parameters: the intercept β_0, the coefficient for the WBC count β_1, and the coefficient for the AG-group β_2. Table 1 lists the Bayes estimates for different prior precisions approximated by the MCMC method using 10,000 iterations and 5 replications.

We computed the survival curves for the two groups, AG-positive and AG-negative, using method 2 in section 3, because for the normal kernel we can derive almost the exact expression with a univariate numerical integration. We drew the predictive survival curves from left to right in Figure 1 for new patients with upper, median, and lower quantiles of WBC counts for the two groups. That is, for the AG-negative group, the WBC count is fixed at 31, 9, 4.4; for the positive group it is 35, 10.5, 5.4. We use solid lines for the AG-negative group, and dotted line for the AG-positive group. From the plot it's clear the survival probability decreases more sharply for higher WBC counts and it also drops more sharply for the AG-negative patients than the positive ones. Also interesting to note is that for the AG-negative group the survival probabilities are near to each other for the two higher WBC counts but significantly different for the lowest one, which signifies the effect of WBC counts on the AG-negative group. As M increases, the survival curves are becoming smoother and approach an exponential curve as expected.

We also apply our method in Section 2.2 to a data set with censored observations. This data set with censored observations is modified from the previous data set by uniformly sampling a survival time U_i from 0 to t_i, then replacing the t_i with U_i as the censored observation for $i = 1, \ldots, 3$, and $i = 18, 19$. That is we modify 5 observations; first 3 observations are from the AG-positive group and the other two observations are from the AG-negative group. The censored observations are given in the last column of Table 3.

The Bayes estimates of $\boldsymbol{\beta}$ are shown in Table 2 for each of the two data sets, one is the original one, the other is the modified one with censored data. For this table, we run the Gibbs samplers with the precision parameter M unknown. Like Escobar and West (1995), we put a gamma prior for M, that is for δ, $\rho > 0$, $M \sim$ gamma(δ, ρ). We take $\delta = \rho = 1$. The Gibbs samplers in previous sections can be extended straightforwardly for this case. The drawing of M is done as explained in Escobar and West. The MCMC algorithm was first run in order to ascertain at which point suitable convergence was obtained and then a further chain was run in order to obtain 3000 samples for inference.

Additionally, for the censored data, we give the Bayes estimates, on a log scale, for the true survival times for those five patients for which we sampled censoring times. These are given in Table 3.

Predictive survival curves for new patients with varying sets of covariate values for the unknown M are shown in Figures 2-5. Figures 2 and 3 are for the original uncensored data while Figures 4 and 5 are for the censored data. The sets of covariates are fixed as in Figure 1. The figures suggest similar patterns as the ones in Figure 1 with M known. The survival curves with negative AG factor drop faster than the ones with positive AG factor. Moreover, the survival curves with the censored data set drop slower than the ones with the original data set.

6. Final Comments

We have described a Bayesian approach to a semiparametric model where we assume covariates act multiplicatively on the survival time. Instead of the log linear relationship studied here, we can consider other transformations of t used in generalized linear models. Our methodology can be modified straightforwardly to incorporate these known relationships between the survival time and covariates. Moreover, we can extend our Gibbs sampler with minor modification to treat the additive model, where the known relationship can be given in componentwise, for example, $y_i = \sum_{j=1}^{k} f_j(x_{ij}\beta_j) + w_i$, where $\{f_j\}$ are a set of known functions.

References

Aitkin, M., Anderson, D., Francis, B., and Hinde, J. (1989). *Statistical Modeling in GLIM.* Oxford: Oxford University Press.

Antoniak, C.E. (1974). Mixture of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* 2, 1152-1174.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: The Johns Hopkins University Press.

Brunner, L.J. (1995). Bayesian linear regression with error terms that have symmetric unimodal densities. *J. Nonparametric Statist.* 4, 335-348.

Brunner, L. J. and Lo, A. (1989). Bayes methods for a symmetric unimodal density and its mode. *Ann. Statist.* 17 1550-1566.

Burridge, J. (1981). Empirical Bayes analysis of survival time data. *J. R. Statist. Soc.*, B 43, 65-75.

Bush, c.A., and MacEachern, S.N. (1996). A semiparametric Bayesian model for randomized block designs. *Biometrika*, 83, 275-285.

Christensen, R., and Johnson, W. (1988). Modeling accelerated failure time with a Dirichlet process. *Biometrika,* 77, 693-704.

Clayton, D. G. (1991). A Monte Carlo methods for Bayesian inference in frailty methods. *Biometrics* 47, 467-485.

Cox, D.R. (1972). Regression models and life-tables (with discussion), *J. R. Statist. Soc.* B 34, 187-220.

Doss, H. (1994). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *Ann. Statist.,* 22, 1763-1786.

Erkanli, A., Müller, P., and West, M. (1992). Curve fitting by a mixture of Dirichlet process model. Technical Report 92-A09, Institute of Statistics and Decision Sciences, Duke University.

Erkanli, A., Stangl, D., and Müller, P. (1993). A Bayesian analysis of ordinal data using mixtures. *ASA Proc. Bayes*, 51-56.

Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *J. Am. Statist. Assoc.* 89, 268-277.

Escobar, M.D., and West, M. (1995). Bayesian density estimation and inference using mixtures, *J. Am. Statist. Assoc.* 90, 577-588.

Feigl, P. and Zelen, M. (1965). Estimation of exponential probabilities with concomitant information. *Biometrics,* 21, 826-838.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1, 209-230.

Ferguson, T.S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics,* eds. H. Rizvi and Rustagi, New York: Academic Press, 287-302.

Ferguson, T.S., Phadia, E.G., and Tiwari, R.C. (1992), Bayesian nonparametric inference *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, eds Ghosh, M. and Pathak, P.K., Hayward: Institute of Mathematical Statistics, 127-150.

Geisser, S., and Eddy, W. (1979), A predictive approach to model selection. *J. Am. Statist. Assoc.* 74, 153-160.

Gelfand, A., and Mallick, B. (1995), Bayesian analysis of proportional hazard models built from monotone functions. *Biometrics,* 51, 843-852.

Gelman, A.E., and Rubin D. (1992). Inference from iterative simulation using multiple sequences, *Statist. Sci.* 7, 457-472.

Gilks, W.R., Best, N.G., and Tan, K.K.C. (1994). Adaptive rejection Metropolis sampling for Gibbs sampling, preprint.

Johnson, W. & Christensen, R. (1989). Nonparametric Bayesian analysis of the accelerated failure time model, *Statistics & Probability Letters*, 7, 179-184.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *J. R. Statist. Soc.* B, 40, 214-221.

Kalbfleisch, J. D., and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kuo, L. (1986a). Computation of mixtures of Dirichlet processes. *Siam. J. Sci. Comput.* 7, 60-71.

Kuo, L. (1986b). A note on Bayes empirical Bayes estimation by means of Dirichlet processes, *Statistics & Probability Letters,* 4, 145-190.

Kuo, L. and Mallick, B. (1994). Variable selection for regression models. Technical Report 94-26, Statistics Department, University of Connecticut.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semiparametric linear regression with censored data. *Journal of Computational and Graphical Statistics,* 1, 77-90.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J. Ameri. Statist. Assoc.,* 89, 958-966.

Lo, A. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *Ann. Statist.,* 12, 351-357.

MacEachern, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput.*, 23, 727-741.

MacEachern, S. N., and Müeller, P. (1994). Estimating mixture of Dirichlet process models. Technical Report 94-11, Institute of Statistics and Decision Sciences, Duke University.

Mallick, B. K., and Gelfand, A. E. (1994). Generalized linear models with unknown link functions. *Biometrika,* 81, 237-246.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equations of state calculations by fast computing machines, *J. Chemical Physics,* 21, 1087-1091.

Mukhopadhyay, S. and Gelfand, A. (1997). Dirichlet process mixed generalized linear models. *J. Amer. Statist. Assoc.* (to appear). Technical Report 95-11, Statistics Department, University of Connecticut.

Newton, M.A. Czado, C. and Chapell, R. (1996). Bayesian inference for semiparametric binary regression. *J. Amer. Statist. Assoc.*, 91, 142-153.

Oakes, D. (1988). Semi-parametric models. In *Encyclopedia of Statistical Sciences* Volume 8 eds: Kotz, S. and Johnson, N. L., New York: Wiley, 367-369.

Sethuraman, J. and Tiwari, R.C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In *Statistical Decision Theory and Related Topics III 2,* eds. Gupta and Berger, New York: Academic Press.

West, M., Müller, P. & Escobar, M.D. (1994). Hierarchical priors and mixture models with application in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley.* eds: A.F.M. Smith and P. Freeman, 363-368.

Lynn Kuo
Department of Statistics
University of Connecticut
Storrs, CT 06269-3120, U.S.A.

Bani Mallick
Department of Mathematics
Imperial College
London SW7 2BZ, United Kingdom

Table 1. Bayes estimates for the coefficients for the model

$\log t = -\beta_0 - \beta_1 WBC - \beta_2 AG + \log V$

estimate	β_0	β_1	β_2
M=.05	-5.32	1.15	.84
M=50	-5.43	1.07	1.01
M=500	-5.84	1.02	1.2
M=5000	-5.95	1.00	1.25

Table 2. Bayes estimates for $\boldsymbol{\beta}$ for both data sets with $M \sim \Gamma(1,1)$

estimate	β_0	β_1	β_2
Uncensored	-5.31	1.11	.91
Censored	-5.52	1.46	.77

Table 3. Bayes estimates for the true survival times from the modified data set

	Estimate	True value	Observed censored data
patient 1	4.62	4.17	4.10
patient 2	4.21	5.05	3.26
patient 3	5.17	4.60	4.42
patient 18	3.29	4.02	2.11
patient 19	3.01	4.17	1.20