Nonparametric Bayesian Estimation Based on Beta Prior in Cure Model

Xiaobing Zhao and Cuiliu Xiao
School of Science, Jiangnan University
1800 Lihu Road, Wuxi, Jiangsu Province, CHINA 214122
Email: maxbzhao@hotmail.com

Xian Zhou
Department of Actuarial Studies
Macquarie University, NSW, Australia
Email: xzhou@efs.mq.edu.au

***Abstract*—This paper is to study nonparametric Bayesian estimation for a proportional hazards model with "long-term survivors". The cumulative hazards function is modeled by a beta process, and the priors of the cure rate and coefficient of covariates can be improper distributions under the proposed model. The posterior estimators of the cure rate, the coefficient for covariates and the survival function are estimated from the cases of discrete-time, continuous-time and grouped survival data. A set of leukemia data are re-analyzed to illustrate the proposed model and statistical inference via a Markov chain Monte Carlo (MCMC) algorithm with Gibbs sampling.**

I. Introduction

Survival models with long-term survivors, often referred to as cure (or mixture) models, are increasingly popular in analyzing data from clinical trial of cancer treatment and other medical studies. The cure model has been used for time-to-event data from many types of cancers, including breast cancer, non-Hodgkin lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer. For these cancers, a significant proportion of patients are "long-term survivors" or "cured", who will not die of the cancer concerned. In this model, the survival function for the population is given by

$$S(t) = 1 - pF_0(t), \tag{1}$$

where $p \in (0, 1]$ is the proportion of the susceptibles (non-cured) in the population and $F_0(t)$ is a proper cumulative distribution function (cdf). One can easily find that $S(t)$ is an "improper" survival function in the sense that $S(\infty) > 0$ (if $p < 1$). This model has been extensively discussed in the literature by many authors, including Farewell (1982), Kuk and Chen (1992), among others. The book by Maller and Zhou (1996) gives an extensive discussion of frequentist methods of inference for this cure model.

Although cure model (1) is attractive and widely used, it has some drawbacks. For example, if covariates are modeled through p, it does not have a proportional hazards (PH) structure, which is a desirable property in survival analysis models with covariates. Moreover, as pointed out by Ibrahim *et al.* (1999), model (1) yields improper posterior distributions, including the case of a noninformative (uniform) prior. This is a crucial drawback since it makes it difficult to carry out Bayesian analysis with noninformative history – which is quite common in practice. More recently, a Bayesian analysis is introduced by Chen *et al.* (1999) based on an alternative cure rate model, referred to as "non-mixture model", given by

$$S(t) = \exp(-\theta F_0(t)), \tag{2}$$

where $\theta \in [0, \infty)$ and $F_0(t)$ is a proper cdf. This model is extensively investigated by Ibrahim *et al.* (1999) and Chen *et al.* (2002) using Bayesian method. Model (2), however, does not have a PH structure either if the standard survival function is modeled by a PH structure.

In order to overcome the above drawback, Maller and Zhou (1996) raised an idea to incorporate Cox PH models with cure models in the form

$$\lambda(t) = \lambda_0(t) \exp(z^\top \beta), \tag{3}$$

where $\lambda_0(t)$ is the baseline hazard of an "improper" survival function and z is covariate. When the baseline $\lambda_0(t)$ corresponds to $1 - pF_0(t)$, the survival function of (3) is

$$S(t) = (1 - pF_0(t))^{\exp(z^\top \beta)}. \tag{4}$$

This improper PH model, which has been further investigated by Zhao and Zhou (2006, 2008), can overcome the drawback just mentioned and *does* have PH structure.

Bayesian analysis for model (4) without cure fraction has been considered by Kalbfeisch (1978) using a gamma process prior on the logarithm of the baseline survival function. Clayton (1991) discussed a Monte Carlo method for related frailty models. Designed for use in various survival analysis problems, the beta process was suggested by Hjort (1990). However, Bayesian estimation has not been investigated for the PH model with cure fraction as model (4).

In this paper, we address Bayesian analysis based on model (4) with a beta process to model the accumulative hazard function (CHF), using Monte Carlo methods describes in Smith and Roberts (1993).

Section 2 next describes a nonparametric Bayesian analysis based on discrete survival time with a beta process, and a nonparametric Bayesian analysis for grouped (interval) data is given in Section 3. Section 4 provides a continuous-time nonparametric Bayesian analysis. Finally, Section 5 demonstrates the application of our model by reanalyzing a set of leukemia data.

978-1-4244-2902-8/09/$25.00 ©2009 IEEE

II. Nonparametric Discrete-time Survival Analysis

A. discrete-time model with censoring

A nonparametric Bayesian survival analysis without covariates and long-term survivors was discussed by Hjort (1990). We follow Hjort's notes and let X be a random variable taking values in $(0, b, 2b, \dots)$ with a (proper) distribution function $F(t)$, and let

$$f(jb) = Pr(X = jb), F(jb) = Pr(X \le jb) = \sum_{i=0}^{jb} f(ib),$$

$$\alpha(jb) = Pr(X = jb | X \ge jb) = f(jb)/F[jb, \infty) \quad (5)$$

and $\Lambda(jb) = \sum_{i=0}^{j} \alpha(ib)$ for $j \ge 0$, where α is the hazard rate and Λ is the accumulative hazard rate. Note that $F(t)$ and $f(t)$ can be recovered from knowledge on Λ:

$$F(jb) = 1 - \prod_{i=0}^{j} \{1 - \alpha(ib)\}, \quad (6)$$

$$f(jb) = \left[\prod_{i=0}^{j-1} \{1 - \alpha(ib)\}\right] \alpha(jb), \quad j \ge 0, \quad (7)$$

where (6) follows (5) and

$$1 - \alpha(ib) = \frac{Pr\{X \ge (i+1)b\}}{Pr\{X \ge ib\}}.$$

Let $X_1, \dots, X_n$ be an i.i.d. random sample from the given distribution, but subjected to right censoring. Thus the observations are $(T_1, \delta_1), \dots, (T_n, \delta_n)$, where $T_i = \min(X_i, C_i)$, C_i is the censoring time for individual i, and $\delta_i = I\{X_i \le C_i\}$. Consider the counting process N and the number-at-risk process Y, given by $N_i(jb) = I\{T_i \le jb, \delta_i = 1\}$ and $Y_i(jb) = I\{T_i \ge jb\}$, $j \ge 0$. We sometimes write $dN(jb)$ for the increment $N(jb) - N((j-1)b)$ at time jb and $N(jb) = \sum_{i=1}^{n} N_i(jb)$, $Y(jb) = \sum_{i=1}^{n} Y_i(jb)$. The likelihood of the observed sample from distribution $F(t)$ can be written

$$\begin{aligned} L &= \prod_{i=1}^{n} [f(t_i b)]^{I(\delta_i = 1)} [F(t_i b, +\infty)]^{I(\delta_i = 0)} \\ &= \prod_{j=0}^{\infty} \left[(1 - \alpha(jb))^{Y(jb) - N(jb)} \alpha(jb)^{dN(jb)} \right]. \end{aligned}$$

If Hjort's model is extended to include covariate z (assumed one-dimensional in the sequel), the distribution function for individual i is $F(t) = 1 - (1 - F_0(t))^{\exp(z_i \beta)}$, and we have

$$F(jb) = Pr(X \le jb) = 1 - \prod_{k=0}^{jb} (1 - \alpha(kb))^{\exp(z_i \beta)}.$$

Then we can obtain $f(jb)$ by (5) and (6):

$$f(jb) = \prod_{k=0}^{j-1} \{1 - \alpha(kb)\}^{\exp(z_i \beta)} \left[1 - (1 - \alpha(kb))^{\exp(z_i \beta)} \right].$$

The likelihood function can be written

$$\begin{aligned} L = \prod_{j=0}^{\infty} \{1 - \alpha(jb)\}^{\sum_{i \in R_{jb} - D_{jb}} \exp(z_i \beta)} \\ \times \prod_{l \in D_{jb}} \left(1 - \{1 - \alpha(jb)\}^{\exp(z_l \beta)} \right), \end{aligned} \quad (8)$$

where R_{jb} and D_{jb} denote the risk set at point jb and the number of the total failure data up to time jb, respectively.

If model (8) is allowed to accommodate long-survivors (cf. (4)), we can similarly get its likelihood:

$$\begin{aligned} L = \prod_{j=0}^{\infty} \{1 - h(jb)\}^{\sum_{i \in R_{jb} - D_{jb}} \exp(z_i \beta)} \\ \times \prod_{l \in D_{jb}} \left(1 - \{1 - h(jb)\}^{\exp(z_l \beta)} \right), \end{aligned} \quad (9)$$

where R_{jb} and D_{jb} correspond respectively to the survival function $S(t) = 1 - (1 - pF_0(t))^{\exp(z_i \beta)}$ and the baseline hazard rate

$$h(jb) = \frac{p f_0(jb)}{1 - p F_0(jb)} = \frac{p \prod_{i=0}^{j-1} \{1 - \alpha(ib)\} \alpha(jb)}{1 - p \prod_{i=0}^{j} \{1 - \alpha(ib)\}}. \quad (10)$$

Then we can obtain the likelihood function for parameter p, β and α based on model (4) by incorporating (9) with (10).

B. Prior and posterior distribution

In this subsection, we discuss noninformative priors for p and β, and examine their properties. The noninformative priors of p and β are given respectively by $\pi(p) = 1$, $p \in (0, 1)$; and $\pi(\beta) = 1$, $\beta \in (-a_0, a_0)$, $a_0 > 0$. We first assume that p and β are independent random variables. Let $\alpha_j = \alpha(jb)$, $j = 0, 1, 2, \dots$, and $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_j, \dots)$. Further assume the observed values $t_1, t_2, \dots, t_n$ to be distinct. We specify an independent beta prior for the $\alpha' s$ by taking $\alpha_j \sim Beta(c_{0j}\alpha_{0j}, c_{0j}(1 - \alpha_{0j}))$, independently over $j = 0, 1, 2, \dots$. While it is reasonable to assume that the $\alpha' s$ are independent from each other conditional on the prior, the assumption of a beta distribution for the $\alpha' s$ is only an approximation to the true time-continuous beta process.

Under the discrete beta process defined here, the joint prior density of α is given by

$$\pi(\alpha) \propto \prod_{j=0}^{\infty} {\alpha_j}^{c_{0j}\alpha_{0j} - 1} (1 - \alpha_j)^{c_{0j}(1 - \alpha_{0j}) - 1},$$

and we assume that α is independent of p and β, where α_{0j} is the prior at the hazard and c_{0j} can be interpreted as the number at risk in an imagined prior sample. The c_{0j} and α_{0j} are referred to as *hyperparameters*.

With this specification, the posterior distribution of the parameter vector $\theta = (p, \beta, \alpha)$ based on the observed data $D = (t_i, \delta_i)$ is given by

$$\begin{aligned} L(\theta | D) = \prod_{j=0}^{\infty} \{1 - h(jb)\}^{\sum_{i \in R_{jb} - D_{jb}} \exp(z_i \beta)} \\ \times \prod_{l \in D_{jb}} (1 - \{1 - h(jb)\}^{\exp(z_l \beta)}) \pi(\alpha), \end{aligned} \quad (11)$$

where $h(jb)$ is defined in (10). We use the Gibbs sampling method to draw sample from the joint posterior distribution of (p, β, α). To sample from the full condition of $\pi[\beta|p, \alpha, D]$, we can use the property in following theorem.

Theorem 2.1: The conditional posterior density of β, given p, α, D, is log-concave.

Proof: Write $\Delta_{j1} = R_{jb} - D_{jb}$, $\Delta_{j2} = D_{jb}$ and $q_j = 1 - h(jb)$. We note that the full condition distribution for β is given by

$$\pi[\beta|\cdot] \propto \left[\prod_{j=0}^{\infty} (q_j)^{\sum_{i\in\Delta_{j1}} \exp(z_i\beta)} \prod_{l\in\Delta_{j2}} \left(1 - (q_j)^{\exp(z_l\beta)}\right)\right],$$

hence $\log \pi[\beta|\cdot]$ is proportional to

$$\sum_{j=0}^{\infty}\left[\sum_{i\in\Delta_{j1}} \exp(z_i\beta)\log(q_j) + \sum_{l\in\Delta_{j2}} \log\left(1 - (q_j)^{\exp(z_l\beta)}\right)\right].$$

Its second derivative with respect to β is

$$\sum_{j=0}^{\infty}\left\{\sum_{i\in\Delta_{j1}} z_i^2 \exp(z_i\beta)\log(q_j) - \sum_{l\in\Delta_{j2}} \frac{\Omega_1 + \Omega_2}{\left[1 - q_j{}^{\exp(z_l\beta)}\right]^2}\right\}, \tag{12}$$

where

$$\Omega_1 = q_j{}^{\exp(z_l\beta)}\{\log(q_j) z_l \exp(z_l\beta)\}^2 \left(1 - q_j{}^{\exp(z_l\beta)}\right)$$

and

$$\Omega_2 = \left\{\left(1 - q_j{}^{\exp(z_l\beta)}\right)\log(q_j) z_l \exp(z_l\beta)\right\}^2.$$

It is not difficult to see that (12) is negative and this completes the proof. ∎

III. Nonparametric Grouped Data Survival Analysis

A. grouped data model with censoring

Partition the time axis $(0, \infty)$ into a finite number of disjoint intervals $I_1, I_2, \ldots, I_{g+1}$, where $I_j = (a_{j-1}, a_j]$ for $j = 1, 2, \ldots, g+1$ with $a_0 = 0$ and $a_{g+1} = \infty$. In practice, the grids I_j are often decided either by data collection, patient monitoring and measurement recording; or by the design of the study. For example, patients being monitored on a weekly schedule suggests that the natural choice for the length of each I_j is one week (Sinha and Dey, 1997).

If the survival function is absolutely continuous, we can also carry out a Bayesian analysis just as Chen *et al.* (1999) did. When the survival function is not absolutely continuous, we can still consider the discrete version of Cox proportional hazards model. Let h_i be the discrete baseline hazard rate for the interval without covariate z; that is,

$$h_i = Pr(a_{i-1} < T \le a_i | T > a_{i-1}).$$

Then

$$1 - h_i = \frac{Pr(T > a_i)}{Pr(T > a_{i-1})}.$$

Hence the survival function is given by

$$S(a_j) = Pr(T > a_j) = \prod_{i=1}^{j}\{1 - h_i\}, \quad j = 1, 2, \ldots, g.$$

If we momentarily think of an observed life time t_i as the event $X \in (a_{i-1}, a_i]$, then the contribution of observation i to the likelihood is

$$F(a_{i-1}, a_i] = Pr(a_{i-1} < T \le a_i) = \prod_{j=1}^{i-1}(1 - h_j)h_i$$

if $\delta_i = 1$; or $F(a_i, \infty) = \prod_{j=1}^{i}(1 - h_j)$ if $\delta_i = 0$. Hence the total likelihood is

$$L = \prod_{j=1}^{J}\left\{\left[(1 - h_j)^{\sum_{i\in R_j - D_j}}\right]\prod_{l\in D_j} h_j\right\},$$

where R_j is the risk set and D_j is the failure set of the j^{th} interval I_j.

With covariate z, $S(a_i) = \prod_{j=1}^{i}(1 - h_j)^{\exp(z_k\beta)}$ and

$$F(a_{i-1}, a_i] = \prod_{j=1}^{i-1}(1 - h_j)^{\exp(z_k\beta)}\left[1 - (1 - h_i)^{\exp(z_k\beta)}\right].$$

Then the likelihood function L is equal to

$$\prod_{j=1}^{J}\left[(1 - h_j)^{\sum_{i\in R_j - D_j}\exp(z_i\beta)}\right]\prod_{l\in D_j}\left[1 - (1 - h_j)^{\exp(z_l\beta)}\right].$$

If long-term survivors are included, the likelihood L can be written as

$$\prod_{j=1}^{J}\left[(1 - \bar{h}_j)^{\sum_{i\in R_j - D_j}\exp(z_i\beta)}\right]\prod_{l\in D_j}\left[1 - (1 - \bar{h}_j)^{\exp(Z_l\beta)}\right], \tag{13}$$

where

$$\bar{h}_j = \frac{pF_0(a_{j-1}, a_j]}{1 - pF_0(a_{j-1})} = \frac{p\prod_{i=1}^{j-1}\{1 - h_i\}h_j}{1 - p + p\prod_{i=1}^{j-1}\{1 - h_i\}}. \tag{14}$$

Then we can obtain the likelihood function for parameter p, β and h based on a proportional hazard model with long-term survivors by incorporating (13) with (14).

B. Prior and posterior distribution

Now we consider a noninformative prior for p: $\pi(p) = 1$, $p \in (0, 1)$. A typical prior for β is $N(\mu_0, \sigma_0^2)$ prior (Chen, *et al.* 2002), or we can take a noninformative prior for β: $\pi(\beta) \propto 1$, $\beta \in [-a_0, a_0]$, where $a_0 > 0$. p and β are assumed to be independent. The prior of h is given by

$$\pi(h) \propto \prod_{j=1}^{J} h_j{}^{c_{0j}h_{0j}-1}(1 - h_j)^{c_{0j}(1-h_{0j})-1}.$$

That is, $h_j \sim Beta(c_{0j}h_{0j}, c_{0j}(1 - h_{0j}))$ and $h_1, \ldots, h_J$ are independent. Moreover, we assume that h is independent of p

and β. Then the joint posterior of $\theta = (p, \beta, h)$ can be written as

$$\pi(\theta|D) = \prod_{j=1}^{J} \left[(1-\bar{h}_j)^{\sum_{i \in R_j - D_j} \exp(z_i\beta)} \right] \times \prod_{l \in D_j} \left[1 - (1-\bar{h}_j)^{\exp(z_l\beta)} \right] \pi(h),$$

where $\bar{h}_j$ is define in (14). Similar to Theorem 2.1, we can prove that the full condition distribution $\pi[\beta|p, h, D]$ is log-concave.

IV. Nonparametric Continuous-time Survival Analysis

A. Continuous-time model with censoring

Let T be a random variable with cdf $F(t) = Pr(T \le t)$ with $F(0) = 0$. The cumulative hazard rate for T or F is a nonnegative, nondecreasing, right continuous function Λ on $[0, \infty)$ which satisfies

$$d\Lambda(s) = Pr(T \in [s, s+ds) | T \ge s) = dF(s)/F[s, \infty).$$

We define

$$\Lambda[a, b) = \int_{[a,b)} \frac{dF(s)}{F[s, \infty)}. \tag{15}$$

Then

$$F[a, b) = \int_{[a,b)} F[s, \infty) d\Lambda(s), 0 \le a \le b < \infty. \tag{16}$$

If F is absolutely continuous, then $\Lambda(t) = -\log\{1 - F(t)\}$ or $F(t) = 1 - \exp\{-\Lambda(t)\}$. This relation, however, no longer holds if the cdf has jumps, and so Hjort (1993) prefers (15) as the starting point for interpretational reasons.

It is not difficult to see that $F(t)$ is uniquely determined by $\Lambda(t)$. F can be restored from equation (16), which is given by the product integral (see Gill and Johansen (1990))

$$F(t) = 1 - \prod_{[0,t]} \{1 - d\Lambda(s)\}, t \ge 0. \tag{17}$$

It can be shown that

$$\prod_{[a,b]} \{1 - d\Lambda(s)\} = \exp\{-\Lambda[a, b]\}$$

if and only if Λ is continuous. Hence $\Lambda = -\log(1 - F)$ holds under this condition. In general, as $1 - x \le \exp(-x)$, we have $F(t) \ge 1 - \exp\{-\Lambda(t)\}$ or $-\log\{1 - F(t)\} \ge \Lambda(t)$. We also mention that, the most important property of product-integration is the multiplicative relation over disjoint intervals, which follows easily from the definition:

$$\prod_{(s,u]} (1 + dX) = \prod_{(s,t]} (1 + dX) \prod_{(t,u]} (1 + dX) \tag{18}$$

for $0 \le s \le t \le u$ (see Andersen *et al.* (1993)).

Let t_i denote the observed life time of individual i and the event $X \in [t_i, t_i + \varepsilon]$. Then the contribution of i to the likelihood is $F[t_i, t_i + \varepsilon] = F(t_i + \varepsilon) - F(t_i)$ if $\delta_i = 1$. Using (18), we have

$$F[t_i, t_i + \varepsilon] = \prod_{[0,t_i)} \{1 - d\Lambda(s)\} \left[1 - \prod_{[t_i, t_i+\varepsilon]} \{1 - d\Lambda(s)\} \right],$$

where $\prod_{[t_i, t_i+\varepsilon)} \{1 - d\Lambda(s)\} = d\Lambda(t_i)$, which follows from (17) and the definition of $d\Lambda(s)$. If $\delta_i = 0$, the contribution of i is

$$F(t_i, \infty) = \prod_{[0,t_i]} \{1 - d\Lambda(s)\}. \tag{19}$$

A discussion in Section 6 of Hjort (1990) or Sinha and Dey (1997) leads to the total likelihood for Λ and β:

$$L = \prod_{[0,\infty)} \{1 - d\Lambda(s)\}^{Y(s)} \prod_{i=1}^{n} \left[\{1 - d\Lambda(t_i)\}^{-1} - 1 \right]^{\delta_i}, \tag{20}$$

where $Y(s) = \sum_{i=1}^{n} I(t_i \ge s)$, and we replace $d\Lambda(t_i)$ with $\Lambda\{t_i\}$, where $\Lambda\{t_i\} = \Lambda(t_i) - \Lambda(t_i-)$ if $\Lambda(t)$ is discontinuous at point t_i for each i when $\delta_i = 1$ (see Hjort (1990)).

If model (20) is extended to include covariate z, then the survival function is $S(t) = [S_0(t)]^{\exp(z\beta)}$. Let Λ denote the cumulative hazard rate of S. By the definition of $d\Lambda(t)$,

$$1 - d\Lambda(t) = \frac{Pr(T \ge t + dt)}{Pr(T \ge t)} = \frac{S(t + dt)}{S(t)}, \tag{21}$$

$S(t) = [S_0(t)]^{\exp(z\beta)}$ and $S(t + dt) = [S_0(t + dt)]^{\exp(z\beta)}$. Hence

$$1 - d\Lambda(t) = [1 - d\Lambda_0(t)]^{\exp(z\beta)}. \tag{22}$$

If $R(t, \beta) = \sum_{j=1}^{n} \exp(z_j\beta) I(t_j \ge t)$, by (20)-(22), the total likelihood function L is given by

$$\prod_{[0,\infty)} \{1 - d\Lambda_0(s)\}^{R(t,\beta)} \prod_{i=1}^{n} \left[\{1 - d\Lambda_0(t_i)\}^{-\exp(z_i\beta)} - 1 \right]^{\delta_i}.$$

Similarly, if model (20) allows long-term survivors and covariate z, that is, $S(t) = (1 - pF_0(t))^{\exp(z\beta)}$, then the likelihood function L can be written as

$$\prod_{[0,\infty)} \{1 - d\bar{\Lambda}(s)\}^{R(t,\beta)} \prod_{i=1}^{n} \left[\{1 - d\bar{\Lambda}(t_i)\}^{-\exp(z_i\beta)} - 1 \right]^{\delta_i}, \tag{23}$$

where

$$\begin{aligned} d\bar{\Lambda}(s) &= \frac{p(1 - F_0(s))}{1 - pF_0(s)} d\Lambda_0(s) \\ &= \frac{p \prod_{[0,s]} \{1 - d\Lambda_0(t)\}}{1 - p + p \prod_{[0,s]} \{1 - d\Lambda_0(t)\}} d\Lambda_0(s). \end{aligned} \tag{24}$$

Then we can obtain the likelihood function for parameter p, β and $\Lambda_0(t)$ based on a proportional hazard model with "long-term" survivors by incorporating (23) with (24).

B. Prior and posterior distribution

Assume that p and β are mutually independent, with priors $\pi(p) = 1$, $p \in (0,1)$ and $\pi(\beta) = 1$, $\beta \in [-a_0, a_0], a_0 > 0$. Hjort (1990) and Sinha and Dey (1997) discussed model (23) with a Beta process prior. Let Λ_0 be a cumulative hazard function with a finite number of jumps at $t_1, t_2, \ldots, t_m$ and $c(\cdot)$ a piecewise continuous, nonnegative function on $[0, \infty)$. A prior for Λ_0 is given by

$$\Lambda_0 \sim Beta(c(\cdot), \tilde{\Lambda}(\cdot)), \tag{25}$$

$$\Lambda_0(t_j) \sim Beta((c(t_j)\tilde{\Lambda}(t_j), c(t_j)(1 - \tilde{\Lambda}(t_j))), \tag{26}$$

where Λ_0 is a prior at the cumulative hazard, and $c(t)$ can be interpreted as the number at risk at t in a margined prior sample. The above definition implies that the Beta process has independent increments that have Beta distributions at fixed points of discontinuity.

We follow Hjort (1990) and assume that the prior for Λ_0 is given by (25) and (26) in model (23). Then the joint posterior $L(p, \beta, \Lambda_0|\cdot)$ is given by the following multiplier of $\pi(\Lambda_0)$:

$$\prod_{[0,\infty)} \{1 - d\bar{\Lambda}(s)\}^{R(s,\beta)} \prod_{i=1}^{n} \left[\{1 - d\bar{\Lambda}(t_i)\}^{-\exp(z_i\beta)} - 1\right]^{\delta_i},$$

where $\bar{\Lambda}$ is given by (24). Similar to Theorem 2.1, the full condition distribution $\pi[\beta|p, h, D]$ is log-concave.

V. Re-analysis of the Leukemia Data

It is possible to simulate the beta process for the baseline cumulative hazard defined under a cox model with continuous-time data. In practice, however, survival data are commonly grouped within some grid intervals, where the grid size is determined by the data and trial design. So for practice purpose, it is more often and sufficient to use a discretized version of beta process along with grouped survival data. In this section, we consider the leukemia data in Maller and Zhou (1996) to demonstrate the application of the techniques proposed in this article.

The leukemia data are on the relapse times of leukemia and have been analyzed extensively in Maller and Zhou (1996) from the classic point of view. It has also been analyzed by Cancho and Bolfaring (2001) using Bayesian inference. The data consist of the relapse times of patients who had been treated by two types of transplants called *allogeneic* (Group 1 with covariate $z = 0$) and *autologous* (Group 2 with covariate $z = 1$). Maller and Zhou (1996) reported successful application of an exponential cure model to the data set in Group 1. For Group 2, however, it appears that even Weibull cure model does not present a reasonable fit. So Cancho and Bolfaring (2001) model the presence of immunes by using an exponentiated-Weibull model to fit the data in Group 2, and conduct classic and Bayesian methods to draw inference on the parameters. The inference and the revision for the "cure rate" in Cancho and Bolfaring (2001) are found to be reasonable. More recently, Shao and Zhou (2004) reanalyzed this data by using the Burr XII distribution to improve data fitting. To be more flexible and robust for the inference with less reliance on parametric assumptions, we consider a nonparametric Bayesian inference without specification of the survival function in this section. Bayesian inference for the parameters produces similar results to those of Cancho and Bolfaring (2001) for Group 2 data.

In the analysis below, the prior distribution for β is taken to be normal with mean 0.2 and standard deviation 0.1. Similar results are obtained with a noninformative prior, but are not reported here to limit the length of the paper. The prior for p is taken to be a uniform distribution between 0 and 1. The prior for the baseline cumulative hazard function is a beta process. We chose $J = 5$ and $J = 10$ to carry out the Bayesian inference. It is well known that the selection of prior hyperparameters is crucial for this nonparametric Bayesian inference, which is determined as follows. First, the baseline distribution function of the failure time is chosen to be $F_0(t) = 1 - \exp(-0.4t)$. Then the beta process prior for the baseline cumulative hazard function is taken as $\Lambda(t) = 0.4t$ with $c(t) = k\exp(-0.4t)$ (cf. Section 3.2), where k varies from 0.01 to 10. The prior hyperparameters h_{0j} and c_{0j} are then determined by this assumption for $j = 1, 2, \ldots, 5$. or $j = 1, 2, \ldots, 10$.

It is should be mentioned that, to sample from the full condition distributions $\pi[p|\beta, h, D]$, $\pi[\beta|p, h, D]$ and $\pi[h_j|p, \beta, D]$, the Metropolis-Hastings algorithm (Gilks and Wild, 1992) can be utilized, which is similar to that described in Chen *et al.* (2001). However, a simpler method to generate a sample from the posterior is given in Smith and Gelfand (1992): for each $\theta = (p, \beta, h_1, \ldots, h_{10})$ in the prior sample, accept θ into the posterior sample with probability $L(\theta, \cdot)/L(\hat{\theta}, \cdot)$, where the likelihood function is given by the equations such as (3.2), and $\hat{\theta}$ maximizes $L(\theta, \cdot)$.

With these priors and the data given by the leukemia dataset, the Gibbs sample was implemented and run with each setting of k. Saving 8000 iterations after discarding the first 2000, the results for parameter p and coefficients of regression β are summarized in Table 1 below (only the case of $J = 10$ is reported for saving space), where p_1 denotes the proportion of susceptibles in Group 1 and p_2 in Group 2. It is interesting to note that the substantial reduction on the posterior mean of p for comparison with the classic estimation in Maller and Zhou (1996). The posterior means of p are also remarkably close to the results based on the parametric Bayesian methodology proposed by Cancho and Bolfaring (2001). Small values of k represent little prior input whereas large values correspond to a strong prior belief that the baseline hazard rate is constant at 0.4 (i.e., an exponential distribution with mean 2.5). These posterior means for p are stable. These posterior means for β are also remarkably stable over a large range of k. In Figures below (Fig. 1 and Fig.2), we show the posterior distribution of p, β. As k increases, the shape approaches the exponential distribution with the chosen prior.

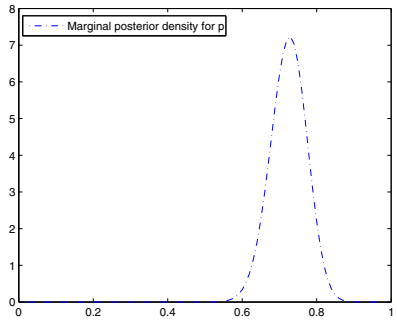

Fig. 1. Marginal Posterior densities of p with k=3 and J=10

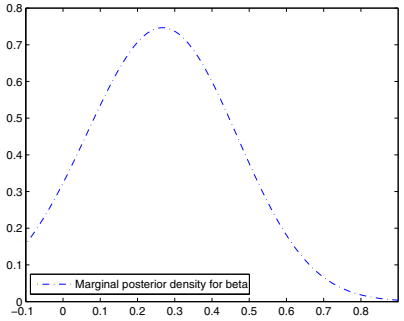

Fig. 2. Marginal Posterior densities of β with k=3 and J=5

TABLE I
POSTERIOR SUMMARIES WITH J=10

k	0.01			0.1		
Variable	p_1	β	p_2	p_1	β	p_2
Mean	0.7193	0.1921	0.7855	0.7197	0.1917	0.7858
Std	0.0023	0.0490	0.0045	0.0045	0.0480	0.0031
k	1			3		
Variable	p_1	β	p_2	p_1	β	p_2
Mean	0.7201	0.1919	0.7862	0.7221	0.1915	0.7879
Std	0.0023	0.0510	0.0032	0.0023	0.0487	0.0026
k	5			7		
Variable	p_1	β	p_2	p_1	β	p_2
Mean	0.7370	0.1917	0.8017	0.7474	0.1918	0.8084
Std	0.0033	0.0509	0.0032	0.0033	0.0521	0.0042

VI. CONCLUSION

In this paper, nonparametric Bayesian analysis for censored data in Hjort (1990) has been extended to model survival data with long-term survivors. A set of leukemia data are reanalyzed by using the proposed nonparametric Bayesian models, in which a simple method to generate a sample from the posterior given in Smith and Gelfand (1992) is employed to obtain the Gibbs sample.

In Section 5, the reanalysis of the leukemia data is based on the nonparametric grouped data model proposed in Section 2. However, the nonparametric continuous-time survival analysis in Section 4 is an attractive alternative of interest. Another important issue with the cure model is the model comparison. It is of considerable interest to compare the proposed models in Sections 3 and 4 with the model in Section 2 to analyze the leukemia data. As mentioned in Chen *et al.* (2002), Bayesian factors are not suitable for cure models since they do not allow a noninformative improper prior distribution. As a result, the conditional predictive ordinate (CPO) cannot be used to do formal model comparisons (cf. Chen *et al.*, 2002). These will be interesting issues for further studies.

ACKNOWLEDGMENT

This work was partially supported by the NSFC under grant 10871084, and by PIRTJiangnan.

REFERENCES

[1] P. K. Andersen, Φ. Borgan and R. D. Gill, *Statistical models based on counting processes.* New York: Springer Verlag, 1993
[2] V. G. Cancho and H. Bolfaring, Modeling the presence of immunes by using the exponentiated Weibull model, *Journal of Applied Statistics*, 2001, 28, 659-671.
[3] M. H. Chen, J.G. Ibrahim and D. Sinha, A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 1999,94, 909-919.
[4] M. H. Chen and G. J. Ibrahim, Bayesian cure rate models for malignant melano- ma, a case-study of cooperative oncology group trial E1690. *Applied Statist,* 2002, 51, 135-150.
[5] D. G. Clayton, A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 1991, 47, 467-485.
[6] V. T. Farewell, The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 1982, 38, 1041-1046.
[7] R. D. Gill and S. Johansen, A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, 1990, 18, 1501-1555.
[8] W. R. Gilks and D. Wild, Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 1992, 2, 337-348.
[9] N. L. Hjort, Nonparametric Bayesian estimators based on beta Processes in models for life history data. *The Annals of Statistics*, 1990, 18, 1259-1294.
[10] N. L. Hjort, *Dynamic likelihood hazard rate estimation*. Technical Report 4, Institute of Mathematics, University of Oslo, 1993
[11] J.G. Ibrahim, M.H. Chen and D. Sinha, Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 1999, 57, 383-388.
[12] J. D. Kalbfleisch, Non-parameteric Bayesian analysis of survival time data. *Journal of the Royal statistical Society B*, 1978, 40, 214-221.
[13] A. Y. C. Kuk and C. H. Chen, A mixture models combining with logistic regression with proportional hazards regression. *Biometrika*, 1992, 79, 531-541.
[14] R. A. Maller and X. Zhou, *Survival analysis with long-term survivors.* John Wiley and Sons, 1996
[15] A. F. M. Smith and G.O. Roberts, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, 1993, 55, 3-23.
[16] A. F. M. Smith and A.E. Gelfand, Bayesian statistics without tears: a sampling resampling perspective. *The American Statistician*, 1992, 46, 84-88.
[17] Q. X. Shao and X. Zhou, A new parametric model for survival data with long-term survivors. *Statistics in Medicine*, 2006, 23,3525-3543
[18] D. Sinha and D. K. Dey, Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 1997, 92, 1195-1212.
[19] X. B. Zhao and X. Zhou, Proportional hazards models for survival data with long-term survivors. *Statistics and Probability Letters*, 2006, 76, 1685-1893.
[20] X. B. Zhao and X. Zhou, Discrete-time survival analysis for survival data with long-term survivors. *Statistics in Medicine*, 2008, 27,1261-1281.