Journal of the Korean Statistical Society (2001), **30**: *1, pp 127–138*

Mixtures of Beta Processes Priors for Right Censored Survival Data[†]

Yongdai Kim[1]

ABSTRACT

In order to combine parametric and nonparametric approaches together for survival analysis with censored observations, a new class of priors called mixtures of beta processes is introduced. It is shown that mixtures of beta processes priors generalize the well known priors - mixtures of Dirichlet processes, and they are conjugate with right censored observations. Formulas for computing the posterior distribution are derived. Finally, a real data set is analyzed for illustrational purpose.

Keywords: Beta processes; Dirichlet processes; Censored observations.

1. INTRODUCTION

When one estimates the survival function with right censored data, Kaplan-Meier estimator (Kapaln-Meier, 1958) is used widely since it is the nonparametric maximum likelihood estimator. However, there are many cases where there are physical reasons that indicate specific parametric families. Exponential distributions arise in a very large number of contexts; extreme value distributions arise frequently in reliability theory because they are the limiting distributions of the lifelengths of series or parallel systems with a large number of identically distributed components. In these situations there would be a loss of efficiency if one were to use a nonparametric estimator instead of using a parametric estimator. In fact, Miller (1983) showed that Kaplan-Meier estimator would be inefficient if one knows the parametric family. Another drawback of the nonparametric estimator is that it fails to estimate tail probabilities when large observations are heavily censored. In particular, if the largest observation is censored, then the Kaplan-Meier estimator does not yield a proper distribution function. On the

[†]This research was supported (in part) by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

[1]Department of Statistics, Hankuk University of Foreign Studies, 449-791, Kyunggi, Korea

other hand, a parametric model provides a description of data approximately, and if this approximation is crude, the estimator would perform poorly and may lead us to a wrong conclusion.

To overcome such deficiencies, Doss (1994) proposed a Bayesian approach based on mixtures of Dirichlet processes priors (Antoniak, 1974) as follows. Let F be an unknown distribution function. For a prior distribution of F, choose a parametric family H_θ. Then for a given parametric family H_θ and a given family of positive constants α_θ, $\theta \in \Theta \subset R^p$, first θ is chosen according to some prior distribution ν on Θ and then F is chosen from $\mathcal{D}_{\alpha_\theta H_\theta}$ where $\mathcal{D}_{\alpha_\theta H_\theta}$ stands for a Dirichlet process with parameter $\alpha_\theta H_\theta$. The model with a mixture of Dirichlet processes prior can avoid the loss of efficiency due to ignoring partial information about parametric family by choosing an appropriate parametric family H_θ but that at the same time can avoid the pitfalls connected with an incorrectly specified parametric model by putting prior mass nonparametrically to small neighborhood of an entire parametric family.

However, it is well known that Dirichlet processes priors are not conjugate with right censored observations and so are not mixtures of Dirichlet processes. To overcome such deficiency, we propose a new class of priors - mixtures of beta processes. The proposed priors put beta processes, instead of Dirichlet processes, around a given parametric family of cumulative hazard functions. Beta processes are suggested by Hjort (1990) as a conjugate family of nonparametric priors for right censored data. It is proved in this paper that the proposed priors - mixutres of beta processes, are also conjugate with right censored data by deriving the closed form of the posterior distribution. It is worth noting that our proposed priors generalize mixtures of Dirichlet processes priors since Dirichlet processes are special forms of beta processes (Hjort, 1990).

Having a conjugate family of priors and obtaining the closed form of the posterior distribution are important in Bayesian analysis. The closed form solution of the posterior distribution helps not only to lessen computational burden but also to study theoretical properties of the posterior distribution. First, we can avoid the complicated MCMC (Markov Chain Monte Carlo) algorithm suggested by Doss (1994). Second, large sample properties of the posterior distribution can be studied by exploring the closed form of the posterior distribution, which cannot be done in the framework of MCMC. In Section 4 of this paper, by investigating the formula of the posterior distribution θ, we find that the choice of the precision parameter in the mixture of beta processes priors is crucial for the posterior distribution to have desirable large sample properties such as posterior consistency.

The paper is organized as follows. In Section 2, mixtures of beta processes priors are introduced. In Section 3, formulas of the posterior distribution with right censored data are derived. In Section 4, we illustrate our model on a real data set of survival times of prostate cancer patients, and discussions follow in Section 5.

2. MIXTURE OF BETA PROCESSES PRIORS

First, we give the definition of beta processes (Hjort, 1990). Let $\Lambda(\cdot)$ be a cumulative hazard function (chf for abbreviation) with a finite number of jumps taking place at $t_1, \cdots, t_m$ and let $c(\cdot)$ be a piecewise continuous nonnegative function on $[0, \infty)$. A process A is called a beta process with mean $\Lambda(\cdot)$ and precision parameter $c(\cdot)$, denoted

$$A \sim \mathcal{B}(\Lambda(\cdot), c(\cdot)), \tag{2.1}$$

if for any $\theta > 0$ it has Lévy representation

$$E(e^{-\theta A(t)}) = \prod_{j:t_j \leq t} E(e^{-\theta S_j}) \exp[-\int_0^1 (1 - e^{-\theta s}) dL_t(s)],$$

with

$$S_j \sim Beta(c(t_j)\Delta\Lambda(t_j), c(t_j)(1 - \Delta\Lambda(t_j)))$$

and

$$dL_t(s) = \int_0^t c(z) s^{-1} (1 - s)^{c(z)-1} d\Lambda_c(z) ds$$

for $t \geq 0$ and $0 < s < 1$, where Λ_c is the continuous part of Λ, $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$, and $Beta(a, b)$ denotes the beta distribution with parameters a and b .

Mixtures of beta processes are defined similarly to mixtures of Dirichlet processes as follows. First, choose a parametric family Λ_θ of chf. Then for a given parametric family of chf Λ_θ and a given family of piecewise continuous nonnegative functions $c_\theta(t)$, $\theta \in \Theta \subset R^p$, first θ is chosen according to a certain prior distribution ν on Θ and then A is chosen from $\mathcal{B}(\Lambda_\theta(\cdot), c_\theta(\cdot))$. We denote a mixture of beta processes A by

$$A \sim \int \mathcal{B}(\Lambda_\theta, c_\theta) \nu(d\theta). \tag{2.2}$$

Remark 2.1. A mixture of beta processes puts its mass on the space of cumulative hazard functions, denoted by $\mathcal{A}$, instead of the space of distribution functions, denoted by $\mathcal{F}$. However, we can recover the distribution function from a given cumulative hazard function by use of the product integration and so we can translate a prior given on $\mathcal{A}$ to one on $\mathcal{F}$. Refer to Walker and Muliere (1997) to see how to define a prior on $\mathcal{F}$ with beta processes priors. We keep, however, using $\mathcal{A}$ instead of $\mathcal{F}$ since firstly the posterior distribution is more nicely presented on $\mathcal{A}$ rather than $\mathcal{F}$, and secondly the distribution of survival data is in many cases modeled through the hazard function. For example, an exponential distribution can be characterized by a distribution with a constant hazard function.

3. POSTERIOR DISTRIBUTION WITH RIGHT CENSORED DATA

Suppose that a priori A is a mixture of beta processes given in (2.2). For a given A, we have $X_1, \ldots, X_n$, i.i.d. survival times with the common chf A. The Xs are not directly available. Instead, $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ are observed where $T_i = \min(C_i, X_i)$, $\delta_i = I(X_i \leq C_i)$, and $C_1, \ldots, C_n$ are i.i.d. random variables with a common distribution G and are independent with $X_1, \ldots, X_n$.

Let

$$N_n(t) = \sum_{i=1}^{n} I(T_i \leq t, \delta_i = 1),$$

$$Y_n(t) = \sum_{i=1}^{n} I(T_i \geq t)$$

and $d_n(t) = \sum_{i=1}^{n} I(T_i = t, \delta_i = 1)$. Let $T_{(1)} \leq \cdots \leq T_{(n)}$ be ordered values of $T_1, \ldots, T_n$. Let q_n be the number of distinct uncensored observations among $T_1, \ldots, T_n$ and let $v_1, \ldots, v_{q_n}$ be the distinct uncensored observations. Let $T^n = (T_1, \ldots, T_n)$ and $\delta^n = (\delta_1, \ldots, \delta_n)$.

Theorem 3.1. *Suppose that a priori A is a mixture of beta processes given in (2.2). Suppose that $\Lambda_\theta(t)$ is absolutely continuous with $\lambda_\theta(t)$ as its first derivative in t. Then, the posterior distribution of A given (T^n, δ^n) is also a mixture of beta processes given by*

$$\int \mathcal{B}(\Lambda_\theta^p(\cdot), c_\theta^p(\cdot))\nu^p(d\theta | T^n, \delta^n)$$

where

$$\begin{aligned}\nu^p(d\theta|T^n,\delta^n) &\propto \exp\left\{-\sum_{i=1}^{n}\int_0^{T_{(i)}}\frac{c_\theta(s)}{c_\theta(s)+(n-i)}d\Lambda_\theta(s)\right\}\\ &\times \prod_{k=1}^{q_n} c_\theta(v_k)\lambda_\theta(v_k)\left[\prod_{j=1}^{d_n(v_k)}(c_\theta(v_k)+Y_n(v_k)-j)\right]^{-1}\\ &\times \nu(d\theta), \end{aligned} \tag{3.1}$$

and

$$d\Lambda_\theta^p(s) = \frac{c_\theta(s)d\Lambda_\theta(s)+dN_n(s)}{c_\theta(s)+Y_n(s)}, \tag{3.2}$$

$$c_\theta^p(s) = c_\theta(s)+Y_n(s). \tag{3.3}$$

Proof: Since the posterior distribution of A given (T^n,δ^n) and θ is a beta process with parameters (3.2) and (3.3) (Hjort, 1990), it suffices to show that the posterior distribution of θ is given by (3.1).

Let $P_n(\cdot|\theta,A)$ be the probability measure of (T^n,δ^n) given θ and A and let $P_n(\cdot|\theta)$ be the probability measure of (T^n,δ^n) given θ only. Then we have

$$P_n(\cdot|\theta) = \int_{\mathcal{A}} P_n(\cdot|\theta,A)P_{\mathcal{A}}(dA)$$

where $\mathcal{A}$ is the space of the cumulative hazard functions equipped with the Borel σ-field generated by cylinder sets and $P_{\mathcal{A}}$ is the probability measure of A (i.e. the prior measure of A).

The main point of obtaining the posterior distribution of θ given (T^n,δ^n) is to find out a σ-finite measure μ such that $P_n(\cdot|\theta)$ is absolutely continuous with respect to μ. Then Bayes theorem yields that

$$\nu(d\theta|T^n,\delta^n) \propto f(T^n,\delta^n|\theta)\nu(d\theta)$$

where

$$f(T^n,\delta^n|\theta) = \frac{dP_n(\cdot|\theta)}{d\mu}(T^n,\delta^n).$$

However, finding μ is not straightforward since first $T_1,\ldots,T_n$ can have ties and second T_is are not independent. The idea of finding μ is as follows. Let

$T^i = (T_1, \ldots, T_i)$ and $\delta^i = (\delta_1, \ldots, \delta_i)$ with $T^0 = 0$ and $\delta^0 = 0$. Write

$$P_n(B_1 \times \cdots \times B_n, U_1 \times \cdots \times U_n|\theta) = \int_{B_n \times U_n} \cdots \int_{B_1 \times U_1} \prod_{i=1}^{n} P(dt_i, du_i | t^{i-1}, u^{i-1}, \theta) \tag{3.4}$$

where B_i are Borel subesets of $[0, \infty)$ and U_i are subsets of $\{0, 1\}$ and $P(\cdot|t^{i-1}, u^{i-1}, \theta)$ are probability measures of (T_i, δ_i) given $(T^{i-1}, \delta^{i-1}) = (t^{i-1}, u^{i-1})$ and θ. Then it can be shown by use of Hjort's result (1990) that

$$\begin{aligned} P(B_i, U_i|t^{i-1}, u^{i-1}, \theta) &= \sum_{u \in U_i} \int_{B_i} (1 - G(t)) dF_i(t|\theta) I(u = 1) \\ &+ \sum_{u \in U_i} \int_{B} (1 - F_i(t|\theta)) dG(t) I(u = 0) \end{aligned}$$

where

$$F_i(t|\theta) = 1 - \prod_{s \in [0,t]} \left[1 - \frac{c_\theta(s) d\Lambda_\theta(s) + dN_{i-1}(s)}{c_\theta(s) + Y_{i-1}(s)} \right].$$

Here

$$N_i(t) = \sum_{j=1}^{i} I(T_j \leq t, \delta_j = 1)$$

and

$$Y_i(t) = \sum_{j=1}^{i} I(T_j \geq t)$$

with $N_0(t) = 0$ and $Y_0(t) = 0$. Now, $P(B_i, U_i|t^{i-1}, u^{i-1}, \theta)$ is dominated by μ_n given by

$$\mu_n(B, U) = \sum_{u \in U} \left\{ \left[\sum_{j=1}^{q_n} I(v_j \in B) + \pi(B) \right] I(u = 1) + \int_B dG(t) I(u = 0) \right\}$$

where π is the Lebesgue measure on $[0, \infty)$. Furthermore, direct calculation yields that the density $f_i(t, u|t^{i-1}, u^{i-1}, \theta)$ of $P(\cdot|t^{i-1}, u^{i-1}, \theta)$ with respect to μ_n is given by

$$\begin{aligned} f_i(t, u|T^{i-1}, \delta^{i-1}, \theta) &= [\Delta F_i(t|\theta) I(t \in B_n) + f_i(t|\theta) I(t \notin B_n)] (1 - G(t)) I(u \quad (3.5) \\ &+ (1 - F_i(t|\theta)) I(u = 0), \end{aligned}$$

where

$$\begin{aligned}\Delta F_i(t|\theta) &= \exp\left(-\int_0^t \frac{c_\theta(s)d\Lambda_\theta(s)}{c_\theta(s)+Y_{i-1}(s)}\right)\frac{\Delta N_{i-1}(t)}{c_\theta(s)+Y_{i-1}(t)}\\ &\quad\times\prod_{s<t}\left[1-\frac{dN_{i-1}(s)}{c_\theta(s)+Y_{i-1}(s)}\right],\end{aligned}$$

$$\begin{aligned}f_i(t|\theta) &= \exp\left(-\int_0^t \frac{c_\theta(s)d\Lambda_\theta(s)}{c_\theta(s)+Y_{i-1}(s)}\right)\frac{c_\theta(t)\lambda_\theta(t)}{c_\theta(t)+Y_{i-1}(t)}\\ &\quad\times\prod_{s<t}\left[1-\frac{dN_{i-1}(s)}{c_\theta(s)+Y_{i-1}(s)}\right]\end{aligned}$$

and $B_n = \{v_1, \ldots, v_{q_n}\}$. Then (3.4) can be rewritten as

$$P(B_1\times\cdots\times B_n, U_1\times\cdots\times U_n|\theta) = \int_{B_n\times U_n}\cdots\int_{B_1\times U_1}\prod_{i=1}^n f_i(t_i, u_i|t^{i-1}, u^{i-1}, \theta)\mu_n(dt_i, du_i).$$

Hence if we let $\mu = \mu_n \times \cdots \times \mu_n$, the probability density function $f(t^n, u^n|\theta)$ of (T^n, δ^n) given θ with respect to μ can be written as

$$f(t^n, u^n|\theta) = \prod_{i=1}^n f_i(t_i, u_i|t^{i-1}, u^{i-1}, \theta),$$

and so we have

$$\begin{aligned}f(T^n, \delta^n|\theta) &\propto \exp\left\{-\sum_{i=1}^n\int_0^{T_{(i)}}\frac{c_\theta(s)d\Lambda_\theta(s)}{c_\theta(s)+(n-i)}\right\}\\ &\quad\times\prod_{k=1}^{q_n} c_\theta(v_k)\lambda_\theta(v_k)\frac{\Gamma(c_\theta(v_k)+Y_n(v_k)-d_n(v_k))}{\Gamma(c_\theta(v_k)+Y_n(v_k))}.\end{aligned}$$

Since the posterior distribution ν^p of θ given (T^n, δ^n) is given by

$$\nu^p(d\theta|T^n, \delta^n) \propto f(T^n, \delta^n|\theta)\nu(d\theta),$$

the proof is done. □

Before going further, we explain that how the posterior distribution of θ affects the posterior distribution of tail probabilities of the survival function. Let $S(t|u) = (F(t)-F(u))/(1-F(u))$ for $t \geq u$, which is the conditional survival function given that the survival time is greater than u. Then, we can write $S(t|u) = \prod_{s\in(u,t]}(1 - dA(s))$ and so the Bayes estimator of $S(t|u)$ becomes

$$\mathrm{E}(S(t|u)|T^n, \delta^n) = \int_\Theta \mathrm{E}(S(t|u)|T^n, \delta^n, \theta)\nu^p(d\theta|T^n, \delta^n)$$

where

$$\begin{aligned}\mathrm{E}(S(t|u)|T^n, \delta^n, \theta) &= \prod_{s\in(u,t]} (1 - d\mathrm{E}(A(s)|T^n, \delta_n, \theta)) \\ &= \prod_{s\in(u,t]} \left(1 - \frac{dN_n(s)}{c_\theta(s) + Y_n(s)}\right) \exp\left(-\int_u^t \frac{c_\theta(s)d\Lambda_\theta(s)}{c_\theta(s) + Y_n(s)}\right).\end{aligned}$$

The first equality is due to the fact that conditional on T^n, δ^n and θ, A is a Lévy process. Since $N_n(t) = 0$ and $Y_n(t) = 0$ for $t > T_{(n)}$, we have

$$\mathrm{E}(S(t|T_{(n)})|T^n, \delta^n) = \int_{\theta\in\Theta} \exp\left(-\int_u^t d\Lambda_\theta(s)\right) d\nu^p(\theta|T^n, \delta^n)$$

for $t > T_{(n)}$ which depends only on the posterior distribution of θ.

4. APPLICATION TO RIGHT CENSORED SURVIVAL DATA

In this section, we apply our method to a data set involving a clinical trial of 211 individuals who had stage IV prostate cancer. The data set is studied by Koziol and Green (1976), Hollander and Proschan (1979), and Doss (1994). One feature of this data set is that the large observations are heavily censored (in particular, there are only two unsensored deaths beyond 120 months and only one beyond 150 months, 158 being the last uncensored observation). Doss (1994) mentioned that nonparametric Bayes methods are desirable for this data set since frequentist's methods failed to estimate a large portion of the right tail due to the heavy censoring of the large observations.

To analyze the prostate cancer data set with mixtures of beta processes priors, we choose the exponential distribution with parameter θ for a parametric family Λ_θ, that is $\Lambda_\theta(t) = \theta t$. For $c_\theta(t)$, we choose a constant function, that is $c_\theta(t) = c$ for all $\theta \in \Theta$ and for all t. We take ν to be $\mathrm{G}(a, b)$ where $\mathrm{G}(a, b)$ is the Gamma distribution with shape parameter a and scale parameter b. Then the posterior distribution (3.1) of θ becomes $\mathrm{G}(\alpha, \beta)$ where

$$\alpha = q_n + a$$

$$\beta = c \sum_{i=1}^{n} \frac{T_{(i)}}{n-i+c} + b.$$

For numerical computation, we choose $a = 33, b = 3300$. The brief explanation of the value a and b can be found in Doss (1994). For c, we choose two values 100 and 0.3. $c = 100$ is chosen to be arbitrarily large to see how close the result is to a parametric analysis when c is large. $c = 0.3$ corresponds to a mixture of Dirichlet processes prior with constant α_θ (i.e. $\alpha_\theta \equiv \alpha > 0$) and $\alpha = 1$. This is because theoretically the Dirichlet process with mean H_θ and precision parameter α is equal to the beta process with parameters Λ_θ and $c_\theta(t)$ where Λ_θ is a cumulative hazard function of H_θ and $c_\theta(t) = \alpha(1 - H_\theta(t))$. See Hjort (1990) for the proof. Hence, if we use a mixture of Dirichlet processes with $H_\theta = 1 - \exp(-\theta t)$ (the exponential distribution) and $\alpha > 0$ on $\mathcal{F}$, then the corresponding prior on $\mathcal{A}$ is a mixture of beta processes with parameters $\Lambda_\theta = \theta t$ and

$$c_\theta(t) = \alpha \exp(-\theta t). \tag{4.1}$$

Since we investigate the distribution from 60 to 180 in Table 4.1, we can choose the middle value 120 and plug this value with $\alpha = 1$ and $\theta = 0.01$ (prior mean) to (4.1) and obtain the value 0.3 for c.

Table 4.1: Estimators of $F(t)$. KME: Kaplar Meyer estimator is in the first row with 95% confidence interval in the second row. The others: Posterior means are in the first row with 95% credible region in the second row (MD and MB mean mixture of Dirichlet priors and mixture of beta priors respectively).

t	60	120	150	180
KME	0.528 (0.437,0.605)	0.628 (0.528,0.707)	0.681 (0.531,0.783)	0.786
MD, $\alpha = 1$	0.519 (0.436,0.602)	0.630 (0.540,0.716)	0.683 (0.573,0.815)	0.786 (0.621,1.000)
MD, $\alpha = 100$	0.485 (0.416,0.555)	0.669 (0.592,0.742)	0.736 (0.667,0.809)	0.805 (0.724,0.879)
MB, $c = 0.3$	0.529 (0.440,0.614)	0.633 (0.539,0.716)	0.689 (0.580,0.830)	0.833 (0.647,1.000)
MB, $c = 100$	0.470 (0.410,0.545)	0.669 (0.599,0.743)	0.743 (0.670,0.806)	0.803 (0.737,0.865)

Table 4.1 compares the results from the mixture of beta processes priors with those from the mixture of Dirichlet processes priors as well as the standard frequentist method. To obtain the credible regions for mixtures of beta processes, first generate θs from ν^p by use of Metropolis-Hastings algorithm. Then for given θs, sample paths of the corresponding beta processes are generated by use of the method proposed by Damien et al. (1995). The results for KME and the mixture of Dirichlet processes priors come from Doss (1994). For large value of c, in particular $c = 100$, the results are close to the parametric analysis, which we expect beforehand. Also, it can be easily shown that as $c \to \infty$, the distribution of θ converges to that for the parametric model.

For small value of c, however, the situation is slightly different. The results are pretty much same as those for the mixture of Dirichlet processes priors with $\alpha = 1$ until time 150. But at 180, two methods give quite different results. This is because the posterior distributions of the hyperparameters for the two methods are quite different. In fact, between 150 and 180, there is only one uncensored observation, and so the probability of this interval strongly depends on the posterior distribution of hyperparameters. In the model with the mixture of beta processes prior, as $c \to 0$, the distribution of θ converges to the gamma distribution with $\alpha = q_n + a$ and $\beta = b + T_{(n)}$. On the other hand, for the mixture of Dirichlet processes model, the posterior distribution of θ is

$$\pi(\theta|\mathbf{N_n}) \sim \theta^q \exp\left(-\theta \sum_{i=1}^{q_n} t_i\right) \prod_{i=1}^{p_n} \prod_{j=0}^{d_n(s_i)-1} (\alpha e^{-\theta s_i} + Y_n(s_i) - d_n^*(s_i) + j)$$

where $s_1, \ldots, s_{p_n}$ are the distinct times of censored survival times and $d_n^*(s) = \sum_{i=1}^n I(T_i = s)$. Hence as $\alpha \to 0$, the posterior distribution of θ from the mixture of Dirichlet processes prior converges to the gamma distribution with $\alpha = q_n + a$ and $\beta = \sum_{i=1}^n T_i I(\delta_i = 1) + T_{(n)}(1 - \xi) + b$ where $\xi = \prod_{i \in R_n} \delta_i$ and $R_n = \{i : T_i = T_{(n)}\}$. If we compare the mean, we can see that the mixture of beta processes prior yields larger θ than the mixture of Dirichlet processes prior when the priors are diffuse.

In the above data analysis, we found that the two different choices of c_θ in the mixture of beta process priors yield qualitatively different posterior distributions of θ. That is, the Bayes estimator of θ with $c_\theta(t) = c > 0$ is larger than that of θ with $c_\theta(t) = \alpha(1 - H_\theta(t))$. Now, we may wonder which c_θ should be used. To answer this question, we should study theoretical properties of the posterior distributions. For example, suppose that the true unknown distribution is in the given parametric family H_θ (or equivalently Λ_θ). If we look at the behavior of the posterior distribution when the sample size is getting larger, we can easily find that in both cases the posterior distributions do not converge weakly to the point mass on the true parameter when there are nonnegligibly many censoring observations. That is, the both choices of c_θ yield the inconsistent posterior distributions. In most of literature of nonparametric Bayesian inference of survival analysis, little attention has been put on the choice of c_θ. However, our analysis shows that c_θ plays an important role for the posterior distribution to have desirable properties.

5. CONCLUDING REMARK

In this paper, we introduced a new class of priors - mixtures of beta processes for the chf and obtained the posterior distribution with right censored data. We observed that the choice of the precision parameter c_θ affects seriously the posterior distribution of θ and so the posterior distribution of the tail probabilities. Hence, in practice, c_θ should be selected with great care. One way of selecting a c_θ is to choose one which gives desirable large sample properties such as posterior consistency. Apparently, the two c_θ considered in this paper do not yield consistent posterior distributions. More results for this problem will appear elsewhere.

Mixture of beta process priors can be used for the cumulative intensity function of general counting processes such as left truncated right censored data and Markov processes. The posterior distribution can be shown to be again a mixture of beta processes. This is another advantage of considering the mixture of beta process priors rather than the mixture of Dirichlet process priors. The proof is available from the author.

Besides Dirichlet processes and beta processes, processes neutral to the right introduced by Doksum(1974) have been used for right censored data. F is called a process neutral to right if F is given by $F(t) = 1 - \exp\{-B(t)\}$ where B is a nondecreasing Levy process. Ferguson and Phadia (1979) showed that the posterior distribution of F is still a process neutral to right. We can possibly extend processes neutral to right priors to mixtures of processes neutral to right by use of the exactly same method as we did in this paper. However, the posterior distribution is not derived similarly since $\mathrm{E}(F(t)) \neq 1 - \exp(-\mathrm{E}(B(t))$. More works should be done for this problem.

References

Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152-1174.

Damien, P., Laud, P.W., and Smith, A.F.M. (1995). Approximate random variate generation from infinitely divisible distribution with application to Bayesian inference. *Journal of Royal Statistical Society* **B**, **57**, 547-563.

Doksum, K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Annals of Probability* **2**, 183-201.

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

Ferguson. T.S. and Phadia, E.G. (1979). Bayesian nonparametric estimation based on censored data. *Annals of Statistics* **7**, 163-186.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics* **18**, 1259-1294.

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. *Biometrics* **35**, 393-401.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.

Koziol, J.A. and Green, S.B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika* **63**, 465-474.

Miller, R.G., Jr. (1983). What price Kaplan-Meier? *Biometrics*, **39**, 1077-1081.

Walker, S. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the polya-urn scheme. *Annals of Statistics* **25** 1762-1780.