The Annals of Statistics
2001, Vol. 29, No. 3, 666–686

ON POSTERIOR CONSISTENCY OF SURVIVAL MODELS

BY YONGDAI KIM[1] AND JAEYONG LEE

Hankuk University of Foreign Studies, Korea and Pennsylvania State University

Ghosh and Ramamoorthi studied posterior consistency for survival models and showed that the posterior was consistent when the prior on the distribution of survival times was the Dirichlet process prior. In this paper, we study posterior consistency of survival models with neutral to the right process priors which include Dirichlet process priors. A set of sufficient conditions for posterior consistency with neutral to the right process priors are given. Interestingly, not all the neutral to the right process priors have consistent posteriors, but most of the popular priors such as Dirichlet processes, beta processes and gamma processes have consistent posteriors. With a class of priors which includes beta processes, a necessary and sufficient condition for the consistency is also established. An interesting counter-intuitive phenomenon is found. Suppose there are two priors centered at the true parameter value with finite variances. Surprisingly, the posterior with smaller prior variance can be inconsistent, while that with larger prior variance is consistent.

1. Introduction. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (iid) with an unknown cumulative distribution function (cdf) F on $[0, \infty)$ and suppose the data are subject to right censoring. Bayesian analysis of a model requires an appropriate class of priors and efficient computational methods. Since Dirichlet processes [Ferguson (1973)] were introduced as priors for nonparametric models, many classes of priors have been constructed and applied to survival models such as Dirichlet processes [Susarla and Van Ryzin (1976)], neutral to the right processes [Doksum (1974), Ferguson and Phadia (1979)], extended gamma processes [Dykstra and Laud (1981)], beta processes [Hjort (1990)] and beta-Stacy processes [Walker and Muliere (1997)]. Recently, Kim (1999) employed Poisson measures to represent Lévy process priors for multiplicative counting process models. This approach gives a simple representation of the prior and posterior and covers all neutral to the right process priors (note that Dirichlet processes, gamma processes, beta processes and beta-Stacy processes are all neutral to the right processes). For the computational side, Markov chain Monte Carlo (MCMC) computation schemes for various priors have been studied by many authors including Doss (1994), Damien, Laud and Smith (1996) and Wolpert and Ickstadt (1998).

Received July 1999; revised December 2000.

[1]Supported in part by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

AMS 2000 *subject classifications*. Primary 62G20; secondary 62M05.

Key words and phrases. Survival model, posterior consistency, Lévy process, neutral to the right process.

In contrast to the construction of suitable priors and their computation, theoretical properties of posteriors in survival models have received relatively little attention [see Ghosh and Ramamoorthi (1995) for an exception]. In this paper, we study the issue of posterior consistency in survival models with neutral to the right process priors. Recently, the issue of posterior consistency in nonparametric Bayesian models has been studied intensively and now a body of a fairly general theory exists [see, Schwartz (1965); Barron (1988, 1989), Barron, Schervish and Wasserman (1999); Ghosal, Ghosh and van der Vaart (2000), Shen and Wasserman (1998), Ghosal, Ghosh and Ramamoorthi (1999a, b)]. The theory, however, assumes the existence of a σ-finite measure which dominates all the distributions under consideration. This assumption is crucial for the general theory, because it allows the use of the Bayes theorem by which the posterior can be expressed in very general settings. Unfortunately, neutral to right priors put probability mass 1 to the set of all discrete distributions and the class of distributions under consideration includes all the discrete as well as continuous distributions. Thus, there does not exist a dominating σ- finite measure and the Bayes theorem cannot be used to represent posteriors. For this reason, survival models naturally fall outside the scope of the the theory developed in the papers mentioned above. It is necessary to take another route to study the consistency of posteriors for survival models. The route taken in this paper is to look at the limits of the first and second moments of the posterior using the approach studied in Kim (1999).

It turns out that not all neutral to the right processes have consistent posteriors. To describe an example of posterior inconsistency through simple moment calculations, we introduce a class of priors called *extended beta processes* which admits a relatively simple parametrization. Despite its simplicity, the class is quite large, including Dirichlet processes and beta processes. A necessary and sufficient condition for consistency with the extended beta processes priors can be characterized under very mild conditions. Then a general theorem is given with sufficient conditions for posterior consistency with neutral to right process priors.

In the course of investigation, a surprising phenomenon was found. Consider two priors, prior 1 and prior 2 which happen to be centered at the true parameter value and suppose prior 1 has smaller variance than prior 2. If the posterior with prior 2 is consistent, it is natural to expect that the posterior with prior 1 is also consistent, because prior 1 is expected to be more concentrated on the true value than prior 2. In survival models, however, this may not be the case. This example contradicts the usual belief that "the more mass around the true value a priori, the more mass around the true value a posteriori." This counter-intuitive phenomenon is dealt with in Section 3 in more detail.

In Section 2, we review basic facts of neutral to the right processes. In Section 3, a necessary and sufficient condition for consistency of the class of extended beta processes is given; with this class of prior, the interesting phenomenon is introduced in detail. A general theorem for consistency is given with sufficient conditions in Section 4 and its proof is given in Section 5.

2. Neutral to the right processes. In this section, we review several features of processes neutral to right as prior distributions for F. Let $\mathscr{F}$ be the space of cdfs on $[0, \infty)$. A process F defined on $\mathscr{F}$ is said to be a random distribution function neutral to right if it can be written in the form

$$F(t) = 1 - \exp(-Y(t)), \tag{2.1}$$

where $Y(t)$ is a nondecreasing stochastic process with independent increments (called Lévy process) with $Y(0) = 0$ and $Y(t) \to \infty$ as $t \to \infty$ with probability 1. As is well known, any nondecreasing Lévy process $Y(t)$ is a sum of a deterministic function and a jump process, and we assume that the deterministic function vanishes everywhere. Note that most of the practically important processes such as the aforementioned ones in the introduction are pure jump processes. From what follows, we simply use the term "Lévy process" for such a Lévy process. The definition (2.1) of processes neutral to right was originally introduced by Doksum (1974) and posterior distribution with censored observations was derived by Ferguson and Phadia (1979).

There are two approaches to Bayesian modeling of survival models. One approach is to put a Lévy process prior on Y directly and the other is to put a Lévy process prior on the cumulative hazard function (chf), A, defined by

$$A(t) = \int_0^t \frac{dF(s)}{1 - F(s-)}.$$

An advantage of the first approach is that one can choose the prior mean and variance without restriction, which is not the case for the second approach. Its main weak point is that Y is not really a cumulative hazard function for discrete distributions, although it is for continuous distributions. This can be problematic, because all the Lèvy process priors put probability 1 to discrete distributions.

The second approach was initiated by Hjort (1990), where he defined a process neutral to right by a cumulative hazard function. It can be shown that F is a process neutral to right if and only if A is a nondecreasing Lévy process such that $A(0) = 0, 0 \le \Delta A(t) \le 1$ for all t with probability 1 and either $\Delta A(t) = 1$ for some $t > 0$ or $\lim_{t\to\infty} A(t) = \infty$ a.s. Note that not all Lévy processes can be prior distributions of A since $\Delta A(t)$ should be bounded by 1.

A consequence of modeling A with a Lévy process is

$$\mathrm{Var}(A(t)) < E(A(t)). \tag{2.2}$$

This causes a problem in elicitation of the prior, because one's prior guess of the chf limits one's confidence on the guess to a certain degree. However, if one is willing to elicit one's prior on A by eliciting the hazard of tiny intervals or through dA, it is a natural consequence. A similar parametric example is the beta prior on a probability. Its variance is always smaller than its expectation.

In spite of this disadvantage, we will use Hjort's characterization of the process neutral to right for the following reasons. First, as was pointed out in Hjort (1990), A is as basic as F in survival analysis. Second, it can be easily extended to general counting process models such as left-truncated right-censored data and the Poisson process [see Kim (1999) for details]. Third, in the proof of posterior consistency we will use extensively the first and second moments of the posterior distribution of A. Since $\Delta A(t)$ is bounded by 1 for all t with probability 1, the two moments of A always exist [see Lemma 2 of Chapter IV, Gikhman and Skorokhod (1975)] while no such nice property exists for Y. So A is mathematically more tractable than Y.

Kim (1999) uses the following characterization of Lévy processes whose jump sizes are bounded by 1. This characterization may be dated back to Lévy [see the note in Breiman (1968), page 318]. Similar characterizations can also be found in Theorem 6.3 VIII of Daley and Vere-Jones (1988) and Theorem 3, page 606 of Fristedt and Gray (1997). For any given Lévy process $A(t)$ on $[0, \infty)$ with $0 \le \Delta A(t) \le 1$, there exists a unique random measure μ on $[0, \infty) \times [0, 1]$ such that

$$A(t) = \int_{[0,\, t]\times[0,\, 1]} x\mu(ds, dx). \tag{2.3}$$

In fact, μ is defined by

$$\mu([0, t] \times B) = \sum_{s\le t} I(\Delta A(s) \in B)$$

for any Borel subset B of $[0, 1]$ and for all $t > 0$. Since μ is a Poisson random measure [Jacod and Shiryaev (1987), page 70], there exists a unique σ-finite measure ν on $[0, \infty) \times [0, 1]$ such that

$$\mathrm{E}(\mu([0, t] \times B)) = \nu([0, t] \times B) \tag{2.4}$$

for all $t > 0$. Conversely, for a given σ-finite measure ν such that

$$\int_0^t \int_0^1 x\nu(ds, dx) < \infty \tag{2.5}$$

for all t, there exists a unique Poisson random measure μ on $[0, \infty) \times [0, 1]$ which satisfies (2.4) [Jacod (1979)] and so we can construct a Lévy process A through (2.3). Consequently, we can use ν to characterize a Lévy process A.

Suppose that a given Lévy process A has fixed discontinuity points at $t_1 < t_2 < \cdots$ and that the Lévy formula is given by

$$\mathrm{E}(\exp(-\theta A(t))) = \left[\prod_{t_i\le t} \mathrm{E}(\exp(-\theta\Delta A(t_i)))\right] \exp\left(-\int_0^1 (1 - e^{-\theta x}) dL_t(x)\right),$$

where $L_t(x)$ is the Lévy measure. Then it can be shown [see Theorem II.4.8 in Jacod and Shiryaev (1987)] that

$$\nu([0,t]\times B)=\int_B dL_t(x)+\sum_{t_i\le t}\int_B dG_i(x)$$

for all $t>0$ and for any Borel set B of $[0,1]$ where $G_i(x)$ is the distribution function of $\Delta A(t_i)$. When there are no fixed discontinuities, μ is a Poisson random measure defined on $[0,\infty)\times[0,1]$ with the intensity measure ν and $dL_t(x)=\int_{[0,t]}\nu(ds,dx)$. Hence, the measure ν simply extends dL_t by incorporating the fixed discontinuity points. However, this simple extension provides a convenient notational device. The posterior distribution, which typically has many fixed discontinuity points, can be summarized neatly by the corresponding measure ν without separating the stochastically continuous part and fixed discontinuity points as was done in the previous papers [Ferguson and Phadia (1979) and Hjort (1990)]. For this reason, we call ν simply the "Lévy measure" of A.

From the Lévy measure ν, we can easily calculate mean and variance of the Lévy process by the following two equations [Kim 1999)]:

$$\mathrm{E}(A(t))=\int_0^t\int_0^1 x\nu(ds,dx) \tag{2.6}$$

and

$$\mathrm{Var}(A(t))=\int_0^t\int_0^1 x^2\nu(ds,dx)-\sum_{s\le t}\left(\int_0^1 x\nu(\{s\},dx)\right)^2. \tag{2.7}$$

Suppose $X_1,\dots,X_n$ are iid with a true cdf F^* whose chf is A^*. In the usual random censorship model, we observe $T^n=(T_1,\dots,T_n)$ and $\delta^n=(\delta_1,\dots,\delta_n)$ where $T_i=\min(X_i,C_i)$, $\delta_i=I(X_i\le C_i)$ and $C_1,\dots,C_n$ are iid censoring times with a distribution function G which are independent with X's.

For a prior distribution of A, let A be a Lévy process with a Lévy measure ν. We suppose that the Lévy measure ν can be rewritten as

$$\nu([0,t]\times B)=\int_0^t\int_B f_s(x)\,dx\,ds \tag{2.8}$$

with

$$\lim_{t\to\infty}\int_0^t\int_0^1 xf_s(x)\,dx\,ds=\infty. \tag{2.9}$$

Then the posterior distribution of A given (T^n,δ^n) is again a Lévy process with the Lévy measure ν^p given by

$$\begin{aligned}\nu^p([0,t]\times B)&=\int_0^t\int_B(1-x)^{Y_n(s)}f_s(x)\,dx\,ds\\&\quad+\int_0^t c_n(s)^{-1}\int_B x^{\Delta N_n(s)}(1-x)^{Y_n(s)-\Delta N_n(s)}\\&\qquad\times f_s(x)\,dx\frac{1}{\Delta N_n(s)}dN_n(s),\end{aligned} \tag{2.10}$$

where $N_n(t) = \sum_{i=1}^n I(T_i \le t, \delta_i = 1), Y_n(t) = \sum_{i=1}^n I(T_i \ge t), \Delta N_n(t) = N_n(t) - N_n(t-)$ and

$$c_n(s) = \int_0^1 x^{\Delta N_n(s)}(1-x)^{Y_n(s)-\Delta N_n(s)} f_s(x)\,dx. \tag{2.11}$$

For the proof, see Hjort (1990) or Kim (1999).

In this paper, we have the following two assumptions:

(A1) There exists a positive constant τ such that $F^*(\tau-) < 1$ and $G(\tau-) < 1$.
(A2) A^* is continuous on $[0, \tau]$.

We will study the posterior distribution of A only on $[0, \tau]$. (A1) guarantees that $Y_n(\tau) \to \infty$ as $n \to \infty$, which is necessary for posterior consistency. If (A1) holds, however, for all $\tau > 0$, posterior consistency results are valid on $[0, \infty)$. (A2) is for technical purposes.

3. Posterior consistency of extended beta processes. Let λ_0, α and β be strictly positive continuous functions defined on $[0, \tau]$ and $A_0(t) = \int_0^t \lambda_0(s)\,ds$, for all $t \in [0, \tau]$. Let $b(x\colon a, b)$ be the density of the beta distribution with parameters $a, b > 0$, that is,

$$b(x\colon a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \quad \text{for } 0 < x < 1.$$

Consider a Lévy process with parameters $(A_0(t), \alpha(t), \beta(t))$ whose Lévy measure is given by

$$\nu([0, t] \times B) = \int_0^t \int_B \frac{1}{x} b(x\colon \alpha(s), \beta(s))\,dx\,dA_0(s).$$

We call it an *extended beta process*. Note that the class of extended beta processes includes the beta processes which are characterized by $\alpha(t) \equiv 1$. We first study the issue of posterior consistency in this restricted class of priors, because the class of extended beta processes is large enough to include both consistent and inconsistent priors and mathematically tractable enough to render a simple necessary and sufficient condition for the posterior consistency without much difficulty. The study of the extended beta processes tells us what feature in the prior process is important for consistency.

In this section, we construct a counter-intuitive example mentioned in Section 1 with the class of extended beta processes as priors. We say the posterior is consistent if the posterior probability of A on any ε-neighborhood of A^* with sup-norm converges to 1 with probability 1. That is, for any $\varepsilon > 0$,

$$\Pr\Big\{\sup_{t\le\tau} |A(t) - A^*(t)| < \varepsilon | T^n, \delta^n\Big\} \to 1 \tag{3.1}$$

with probability 1.

REMARK. The standard definition of the posterior consistency of A is that the posterior distribution of A converges weakly to the point mass at A^* in

$D[0, \tau]$ with probability 1 where $D[0, \tau]$ is the space of all real valued functions defined on $[0, \tau]$ which are right continuous and have left limits, equipped with Skorokhod topology. It is easy to verify that (3.1) implies the standard posterior consistency. We use (3.1) to avoid unnecessarily complicated Skorokhod topology.

The following theorem gives a necessary and sufficient condition for the posterior distribution of A to be consistent within the class of extended beta process priors.

THEOREM 3.1. *A priori, let A be an extended beta process with parameters $(A_0(t), \alpha(t), \beta(t))$ with $\lambda_0(t), \alpha(t)$ and $\beta(t)$ bounded and continuous on $t \in [0, \tau]$. Then the posterior distribution of A given $(T_1, \delta_1), \dots, (T_n, \delta_n)$ is consistent if and only if $\alpha(t) \equiv 1$; that is, an extended beta process prior has consistent posterior if and only if it is a beta process.*

PROOF. Since A^* is continuous, there is a set of probability 1 such that on that set $\Delta N_n(t) \le 1$ for all t, and hence by (2.10), the posterior distribution of A becomes a Lévy process with the Lévy measure

$$\begin{aligned}\nu^p([0, t] \times B) = &\int_0^t \int_B (1-x)^{Y_n(s)} \frac{1}{x} b(x\colon \alpha(s), \beta(s))\, dx\, dA_0(s)\\ &+ \int_0^t \int_B b(x\colon \alpha(s), \beta(s) + Y_n(s) - 1)\, dx\, dN_n(s).\end{aligned}$$

By (2.6), (2.7) and the moments formulas for beta distributions, we have

$$\begin{aligned}\mathrm{E}(A(t)|T^n, \delta^n) &= \int_0^t \int_0^1 x \nu^p(ds, dx)\\ &= \int_0^t \int_0^1 (1-x)^{Y_n(s)} b(x\colon \alpha(s), \beta(s))\, dx\, dA_0(s)\\ &\quad + \int_0^t \int_0^1 x b(x\colon \alpha(s), \beta(s) + Y_n(s) - 1)\, dx\, dN_n(s)\\ &= \int_0^t \frac{\Gamma(\alpha(s)+\beta(s))\Gamma(\beta(s)+Y_n(s))}{\Gamma(\beta(s))\Gamma(\alpha(s)+\beta(s)+Y_n(s))}\, dA_0(s)\\ &\quad + \int_0^t \frac{\alpha(s)}{\alpha(s)+\beta(s)+Y_n(s)-1}\, dN_n(s)\end{aligned} \tag{3.2}$$

and

$$\begin{aligned}\mathrm{Var}(A(t)|T^n, \delta^n) &= \int_0^t \int_0^1 x^2 \nu^p(ds, dx) - \sum_{s \le t} \left(\int_0^1 x \nu^p(\{s\}, dx) \right)^2\\ &= \int_0^t \int_0^1 x(1-x)^{Y_n(s)} b(x\colon \alpha(s), \beta(s))\, dx\, dA_0(s)\\ &\quad + \int_0^t \int_0^1 x^2 b(x\colon \alpha(s), \beta(s) + Y_n(s) - 1)\, dx\, dN_n(s)\end{aligned}$$

$$
\begin{aligned}
&-\int_0^t \left(\int_0^1 xb(x\colon \alpha(s), \beta(s)+Y_n(s)-1)\,dx\right)^2 dN_n(s) \\
&= \int_0^t \frac{\Gamma(\alpha(s)+\beta(s))\Gamma(\alpha(s)+1)\Gamma(\beta(s)+Y_n(s))}{\Gamma(\alpha(s))\Gamma(\beta(s))\Gamma(\alpha(s)+\beta(s)+Y_n(s)+1)}\,dA_0(s) \\
&\quad+\int_0^t \frac{\alpha(s)(\alpha(s)+1)}{(\alpha(s)+\beta(s)+Y_n(s)-1)(\alpha(s)+\beta(s)+Y_n(s))}\,dN_n(s) \\
&\quad-\int_0^t \frac{\alpha(s)^2}{(\alpha(s)+\beta(s)+Y_n(s)-1)^2}\,dN_n(s) \\
&= \int_0^t \alpha(s)\frac{\Gamma(\alpha(s)+\beta(s))\Gamma(\beta(s)+Y_n(s))}{\Gamma(\beta(s))\Gamma(\alpha(s)+\beta(s)+Y_n(s)+1)}\,dA_0(s) \\
&\quad+\int_0^t \frac{\alpha(s)(\beta(s)+Y_n(s)-1)}{(\alpha(s)+\beta(s)+Y_n(s)-1)^2(\alpha(s)+\beta(s)+Y_n(s))}\,dN_n(s).
\end{aligned}
\tag{3.3}
$$

By the fact that $Y_n(s) \to \infty$ for $s \le \tau$ together with the dominating convergence theorem, the first term in (3.2) converges to 0; by Lemma 1 in Section 5, the second term converges to $\int_0^t \alpha(s)\,dA^*(s)$. Similarly, the first and the second term in (3.3) converge to 0. Combining these facts, we have that

$$
\mathrm{E}(A(t)|T^n, \delta^n) \to \int_0^t \alpha(s)\,dA^*(s)
$$

and

$$
\mathrm{Var}(A(t)|T^n, \delta^n) \to 0
$$

with probability 1 for all $t \in [0, \tau]$. By Theorem A.1 in the Appendix the posterior probability of $A(\cdot)$ on any ε-neighborhood of $\int_0^{(\cdot)} \alpha(s)dA^*(s)$ with sup-norm converges to 1. Hence, the posterior is consistent if and only if $A^*(t) = \int_0^t \alpha(s)\,dA^*(s)$ for all $t \in [0, \tau]$ or, equivalently, $\alpha(t) \equiv 1$. □

REMARK. We have proved a stronger result than the statement of the theorem. In fact, we have proved:

If $\alpha(t) \equiv 1$, the posterior is consistent at all continuous chf.
If $\alpha(t) \not\equiv 1$, the posterior is inconsistent at all continuous chf.

Theorem 3.1 has interesting implications. Consider two priors, prior 1 and prior 2, which happen to be centered around the true value of the parameter. Suppose prior 1 has smaller variance than prior 2. Then it is natural to expect that the posterior with prior 1 is more concentrated at the true parameter than that with prior 2. Thus, if the posterior with prior 1 is consistent, the posterior with prior 2 is expected to be consistent. In survival model, however, this may not be the case. To take a look at this counterintuitive phenomenon closely, consider two extended beta processes 1 and 2 with parameters $(A^*, \alpha_1 \equiv 1/2, \beta \equiv 1)$ and $(A^*, \alpha_2 \equiv 1, \beta \equiv 1)$, respectively. Using the moment formulas of Lévy processes, it is not hard to see that

$$\mathrm{E}_1(A(t)) = \mathrm{E}_2(A(t)) = A^*(t) \quad \text{for all } t \in [0, \tau]$$

and

$$\begin{aligned} \mathrm{Var}_1(A(t)) &= \tfrac{1}{3}A^*(t), \\ \mathrm{Var}_2(A(t)) &= \tfrac{1}{2}A^*(t) \quad \text{for all } t \in [0, \tau], \end{aligned}$$

where the subscript i represents that the expectation and variance are with respect to the extended beta process i, for $i = 1, 2$. Thus, both processes have the same prior mean A^*, but process 1 has smaller prior variance. However, Theorem 3.1 implies that process 1 has an inconsistent posterior, while process 2 has a consistent posterior. This example contradicts the usual belief that "the more mass around the true value a priori, the more mass around the true value a posteriori."

Note that $f_t(x)$ in (2.8) governs the number as well as the sizes of jumps of a Lévy process. Since $A_0(t) = \int_0^t \lambda_0(s)\,ds$,

$$f_t(x) = \frac{\lambda_0(t)}{x} b(x\colon \alpha(t), \beta(t)),$$

for the extended beta process with parameter (A_0, α, β). The condition $\alpha(t) \equiv 1$ implies that the rate of $f_t(x)$ near 0 is crucial for consistency of the posterior and it has to be exactly

$$f_t(x) \approx c(t)\frac{1}{x} \quad \text{for } x \text{ near } 0, \tag{3.4}$$

for some positive function $c(t)$. Since $\int_0^1 1/x\,dx = \infty$, the Lévy process prior should have infinitely many infinitesimal jumps; however, too many infinitesimal jumps [e.g., $f_t(x) \approx c(t)/x^{3/2}$ or $\alpha(t) \equiv 1/2$ for extended beta processes] leads to an inconsistent posterior.

Note also that the extended beta process with $\alpha(t) > 1$ results in an inconsistent posterior. An explanation of this posterior inconsistency is that the process has finitely many jumps with probability 1, so it does not put its mass on the parameter space densely enough.

Hjort (1990) defined the beta process as the limit of time discrete models with beta priors. Such an intuitive characterization of the extended beta process with $\alpha(t) < 1$, however, is not available. Usually, time continuous models are considered to be the approximation of time discrete models, and this partially explains why only the beta process has the desirable large sample properties.

It can be postulated from Theorem 3.1 that (3.4) is necessary for consistency of posterior with general Lévy process priors. Indeed, in the next section, we use a similar condition to prove consistency of posterior with general Lévy process priors.

4. Main results. In this section, we give sufficient conditions for the consistency of posterior of A when the prior is a general Lévy process. It will be shown that our sufficient conditions include most of the practically used priors such as Dirichlet processes and gamma processes.

Assume that a priori A is a Lévy process with the Lévy measure given by

$$\nu([0,t]\times B)=\int_0^t\int_B \frac{1}{x} g_s(x)\,dx\,\lambda_0(s)\,ds, \tag{4.1}$$

where $\int_0^1 g_t(x)\,dx = 1$ for all $t\in[0,\tau]$. Assume that $\lambda_0(t)$ is bounded and positive on $(0,\tau)$.

REMARK. Comparing (2.8) and (4.1), we can see that

$$\lambda_0(t)=\int_0^1 x f_t(x)\,dx$$

and

$$g_t(x)=\frac{x f_t(x)}{\lambda_0(t)}$$

provided $\lambda_0(t)>0$.

REMARK. Positiveness of $\lambda_0(t)$ on $t\in(0,\tau)$ is necessary for posterior consistency. Suppose $\lambda_0(t)=0$ for $t\in[c,d]$ where $0<c<d<\tau$. Then, both the prior and posterior put mass 1 to the set of chfs, A, with $A(d)=A(c)$. Hence the posterior distribution cannot be consistent unless $A^*(d)=A^*(c)$ where A^* is the true chf.

For the general consistency result, we need the following two conditions:

(C1) $\sup_{t\in[0,\tau],\,x\in[0,1]}(1-x)g_t(x)<\infty$.

(C2) There exists a function $h(t)$ defined on $[0,\tau]$ such that

$$\lim_{x\to 0}\sup_{t\in[0,\tau]}|g_t(x)-h(t)|=0$$

and

$$0<\inf_{t\in[0,\tau]}h(t)\le\sup_{t\in[0,\tau]}h(t)<\infty.$$

(C2) is the main condition which is basically the same as that $\alpha(t)\equiv 1$ for posterior consistency with extended beta processes. The main idea of the proof is to approximate the posterior with a Lévy process prior by that with an extended beta process prior. (C1) is necessary for the approximation. The following theorem is the main theorem of the paper.

THEOREM 4.1. *Under* (C1) *and* (C2), *the posterior distribution of A given (T^n,δ^n) is consistent.*

The posterior consistency of the distribution itself follows immediately from the main theorem.

COROLLARY 1. *Under the same conditions in Theorem* 4.1, *the posterior distribution of F given* (T^n, δ^n) *is consistent*. *Here*, *the posterior consistency of F means that for any* $\varepsilon > 0$,

$$\Pr\Big\{\sup_{t\le\tau}|F(t) - F^*(t)| < \varepsilon|T^n, \delta^n\Big\} \to 1$$

with probability 1.

PROOF. We recover F from A by $F(t) = \prod_{s\le t}(1-dA(s))$. Since the product integration is a continuous mapping from $D[0, \tau]$ to $D[0, \tau]$ with sup-norm [Theorem 7 in Gill and Johansen (1990)], for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\Big\{\sup_{t\le\tau}|A(t) - A^*(t)| < \delta\Big\} \subset \Big\{\sup_{t\le\tau}|F(t) - F^*(t)| < \varepsilon\Big\}.$$

Since the posterior probability of $\{\sup_{t\le\tau}|A(t) - A^*(t)| < \delta\}$ converges to 1 with probability 1 by Theorem 4.1, so does the posterior probability of $\{\sup_{t\le\tau}|F(t) - F^*(t)| < \varepsilon\}$. $\square$

Posterior consistency of beta process priors (hence, of Dirichlet process priors) can be proved using Theorem 4.1, but the proof is omitted because the consistency is already shown in Theorem 3.1. In the following example, we show that gamma process priors have consistent posteriors.

EXAMPLE (Gamma process). A priori, assume that $Y(t) = -\log(1 - F(t))$ is a gamma process with parameters $(A_0(t), c(t))$ with $A_0(t) = \int_0^t \lambda_0(s)\,dx$, where $\lambda_0(t)$ is a positive bounded function on $t \in (0, \tau)$. Furthermore, assume that $c(t)$ is continuous around $t = 0$ and $0 < \inf_{t\in[0,\tau]} c(t) \le \sup_{t\in[0,\tau]} c(t) < \infty$. Here, the gamma process with parameters $(A_0(t), c(t))$ is defined by

$$Y(t) = \int_0^t \frac{1}{c(s)}\,dX(s),$$

where $X(t)$ is a Lévy process whose marginal distribution of $X(t)$ is a gamma distribution with parameters $(\int_0^t c(s)\,dA_0(s), 1)$. For details of this definition, see Lo (1982). This prior process was used by Doksum (1974), Ferguson and Phadia (1979) and Kalbflesch (1978). Since

$$\log \mathrm{E}(\exp(-\theta Y(t))) = \int_0^t \int_0^\infty (e^{-\theta x} - 1)\frac{c(s)}{x}\exp(-c(s)x)\,dx\,dA_0(s),$$

it can be shown that the chf A of F is a Lévy process with a Lévy measure given by

$$\nu([0, t] \times B) = \int_0^t c^*(s) \int_B \frac{1}{-\log(1 - x)}(1 - x)^{c(s)-1}\,dx d\tilde{A}_0(s),$$

where

$$c^*(t) = \left(\int_0^1 \frac{x}{-\log(1-x)} (1-x)^{c(t)-1}\, dx \right)^{-1}$$

and

$$\widetilde{A}_0(t) = \int_0^t \frac{c(s)}{c^*(s)}\, dA_0(s).$$

Note that $0 < \inf_{t\in[0,\tau]} c^*(t) \le \sup_{t\in[0,\tau]} c^*(t) < \infty$. Therefore, we have

$$g_t(x) = c^*(t) \frac{x}{-\log(1-x)} (1-x)^{c(t)-1}, \qquad 0 \le x \le 1$$

and $h(t) = c^*(t)$. Now (C1) follows immediately. For (C2), by the Taylor expansion, for all $x \in (0, \varepsilon)$ we have

$$\begin{aligned} &\left| 1 - \frac{x}{-\log(1-x)} (1-x)^{c(t)-1} \right| \\ &\quad \le \left| 1 - \frac{x}{-\log(1-x)} \right| + \frac{x}{-\log(1-x)} |c(t)-1| \varepsilon \max\{1, (1-\varepsilon)^{-2}\}. \end{aligned}$$

Since $-x/\log(1-x)$ converges to 1 as x tends to 0, we get (C2). Hence, by Theorem 4.1 the posterior is consistent.

5. Proofs of the main results. In this section, we prove Theorem 4.1 and lemmas necessary for the proofs of Theorems 3.1 and 4.1. As stated in Section 2, we assume (A1) and (A2) throughout this section.

LEMMA 1. *Let $X_1(t), X_2(t), \ldots$ be stochastic processes defined on $[0, \tau]$. Suppose that there exists a continuous function $X(t)$ defined on $[0, \tau]$ such that*

$$\lim_{n\to\infty} \sup_{t\in[0,\tau]} |X_n(t) - X(t)| = 0$$

with probability 1. *Then, for all $t \in [0, \tau]$,*

$$\lim_{n\to\infty} \int_0^t X_n(s) \frac{1}{Y_n(s)}\, dN_n(s) = \int_0^t X(s)\, dA^*(s) \qquad a.s.$$

PROOF. First, we shall prove that

$$\sup_{t\in[0,\tau]} \left| \int_0^t \frac{1}{Y_n(s)}\, dN_n(s) - A^*(t) \right| \to 0 \tag{5.1}$$

with probability 1. The distribution of T_1 is given by

$$H(t) = \Pr(T_1 \ge t) = \exp(-A^*(t) - B(t)),$$

where $B(t) = -\log(1 - G(t))$. Then the joint probability of T_1 and $\delta_1 = 1$ becomes

$$Q(t) = \Pr(T_1 \geq t, \delta_1 = 1) = \int_t^\infty H(s)\, dA^*(s).$$

Now we can write

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

and so the strong law of large numbers implies that for a fixed t,

$$N_n(t)/n \to 1 - Q(t) \quad \text{with probability 1,}$$

Since $Q(t)$ is a nonincreasing function, we can prove

$$\sup_{t \in [0, \tau]} |N_n(t)/n - (1 - Q(t))| \to 0 \tag{5.2}$$

with probability 1 similarly to the Gilivenko–Cantelli lemma. Again, the Gilivenko–Cantelli lemma implies that

$$\sup_{t \in [0, \tau]} |Y_n(t)/n - H(t)| \to 0 \tag{5.3}$$

with probability 1. Combining (5.2) and (5.3) with Lemma A.2 of Tsiatis (1981), we get

$$\sup_{t \in [0, \tau]} \left| \int_0^t \frac{1}{Y_n(s)}\, dN_n(s) - \int_0^t \frac{1}{H(s)} H(s)\, dA^*(s) \right| \to 0$$

with probability 1 and the proof of (5.1) is done.

The proof of the lemma can be done by combining (5.1) and Lemma A.2 in Tsiatis (1981). □

LEMMA 2. *For $i = 0, 1, \ldots, s \in [0, \tau]$, $c > 0$ and $\lambda > 0$,*

$$\limsup_{n \to \infty} \sup_{s \in [0, \tau]} n^{i+1} \int_{\lambda/n}^{1} x^i (1 - x)^{Y_n(s) - c}\, dx \leq \sum_{k=0}^{i} C_k \lambda^k \exp(-p\lambda) \quad a.s.,$$

where $p = \lim Y_n(\tau)/n$ and C_k are positive constants, for $k = 0, 1, \ldots, i$, depending on i and c, but not on λ and s.

PROOF. Without loss of generality, assume $0 < \lambda/n < 1$ and $Y_n(\tau) - c > 0$. We prove this lemma by induction on i. First, suppose $i = 0$,

$$\begin{aligned} n \int_{\lambda/n}^{1} (1 - x)^{Y_n(s) - c}\, dx &= \frac{n}{Y_n(s) - c + 1} \left(1 - \frac{\lambda}{n}\right)^{Y_n(s) - c + 1} \\ &\leq \frac{n}{Y_n(\tau) - c + 1} \left(1 - \frac{\lambda}{n}\right)^{n(Y_n(\tau) - c + 1)/n}. \end{aligned}$$

Since $Y_n(s)$ is nonincreasing in s and $p = \lim_{n\to\infty} \frac{Y_n(\tau)}{n} > 0$, the result follows for $i = 0$. Suppose the result holds for $i - 1$ with $i = 1, 2, \ldots$. Then, using integration by parts

$$n^{i+1}\int_{\lambda/n}^{1} x^i(1-x)^{Y_n(s)-c}\,dx \le \frac{n^{i+1}}{Y_n(s)-c+1}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{Y_n(s)-c+1} + \frac{ni}{Y_n(s)-c+1}n^i\int_{\lambda/n}^{1} x^{i-1}(1-x)^{Y_n(s)-c+1}\,dx.$$

Using the induction assumption, the result follows for all $i = 0, 1, \ldots$. □

LEMMA 3. *Suppose* (C1) *holds. Then, for* $i = 0, 1, \ldots,$

$$\sup_{t\in[0,\tau]}\int_0^t\int_0^1 x^i(1-x)^{Y_n(s)}g_s(x)\,dx\,\lambda_0(s)\,ds \to 0 \quad \textit{with probability } 1.$$

PROOF. Let $g^* = \sup_{x\in[0,1],\, s\in[0,\tau]}(1-x)g_s(x)$, which is finite by (C1). Using the fact $|x|^i \le 1$ for all $i = 0, 1, \ldots$ and that $Y_n(s)$ is nonincreasing in s, we have

$$\begin{aligned}
&\int_0^t\int_0^1 x^i(1-x)^{Y_n(s)}g_s(x)\,dx\lambda_0(s)\,ds\\
&\quad\le \int_0^\tau\int_0^1 x^i(1-x)^{Y_n(s)}g_s(x)\,dx\lambda_0(s)\,ds\\
&\quad\le \int_0^\tau\int_0^1 (1-x)^{Y_n(s)-1}(1-x)g_s(x)\,dx\lambda_0(s)\,ds\\
&\quad\le \int_0^\tau \frac{g^*}{Y_n(s)}\lambda_0(s)\,ds\\
&\quad\le \frac{g^*A_0(\tau)}{Y_n(\tau)} \to 0,
\end{aligned}$$

where $A_0(\tau) = \int_0^\tau \lambda_0(s)\,ds$. Since $Y_n(\tau) \to \infty$ a.s., the result follows. □

LEMMA 4. *Suppose* (C1) *and* (C2) *hold. Then, for* $i = 0, 1, \ldots,$

$$\sup_{s\in[0,\tau]}\int_0^1 x^i(1-x)^{Y_n(s)-1}|g_s(x)-h(s)|\,dx = o\left(\frac{1}{n^{i+1}}\right) \quad \textit{with probability } 1.$$

PROOF. Let

$$M = \sup_{x\in[0,1], s\in[0,\tau]}(1-x)[g_s(x)+h(s)]$$

and

$$p = \lim_{n\to\infty} Y_n(\tau)/n. \tag{5.4}$$

Note that $M < \infty$ by (C1) and (C2) and $p > 0$. Choose an arbitrary large positive number λ. For all positive integer n such that $\lambda/n < 1$, define

$$d_n = \sup_{x \in [0,\, \lambda/n],\, s \in [0,\, \tau]} |g_s(x) - h(s)|.$$

By (C2), $\lim_{n\to\infty} d_n = 0$. For all n with $\lambda/n < 1$ and $s \in [0, \tau]$, we have

$$\begin{aligned}
&n^{i+1} \int_0^1 x^i (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \\
&\quad \le n^{i+1} \left[\int_0^{\lambda/n} + \int_{\lambda/n}^1 \right] x^i (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \\
&\quad \le n^{i+1} d_n \int_0^{\lambda/n} x^i (1-x)^{Y_n(s)-1}\, dx \\
&\qquad + n^{i+1} \int_{\lambda/n}^1 x^i (1-x)^{Y_n(s)-2} (1-x) |g_s(x) - h(s)|\, dx \\
&\quad \le n^{i+1} d_n \left(\frac{\lambda}{n} \right)^i \int_0^{\lambda/n} dx + n^{i+1} M \int_{\lambda/n}^1 x^i (1-x)^{Y_n(s)-2}\, dx \\
&\quad \le \lambda^{i+1} d_n + n^{i+1} M \int_{\lambda/n}^1 x^i (1-x)^{Y_n(s)-2}\, dx.
\end{aligned}$$

By Lemma 2 and the fact that $\lim_{n\to\infty} d_n = 0$,

$$\limsup_{n\to\infty} \sup_{s\in[0,\, \tau]} n^{i+1} \int_0^1 x^i (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \le \sum_{k=0}^{i} C_k \lambda^k \exp(-\lambda p),$$

for some positive constants C_k independent of λ, for $k = 0, 1, 2, \ldots, i$. Since λ is arbitrarily large, the result follows. □

LEMMA 5. *Suppose* (C1) *and* (C2) *hold. Then, for* $i = 0, 1, \ldots,$

$$\begin{aligned}
&\sup_{s \in [0,\, \tau],\, \Delta N_n(s) = 1} \left| \int_0^1 x^i (1-x)^{Y_n(s)-1} \left(\frac{g_s(x)}{k_n(s)} - Y_n(s) \right) dx \right| \\
&\qquad = o\left(\frac{1}{n^i} \right) \quad \textit{with probability } 1,
\end{aligned}$$

where

$$k_n(s) = \int_0^1 (1-x)^{Y_n(s)-1} g_s(x)\, dx.$$

PROOF. For $s \in [0, \tau]$ with $\Delta N_n(s) = 1$,

$$\begin{aligned}
&\left| \int_0^1 x^i (1-x)^{Y_n(s)-1} \left(\frac{g_s(x)}{k_n(s)} - Y_n(s) \right) dx \right| \\
&\quad \le \frac{1}{k_n(s)} \int_0^1 x^i (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \qquad (5.5)\\
&\qquad + \frac{1}{k_n(s)} \int_0^1 x^i (1-x)^{Y_n(s)-1} |h(s) - k_n(s) Y_n(s)|\, dx.
\end{aligned}$$

It suffices to show supremums of two terms on the right-hand side of (5.5) over all $s \in [0, \tau]$ with $\Delta N_n(s) = 1$ converge to 0 with probability 1. In this proof, sup and inf are the supremum and infimum over all $s \in [0, \tau]$ with $\Delta N_n(s) = 1$, respectively.

First, we have

$$\begin{aligned}
\sup |Y_n(s) k_n(s) - h(s)| &= \sup \left| Y_n(s) \int_0^1 (1-x)^{Y_n(s)-1} g_s(x)\, dx \right. \\
&\qquad \left. - Y_n(s) \int_0^1 (1-x)^{Y_n(s)-1} h(s)\, dx \right| \\
&\le \sup Y_n(s) \int_0^1 (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \qquad (5.6)\\
&\le \sup n \int_0^1 (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \\
&= o(1) \quad \text{with probability 1,}
\end{aligned}$$

where the last equality is due to Lemma 4.

Consider the first term in (5.5). Since $Y_n(s) \le n$ for all n and s,

$$\inf\, n k_n(s) \ge \inf\, h(s) - \sup |h(s) - k_n(s) Y_n(s)|. \qquad (5.7)$$

By (C2) and (5.6) we see that $n k_n(s) > 0$ all but finitely many n. Hence, Lemma 4 implies

$$\sup \frac{n^{i+1}}{n k_n(s)} \int_0^1 x^i (1-x)^{Y_n(s)-1} |g_s(x) - h(s)|\, dx \to 0, \text{ with probability 1.}$$

Finally, consider the second term in (5.5):

$$\begin{aligned}
&n^i \sup \frac{1}{k_n(s)} \int_0^1 x^i (1-x)^{Y_n(s)-1} |h(s) - k_n(s) Y_n(s)|\, dx \\
&\quad \le n^i \sup \frac{|h(s) - k_n(s) Y_n(s)|}{k_n(s)} \int_0^1 x^i (1-x)^{Y_n(s)-1}\, dx \\
&\quad \le \sup \frac{|h(s) - k_n(s) Y_n(s)|}{k_n(s) Y_n(s)} \frac{n^i \Gamma(i+1)}{(Y_n(s)+1) \cdots (Y_n(s)+i)}.
\end{aligned}$$

Again using (C2) and (5.6) together we have

$$\inf\ k_n(s)Y_n(s) \geq \inf\ h(s) - \sup|h(s) - k_n(s)Y_n(s)| > 0,$$

all but finitely many n with probability 1. Hence the second term in (5.5) converges to 0 with probability due to (5.6). This completes the proof. □

PROOF OF THEOREM 4.1 Since the posterior distribution of A is also a Lévy process, by Theorem A.1 it suffices to show that

$$\mathrm{E}(A(t)|T^n, \delta^n) \to A^*(t) \tag{5.8}$$

and

$$\mathrm{Var}(A(t)|T^n, \delta^n) \to 0 \tag{5.9}$$

with probability 1, for all $t \in [0, \tau]$.

Noting that $k_n(s) = c_n(s)/\lambda_0(s)$ for $s > 0$ with $\Delta N_n(s) = 1$ and $N(0) = 0$ with probability 1, we have

$$\begin{aligned} \mathrm{E}(A(t)|T^n, \delta^n) = &\int_0^t \int_0^1 (1-x)^{Y_n(s)} g_s(x)\,dx\lambda_0(s)\,ds \\ &+ \int_0^t \frac{1}{k_n(s)} \int_0^1 x(1-x)^{Y_n(s)-1} g_s(x)\,dx\,dN_n(s). \end{aligned} \tag{5.10}$$

By Lemma 3, the first term on the right-hand side of (5.10) converges to 0 with probability 1. By adding and subtracting the same quantity and using Lemma 5, the second term on the right hand side of (5.10) is rewritten by

$$\begin{aligned} &\int_0^t Y_n(s) \int_0^1 x(1-x)^{Y_n(s)-1} \left|\frac{g_s(x)}{k_n(s)} - Y_n(s)\right| dx \frac{1}{Y_n(s)}\,dN_n(s) \\ &\quad + \int_0^t Y_n(s)^2 \int_0^1 x(1-x)^{Y_n(s)-1}\,dx \frac{1}{Y_n(s)}\,dN_n(s). \end{aligned} \tag{5.11}$$

By Lemmas 1 and 5 and the fact that $Y_n(s) \leq n$, the first term of (5.11) converges to 0 with probability 1. By the beta integral, the second term of (5.11) is

$$\int_0^t \frac{Y_n(s)}{Y_n(s)+1} \frac{1}{Y_n(s)}\,dN_n(s).$$

Since

$$\sup_{s\in[0,\tau]} \left|\frac{Y_n(s)}{Y_n(s)+1} - 1\right| \leq \frac{1}{Y_n(\tau)+1} \to 0 \quad \text{with probability 1},$$

the second term of (5.11) converges to $A^*(t)$ by Lemma 1.

For (5.9), we have

$$
\begin{aligned}
\operatorname{Var}(A(t)|T^n,\delta^n) &= \int_0^t\int_0^1 x(1-x)^{Y_n(s)}g_s(x)\,dx\,\lambda_0(s)\,ds\\
&\quad+\int_0^t\left[\frac{1}{k_n(s)}\int_0^1 x^2(1-x)^{Y_n(s)-1}g_s(x)\,dx\right.\\
&\qquad\left.-\left(\frac{1}{k_n(s)}\int_0^1 x(1-x)^{Y_n(s)-1}g_s(x)\,dx\right)^2\right]dN_n(s).
\end{aligned}\tag{5.12}
$$

The first term on the right-hand side of (5.12) converges to 0 with probability 1 by Lemma 3. Since the integrand of the second term of (5.12) is the variance of the density

$$\frac{1}{k_n(s)}(1-x)^{Y_n(s)-1}g_s(x)\quad\text{for } 0<x<1,$$

it is less than or equal to

$$\frac{1}{k_n(s)}\int_0^1 x^2(1-x)^{Y_n(s)-1}g_s(x)\,dx.$$

Hence, the second term on the right-hand side of (5.12) is less than or equal to

$$
\begin{aligned}
&\int_0^t Y_n(s)\int_0^1 x^2(1-x)^{Y_n(s)-1}\left|\frac{g_s(x)}{k_n(s)}-Y_n(s)\right|dx\,\frac{1}{Y_n(s)}\,dN_n(s)\\
&\quad+\int_0^t Y_n(s)^2\int_0^1 x^2(1-x)^{Y_n(s)-1}\,dx\,\frac{1}{Y_n(s)}\,dN_n(s)\\
&\le\int_0^t n\int_0^1 x^2(1-x)^{Y_n(s)-1}\left|\frac{g_s(x)}{k_n(s)}-Y_n(s)\right|dx\,\frac{1}{Y_n(s)}\,dN_n(s)\\
&\quad+\int_0^t\frac{Y_n(s)\Gamma(3)}{(Y_n(s)+1)(Y_n(s)+2)}\frac{1}{Y_n(s)}\,dN_n(s).
\end{aligned}
$$

By Lemmas 1 and 5, it converges to 0 with probability 1. □

APPENDIX

THEOREM A.1. *Let A_n be a sequence of nondecreasing stochastic processes with values in $D[0,\tau]$ such that for all $t\in[0,\tau]$,*

$$\operatorname{Var}(A_n(t))\to 0.$$

Suppose that there exists a nondecreasing deterministic continuous function A_0 in $D[0,\tau]$ such that

$$\mathrm{E}(A_n(t))\to A_0(t)\quad\textit{for all } t\in[0,\tau].$$

Then for any $\varepsilon > 0$,

$$\Pr\Big\{\sup_{t\le\tau}|A_n(t) - A_0(t)| < \varepsilon\Big\} \to 1.$$

PROOF. Since A_0 is a nondecreasing continuous function, we can choose a grid $0 = t_0 < t_1 < t_2 < \cdots < t_k = \tau$ such that

$$\max_{i=1,\ldots,k}|A_0(t_i) - A_0(t_{i-1})| < \varepsilon/3.$$

First, we will show that for any $\varepsilon > 0$,

$$\Pr\Big\{\max_{i=1,\ldots,k}|A_n(t_i) - A_0(t_i)| > \varepsilon\Big\} \to 0.$$

By Chebyshev inequality and the assumptions of the theorem,

$$\begin{aligned}
&\Pr\Big\{\max_{i=1,\ldots,k}|A_n(t_i) - A_0(t_i)| > \varepsilon\Big\}\\
&\quad\le \Pr\left\{\sqrt{\sum_{i=1}^{k}(A_n(t_i) - A_0(t_i))^2} > \varepsilon\right\}\\
&\quad\le \frac{1}{\varepsilon^2}\sum_{i=1}^{k}((A_0(t_i) - \mathrm{E}A_n(t_i))^2 + \mathrm{Var}(A_n(t_i))) \to 0.
\end{aligned}$$

Now, for given t with $t_{i-1} \le t < t_i$, let $t_* = t_{i-1}$ and $t^* = t_i$. Then we can write

$$\begin{aligned}
|A_n(t) - A_0(t)| &\le |A_n(t) - A_n(t_*)| + |A_n(t_*) - A_0(t_*)| + |A_0(t) - A_0(t_*)|\\
&\le |A_n(t^*) - A_n(t_*)| + |A_n(t_*) - A_0(t_*)| + |A_0(t^*) - A_0(t_*)|\\
&\le |A_n(t^*) - A_0(t^*)| + 2|A_n(t_*) - A_0(t_*)| + 2|A_0(t^*) - A_0(t_*)|\\
&\le 3\max_{i=1,\ldots,k}|A_n(t_i) - A_0(t_i)| + \tfrac{2}{3}\varepsilon.
\end{aligned}$$

Hence,

$$\Pr\Big\{\sup_{t\le\tau}|A_n(t) - A_0(t)| > \varepsilon\Big\} \le \Pr\Big\{\max_{i=1,\ldots,k}|A_n(t_i) - A_0(t_i)| > \varepsilon/9\Big\} \to 0$$

and the proof is done. □

Acknowledgments. The authors thank the Editor, an Associate Editor and the referees for helpful questions, comments and suggestions.

REFERENCES

BARRON, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical report, Univ. Illinois.

BARRON, A. R. (1989). Uniformly powerful goodness of fit tests. *Ann. Statist.* **17** 107–124.

BARRON, A., SCHERVISH, M. J. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536–561.

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, MA.

DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.

DAMIEN, P., LAUD, P. W. and SMITH, A. F. M. (1996). Implementation of Bayesian nonparametric inference based on beta processes. *Scand. J. Statist.* **23** 27–36.

DOKSUM, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201.

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

DYKSTRA, R. L. and LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367.

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

FERGUSON, T. S. and PHADIA, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7** 163–186.

FRISTEDT, B. and GRAY, L. (1997). *A Modern Approach to Probability Theory.* Birkhäuser, Boston.

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999a). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158.

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999b). Consistency issues in Bayesian nonparametrics. In *Asymptotic Nonparametrics and Time Series: A Tribute to Madan Lal Puri* (S. Ghosh, ed.) 639–667. Dekker, New York.

GHOSAL, S., GHOSH, J. K. and VAN DER VAART (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531.

GHOSH, J. K. and RAMAMOORTHI R. V. (1995). Consistency of Bayesian inference for survival analysis with or without censoring. In *Analysis of Censored Data* (H.L. Koul and J. V. Deshpande, eds.) **27** 95–103. IMS, Hayward, CA.

GIKHMAN, I. I. and SKOROKHOD, A. V. (1975). *Theory of Stochastic Processes II*. Springer, Berlin.

GILL, R. D. and JOHANSEN, S. (1990). A survey of product integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555.

HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294.

JACOD, J. (1979). *Calcul stochastique et problémes de martingales. Lecture Notes in Math.* **714**. Springer, Berlin.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, New York.

KALBFLEISCH, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **40** 214–221.

KIM, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27** 562–588.

LO, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrsch. Verw. Gebiete* **59** 55–66.

SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26.

SHEN, X. and WASSERMAN, L. (1998). Rates of convergence of posterior distributions. Technical report, Carnegie Mellon Univ.

SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71** 897–902.

TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108.

WALKER, S. and MULIERE, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.* **25** 1762–1780.

WOLPERT, R. L. AND ICKSTADT K. (1998). Simulation of Lévy random fields. *Practical Nonparametric and Semiparametric Bayesian Statistics*. *Lecture Notes in Statist.* **133** 227–242. Springer, New York.

DEPARTMENT OF STATISTICS
HANKUK UNIVERSITY OF FOREIGN STUDIES
KYUNGGI, 449-791
KOREA
E-MAIL: kimy@stat.hufs.ac.kr

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
326 THOMAS BUILDING
UNIVERSITY PARK, PENNSYLVANIA 16802
E-MAIL: leej@stat.psu.edu