Posterior Consistency for Right Censored Data

Jaeyong Lee*

October 13, 2003

Abstract

Ghosh and Ramamoorthi (1996) studied the 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 the posterior consistency of survival models with neutral to the right process priors which include Dirichlet process priors. A set of sufficient conditions for the 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. For extended beta processes, a necessary and sufficient condition for the consistency is also established.

1 Introduction

Since Ferguson (1973) introduced Dirichlet process priors as a class of nonparametric priors, there have been considerable research interests in this area among Bayesians. In the last three decades, new and more general nonparametric priors such as processes neutral to the right (Doksum 1974), tailfree process priors (Doksum 1974), and beta processes (Hjort 1990) have been developed. These give statisticians more freedom to choose priors appropriate for varying applications.

There is, however, a danger in blind usage of nonparametric priors. In parametric models, the Bayes estimators or the posterior distributions are all asymptotically optimal in frequentist criteria. Diaconis and Freedman (1986) however showed that this is not the case in nonparametric models. In their paper, they showed that in a simple nonparametric model, a location model with nonparametric error distribution, an innocent looking prior might yield a highly misleading

*The Department of Statistics, Seoul National University. Supported in part by KOSEF through Statistical Research Center for Complex Systems at Seoul National University

posterior, i.e. the posterior distribution might be concentrated at a wrong parameter value as the sample size increases. The implication of this result is that as the sample size increases the Bayesian with the prior described in their paper will be more and more sure of the wrong value of the parameter. This surprising observation has led to a host of research in Bayesian asymptotics that clarifies which are reasonable priors in general situations.

The issue of the posterior consistency in nonparametric Bayesian models has been studied intensively and now a body of a fairly general theory exists (see, Barron et al. 1999, Ghosal et al. 1999). 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 use of Bayes theorem by which the posterior can be expressed in a very general setting. 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 Bayes theorem can not be used to represent the posteriors. For this reason, survival models naturally fall outside of the scope of the the theory developed in the papers mentioned above. It is necessary to take another route to study the consistency of the posteriors for survival models. It turned out that the moments of the posterior distribution can be computed explicitly and their limiting behavior governs the consistency result.

To describe an example of the posterior inconsistency through simple moment calculations, we introduce a class of priors called *extended beta processes* which admits a relatively simple parameterization. Despite its simplicity, the class is quite large, including Dirichlet processes and beta processes. A simple necessary and sufficient condition for the consistency with the extended beta processes priors can be characterized under very mild conditions. Then a general theorem is given with sufficient conditions for the posterior consistency with neutral to right process priors.

2 Survival Models and Processes Neutral to the Right

Let $X_1, \cdots, X_n$ be iid survival times with cumulative distribution function (cdf) F and $C_1, \cdots, C_n$ be independent censoring times with cdf G, independent of $X_i's$. Since the observations are subject to right censoring, we observe only $(T_1, \delta_1), \cdots, (T_n, \delta_n)$, where $T_i = \min(C_i, X_i)$ and $\delta_i = I(X_i \le C_i)$. Let $D_n = \{(T_1, \delta_1), \cdots, (T_n, \delta_n)\}$. Let A be the cumulative hazard function (chf) of F, $A(t) = \int_0^t dF(s)/(1 - F(s-))$.

We say that a prior process on cdf F is a process neutral to the right, if the corresponding chf A is a nonstationary subordinator (a positive nondecreasing independent increment process) such that $A(0) = 0$, $0 \leq \Delta A(t) \leq 1$ for all t with probability one and either $\Delta A(t) = 1$ for some $t > 0$ or $\lim_{t\to\infty} A(t) = \infty$ with probability one. See Doksum (1974) for the original definition of processes neutral to the right. In what follows, the term *subordinator* is used for a prior process of chf A which induces a process neutral to the right on F.

For any given subordinator $A(t)$ on $[0,\infty)$, there exists a unique random measure μ on $[0,\infty)\times[0,1]$ such that

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

In fact, μ is defined by

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

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

$$\mathrm{E}(\mu([0,t]\times B)) = \nu([0,t]\times B) \tag{2}$$

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

$$\int_0^t \int_0^1 x\nu(ds,dx) < \infty$$

for all t, there exists a unique Poisson random measure μ on $[0,\infty)\times[0,1]$ which satisfies (2) (Jacod, 1979) and so we can construct a subordinator A through (1). Conclusively, we can use ν to characterize a subordinator A.

The characterization of subordinators with Lévy measures is also convenient in representing the posterior distribution, for the class of processes neutral to the right is conjugate with respect to right censored survival data. Suppose a priori A is a subordinator with Lévy measure

$$\nu(ds,dx) = f_s(x)dxds, \text{ for } s \geq 0 \text{ and } 0 \leq x \leq 1, \tag{3}$$

with $\lim_{t\to\infty} \int_0^t \int_0^1 xf_s(x)dxds = \infty$. Then, the posterior distribution of A given D_n is again a subordinator with Lévy measure ν^p given by

$$\nu^p(ds,dx) = (1-x)^{Y_n(s)} f_s(x)dxds + dH_s(x)\frac{1}{\Delta N_n(s)}dN_n(s), \tag{4}$$

where $H_s(x)$ is a distribution function on $[0,1]$ and is defined by

$$dH_s(x) \propto x^{\Delta N_n(s)}(1-x)^{Y_n(s)-\Delta N_n(s)} f_s(x)dx$$

and $N_n(t) = \sum_{i=1}^n I(T_i \le t, \delta_i = 1)$, $Y_n(t) = \sum_{i=1}^n I(T_i \ge t)$, $\Delta N_n(t) = N_n(t) - N_n(t-)$. Note that the posterior process is the sum of stochastically continuous and discrete parts, which correspond to the first and the second terms in (4), respectively. Note also that H_s is the distribution of jump size at s if $\Delta N_n(s) \neq 0$. This fact is used later. For the proof of (4), see Hjort (1990) or Kim (1999).

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 $b(x : a, b)$ be the density of the beta distribution with parameters $a, b > 0$, i.e.,

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

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

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

We call it an *extended beta process.* Note the class of extended beta processes includes the beta processes which are characterized by $\alpha(t) \equiv 1$. 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 one. That is, for any $\epsilon > 0$,

$$\Pr\{\sup_{t \le \tau} |A(t) - A^*(t)| < \epsilon | T^n, \delta^n\} \to 1 \tag{5}$$

with probability 1.

The next 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 *(Kim and Lee 2001) 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 $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ is consistent if and only if $\alpha(t) \equiv 1$, i.e., an extended beta process prior has consistent posterior if and only if it is a beta process.*

Note that $f_t(x)$ in (3) governs the number as well as sizes of jumps of a Lévy process. Since $A_0(t) = \int_0^t \lambda_0(s)ds$,

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

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 the consistency of posterior and it has to be exactly

$$f_t(x) \approx c(t)\frac{1}{x}, \text{ for } x \text{ near } 0, \tag{6}$$

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 one, so it does not put its mass on the parameter space densely enough.

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 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

$$\nu([0,t] \times B) = \int_0^t \int_B \frac{1}{x} g_s(x) dx \lambda_0(s) ds \tag{7}$$

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 Positiveness of $\lambda_0(t)$ on $t \in (0, \tau)$ is necessary for the 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) $\sup_{t \in [0,\tau], x \in [0,1]} (1 - x) g_t(x) < \infty$

and

(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

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

(C2) is the main condition which is basically the same as that $\alpha(t) \equiv 1$ for the 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 *(Kim and Lee 2001) 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 $\epsilon > 0$,*

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

with probability 1.

Under mild conditions, gamma process prior has also consistent posterior.

References

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

DIACONIS, P. AND FREEDMAN, D.A. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1-26.

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

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

GHOSAL, S., GHOSH, J.K. AND RAMAMOORTHI, R.V. (1999). 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* (Subir Ghosh Ed.), Marcel Dekker, Inc., 639-667.

GHOSH, J.K. AND RAMAMOORTHI R.V. (1996). Consistency of Bayesian inference for survival analysis with or without censoring. in *Analysis of Censored Data* (eds: Koul, H.L. and Deshpande, J.V.)

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 Mathematics, 714. Springer, Berlin.

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

KIM, Y. AND LEE, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29**, 666-686.