Statistica Sinica **6**(1996), 187-197

LIMITING POSTERIOR DISTRIBUTIONS UNDER MIXTURE OF CONJUGATE PRIORS

Lawrence J. Brunner and Albert Y. Lo

University of Toronto and Hong Kong University of Science & Technology

Abstract. Suppose the posterior distribution has a limiting distribution with respect to each member of a family of conjugate priors. Subject to a uniform boundedness condition on the prior parameters, the posterior distribution with respect to a mixture of member priors has the same limiting distribution. This result is used to show that a posterior distribution (given complete data) with respect to a mixture of Dirichlet processes prior can be approximated by a Brownian bridge. It also follows from this result that the limiting posterior distribution (given censored data) with respect to a mixture of beta-neutral processes prior is identical to the limiting sampling distribution of the Kaplan-Meier estimator.

Key words and phrases: Mixtures of conjugate priors, mixtures of beta neutral processes priors, mixtures of Dirichlet processes priors, mixtures of weighted gamma processes priors, limiting posterior distributions.

1. Introduction

Normal approximations to posterior distributions originated in work by Laplace, Bernstein, von Mises, and LeCam (1953); (see LeCam (1986)). The technique developed can be conveniently applied if the parameter of interest lies in a finite dimensional space, and the prior distribution is smooth. In the case of an infinite dimensional parameter space, the usual way of studying normal approximations for nonparametric models consists of showing that a posterior distribution with respect to a class of conjugate priors has a limiting Gaussian distribution. See, for example, Lo (1993) for some recent results and references in this direction. In this paper we show that the limiting posterior distribution with respect to conjugate priors is preserved for posterior distributions with respect to mixtures of conjugate priors. More precisely, we show that, subject to a uniform boundedness condition on the mixing prior parameters, limiting posterior distributions can be obtained for mixture priors. The idea of this approach is rather simple: a posterior distribution with respect to a mixture of a family of member priors is a mixture of posterior distributions with respect to member priors of the family. That is, a posterior with respect to a mixture prior is a mixture of sub-posteriors. A consequence is that if the sub-posteriors are well behaved, so is their mixture.

If a family of priors forms a conjugate family, the posteriors are also members of this family, and the limiting posterior distribution of the mixture of sub-posteriors can be identified. The dimension of the parameter space is irrelevant in this approach, and hence this method can be used to yield asymptotic posterior distributions with respect to mixtures of conjugate priors for parametric and nonparametric models alike. The application of this method to derive a limiting posterior distribution with respect to a mixture of conjugate priors for parametric models (see DeGroot (1970), Dalal and Hall (1983)) is quite direct and will not be given. It should also be pointed out that the Bayesian hierarchical models discussed by Lindley and Smith (1972) concern mixtures of normal priors for linear models, and is another instance where the present method can be applied.

In this paper, the emphasis is on Bayesian nonparametric problems. Our technique is used to transfer the limiting posterior distribution with respect to a conjugate prior to that with respect to a mixture of conjugate priors; two examples are given. Section 2 discusses the case of mixtures of Dirichlet processes priors (Ferguson (1973), Antoniak (1974)) for complete data. In this case, Freedman and Diaconis (1983), Section 5 showed that the posterior distribution is consistent if the prior parameters are subject to a uniform boundedness condition; our technique applied to this case actually identifies the limiting posterior distribution, under the same boundedness condition. Section 3 discusses the right censored data case and a beta-neutral processes prior (Hjort (1990), Lo (1993)). Mixtures of beta-neutral processes priors are defined. A limiting posterior distribution with respect to a mixture of beta-neutral process priors given right-censored data is obtained in this section. As a consequence, the posterior distribution of the survival function given right-censored data converges to a point mass at the "true" survival function, settling a conjecture of Doss (1991).

We discuss some basic techniques of the method in the rest of this section. A mixture prior for a parameter θ is defined by

$$\mathcal{L}\{\phi\} = v(d\phi), \quad \mathcal{L}\{\theta \mid \phi\} = \pi(d\theta \mid \phi). \tag{1.1}$$

In this setting, ϕ can be conveniently regarded as an mixing index of the prior distribution of the parameter θ. The model distribution is defined by

$$\mathcal{L}\{\boldsymbol{X}|\phi, \theta\} = P(d\boldsymbol{x}|\theta). \tag{1.2}$$

[Note that the model indicates that ϕ, θ, $\boldsymbol{X}$ is a three term Markov chain in the order written.] Given $\boldsymbol{X} = \boldsymbol{x}$, the posterior distribution of the parameter of interest is $\pi(d\theta|\boldsymbol{x}) = \int \pi(d\theta|\phi, \boldsymbol{x})v(d\phi|\boldsymbol{x})$, where $v(d\phi|\boldsymbol{x})$ is the conditional distribution of ϕ given $\boldsymbol{X} = \boldsymbol{x}$.

Our method depends on an explicit construction of random variables with distributions identical to the posterior distributions with respect to conjugate priors. Suppose the mixing variable $\phi \in \mathcal{T}$ and that the parameter θ is an element of Θ, which is a subset of a linear space equipped with a convex norm $|\cdot|$.

Lemma 1.1. *Assume model* (1.1) *and* (1.2). *Let* $\boldsymbol{X} = \boldsymbol{x}$ *be given, and* n *be the sample size. Suppose random variables* $\{\theta_{n,\phi} : \phi \in \mathcal{T}\}$ *are defined on a probability space such that, for each* ϕ, $\theta_{n,\phi}$ *has distribution* $\pi(d\theta|\phi, \boldsymbol{x})$. *Let* $E_{\boldsymbol{x}}\{\cdot\}$ *be the expectation* (*conditional on* $\boldsymbol{X} = \boldsymbol{x}$) *on this probability space, and* $\hat{\theta} = \int\int \theta\pi(d\theta|\phi, \boldsymbol{x})v(d\phi|\boldsymbol{x}) = \int E_{\boldsymbol{x}}\{\theta_{n,\phi}\}v(d\phi|\boldsymbol{x})$ *be the posterior mean. Suppose there exist random variables* θ_n *which are independent of* ϕ *with* $\hat{\theta}_n = E_{\boldsymbol{x}}\{\theta_n\}$, *and an increasing sequence* $b(n)$, *such that*

$$\lim_n \sup \sup_\phi E_{\boldsymbol{x}}\{b(n)|\theta_n - \theta_{n,\phi}|\} = 0. \tag{1.3}$$

Then $\mathcal{L}\{b(n)(\theta_n - \hat{\theta}_n)|\boldsymbol{x}\}$ *and the posterior distribution* $\mathcal{L}\{b(n)(\theta - \hat{\theta}_n)|\boldsymbol{x}\}$ *have the same limit;* (1.4)

furthermore, $b(n)|\hat{\theta}_n - \hat{\theta}| \to 0$, *and* (1.4) *remains valid if either one of* $\hat{\theta}_n$ *is replaced by* $\hat{\theta}$.

Proof. It suffices to note that for each bounded and contracting Lipschitz function h,

$$|E_{\boldsymbol{x}}\{h[b(n)(\theta_n - \hat{\theta}_n)]\} - \int\int h[b(n)(\theta - \hat{\theta}_n)]\pi(d\theta|\phi, \boldsymbol{x})v(d\phi|\boldsymbol{x})| \to 0; \tag{1.5}$$

see for example Pollard (1984) and LeCam (1986). The left side of (1.5) equals

$$\begin{aligned} &\Big|E_{\boldsymbol{x}}\{h[b(n)(\theta_n - \hat{\theta}_n)]\} - \int E_{\boldsymbol{x}}\{h[b(n)(\theta_{n,\phi} - \hat{\theta}_n)]\}v(d\phi|\boldsymbol{x})\Big| \\ &= \Big|\int E_{\boldsymbol{x}}\{h[b(n)(\theta_n - \hat{\theta}_n)] - h[b(n)(\theta_{n,\phi} - \hat{\theta}_n)]\}v(d\phi|\boldsymbol{x})\Big| \\ &\leq b(n)\int E_{\boldsymbol{x}}|\theta_n - \theta_{n,\phi}|v(d\phi|\boldsymbol{x}) \leq \sup_\phi E_{\boldsymbol{x}}[b(n)|\theta_n - \theta_{n,\phi}|] \to 0, \end{aligned}$$

where the first inequality follows from Jensen's inequality. Similarly, the second statement follows from

$$b(n)|\hat{\theta}_n - \hat{\theta}| = b(n)\Big|E_{\boldsymbol{x}}\{\theta_n\} - \int E_{\boldsymbol{x}}\{\theta_{n,\phi}\}v(d\phi|\boldsymbol{x})\}\Big|$$
$$\le b(n)\int E_{\boldsymbol{x}}|\theta_n - \theta_{n,\phi}|v(d\phi|\boldsymbol{x}) \le \sup_{\phi} E_{\boldsymbol{x}}\{b(n)|\theta_n - \theta_{n,\phi}|\},$$

where the first inequality follows from Jensen's inequality.

Lemma 1.1 states that, subject to the uniform convergence condition (1.3), the behavior of the mixing distribution $v(d\phi|\boldsymbol{x})$ is irrelevant. This is of interest since Diaconis and Freedman (1986) showed that $v(d\phi|\boldsymbol{x})$ may not behave appropriately in location models; see Section 2.

As we indicated in the beginning, an interesting case for the application of Lemma 1.1 is that $\{\pi(d\theta|\phi) : \phi \in \mathcal{T}\}$ is a conjugate family of priors for the model distribution $\{P(d\boldsymbol{x}|\theta) : \theta \in \Theta\}$. Lemma 1.1 transfers the limiting posterior distribution based on a member of a conjugate family of priors (i.e., the limit of $\mathcal{L}\{b(n)(\theta_n - \hat{\theta}_n)|\boldsymbol{x}\}$) to that with respect to a mixture of conjugate priors (i.e., the limit of $\mathcal{L}\{b(n)(\theta - \hat{\theta}_n)|\boldsymbol{x}\}$). In nonparametric problems where a model density does not exist, the method based on Lemma 1.1 would appear to be promising. A case in point is where Θ is the space of distribution functions or cumulative hazards on a Euclidean space. Here, Θ is a subset of the "cadlag" space, i.e., the space of right continuous functions with left limits equipped with the uniform norm and the projection σ-field. Lemma 1.1 can be applied if the limit of $\mathcal{L}\{b(n)[\theta_n - \hat{\theta}_n]|\boldsymbol{x}\}$ is concentrated on a separable subset of the "cadlag" space. This is the usual case of interest in statistics; (see Pollard (1984)).

In the following sections, we use Lemma 1.1 to derive limiting posterior distributions with respect to mixtures of conjugate nonparametric priors in the setting described above.

Remark 1.1. Lemma 1.1 states that for each sample sequence $\omega = (x_1, \ldots, x_n, \ldots)$ such that Assumption (1.3) is valid, the conclusion of Lemma 1.1 is also valid (note that $\boldsymbol{x}$ is the initial segment of the sequence ω). Therefore, if ω is random and has a (joint) distribution, the validity of Assumption (1.3) with probability one implies that the conclusion of Lemma 1.1 is also valid with probability one. In case Assumptions (1.3) is valid in probability, the conclusion of Lemma 1.1 is also valid in probability.

Remark 1.2. Lemma 1.1 is also useful for sample theorists. Suppose Assumption (1.3) is valid in probability; then the sampling distributions $\mathcal{L}\{b(n)[\hat{\theta}_n - \theta_0]|\theta_0\}$ and $\mathcal{L}\{b(n)[\hat{\theta} - \theta_0]|\theta_0\}$ have the same limit (for each θ_0). This provides an easy proof for those who are interested in the limiting sampling distribution of the posterior mean; (see for example Susarla and Van Ryzin (1979)).

2. Mixtures of Dirichlet Process Priors for Complete Data Models

Suppose $v(d\phi)$ is the distribution of ϕ, $\mathcal{L}\{F|\phi\}$ is a Dirichlet process (Ferguson (1973)) with shape α_ϕ (denoted by $D(dF|\alpha_\phi)$), and $X_1,\ldots,X_n|(\phi,F)$ are iid F. Let $\boldsymbol{x}=\{x_1,\ldots,x_n\}$ be the data (i.e. $\boldsymbol{X}=\boldsymbol{x}$), and let $\hat{F}_n(\cdot)=n^{-1}\sum_i\delta_{x_i}(\cdot)$ be the empirical distribution function, giving mass n^{-1} to each x_i, $i=1,\ldots,n$. (δ_y is a point mass at y.) The posterior distribution of F given $\boldsymbol{x}$ is $\pi(dF|\boldsymbol{x})=\int D(dF|\alpha_\phi+n\hat{F}_n)v(d\phi|\boldsymbol{x})$; (see Antoniak (1974), who also derived $v(d\phi|\boldsymbol{x})$ explicitly). Let Z_i's be iid exponential random variables, and given the data $\boldsymbol{x}$, let $D_n(\cdot)=\sum_i Z_i\delta_{x_i}(\cdot)/(\sum_i Z_i)$.

Given ϕ and the data $\boldsymbol{x}$, let μ_ϕ be a gamma process with shape measure α_ϕ, and let μ_ϕ and the Z_i's be independent. Then

$$F_{n,\phi}(\cdot)=[\mu_\phi(\cdot)+\textstyle\sum_i Z_i D_n(\cdot)]/[\mu_\phi(\infty)+\sum_i Z_i]$$

is a $D(dF|\alpha_\phi+n\hat{F}_n)$ process. Easy computation shows that

$$nE_{\boldsymbol{x}}[\sup_t|F_{n,\phi}(t)-D_n(t)|]\le\alpha_\phi(\infty). \tag{2.1}$$

Therefore,

Theorem 2.1. *Under the uniform boundedness condition* $\sup_\phi\alpha_\phi(\infty)<\infty$,

$$\mathcal{L}\{n^{1/2}[F(\cdot)-\hat{F}_n(\cdot)]|\boldsymbol{x}\}\quad\textit{and}\quad\mathcal{L}\{n^{1/2}[D_n(\cdot)-\hat{F}_n(\cdot)]|\boldsymbol{x}\}$$

have the same limit in $D[-\infty,\infty]$ *with the uniform norm and the projection* σ*-field; the conclusion also holds if either one of* $\hat{F}_n$ *is replaced by the posterior mean* $\hat{F}(t)=\int F(t)\pi(dF|\boldsymbol{x})$.

Freedman and Diaconis (1983) showed that the finiteness of $\sup_\phi\alpha_\phi(\infty)$ ensures consistent posterior distributions. Theorem 2.1 states that the finiteness of $\sup_\phi\alpha_\phi(\infty)$ actually yields a limiting posterior distribution.

Since the simulation of $D_n(\cdot)$ is the basis of the Bayesian bootstrap (Rubin (1981)), Theorem 2.1 implies that the Bayesian bootstrap is asymptotically correct in approximating posterior distributions of F with respect to a mixture of Dirichlet processes priors.

Let us close this section with a discussion of the special case of a nonparametric location model. Suppose $\alpha_\phi(A)=\alpha(A-\phi)$ for any event A where α is a finite measure. Then

$$\mathcal{L}\{F|\phi\}=D(dF|\alpha(\cdot-\phi)),\ \text{and}\ X_1,\ldots,X_n|(\phi,F)\ \text{are iid}\ F$$

is equivalent to

$$\mathcal{L}\{G|\phi\} = D(dG|\alpha(\cdot)), \text{ and } X_1 - \phi, \ldots, X_n - \phi|(\phi, G) \text{ are iid } G(\cdot). \tag{2.2}$$

In this notation, $F(\cdot) = G(\cdot - \phi)$.

Model (2.2) is the location model with a Dirichlet prior on the shape distribution G. Diaconis and Freedman (1986), and Doss (1985), showed that the posterior distribution of the location parameter ϕ, $v(d\phi|\boldsymbol{x})$, does not converge to a point mass at the "true" location ϕ_0 in general. On the other hand, since $\sup_\phi \alpha(A - \phi) \le \alpha(\infty) < \infty$, the above mentioned result of Freedman and Diaconis (1983) implies that the posterior distribution of $F(\cdot) = G(\cdot - \phi)$ (given complete data) is consistent; Theorem 2.1 and the following Remark 2.1 identify the limiting posterior distribution of $F(\cdot)$ as a Brownian bridge.

Remark 2.1. The limit of $\mathcal{L}\{n^{1/2}[D_n(\cdot) - \hat{F}_n(\cdot)]|\boldsymbol{x}\}$ was studied in Lo (1987). The arguments there show that: $\sup_t |\hat{F}_n(t) - F_0(t)| \to 0$ implies $\mathcal{L}\{n^{1/2}[D_n(\cdot) - \hat{F}_n(\cdot)]|\boldsymbol{x}\} \to \mathcal{L}\{B(F_0(\cdot))\}$, where F_0 is a distribution function, and $B(s) = W(s) - sW(1)$, $0 \le s \le 1$, is a Brownian bridge.

Remark 2.2. The choice of $D_n(\cdot)$ to play the role of θ_n in Lemma 1.1 is technically convenient in view of the large-sample theory developed in Lo (1987). On the other hand, $\mathcal{L}\{D_n(\cdot)|\boldsymbol{x}\}$ is not a posterior distribution with respect to some genuine prior distribution. Perhaps a more desirable choice is a particular member of the conjugate family of priors in question. In the present situation, such a choice is given by a fixed $F_{n,\phi}(\cdot)$, say,

$$F_{n,\phi_0}(\cdot) = [\mu_{\phi_0}(\cdot) + nD_n(\cdot)]/[\mu_{\phi_0}(\infty) + n].$$

Theorem 2.1 remains valid since the key inequality (2.1) is

$$nE_{\boldsymbol{x}}[\sup_t |F_{n,\phi}(t) - F_{n,\phi_0}(t)|] \le E_{\boldsymbol{x}}\mu_\phi(\infty) + E_{\boldsymbol{x}}\mu_{\phi_0}(\infty) = \alpha_\phi(\infty) + \alpha_{\phi_0}(\infty).$$

3. Mixtures of Beta-Neutral Processes Priors for Right Censored Data Models

Let α and β be two finite measures on $[0, \infty)$, and μ_α and μ_β be two independent gamma processes with shapes α and β, respectively. In Lo (1993), a beta-neutral $(\alpha; \beta)$ survival process (Hjort (1990)) is defined as

$$S_{\alpha,\beta}(t) = \prod_{y:y\le t} \{1 - \Delta\mu_\alpha(y)/[\mu_\alpha[y, \infty) + \mu_\beta[y, \infty)]\},$$

where $\Delta\mu_\alpha(y)$ is the size of the jump of μ_α at y. Lo (1993) called $S_{\alpha,\beta}$ a Bayesian copy of the Kaplan-Meier function (based on μ_α and μ_β). The beta-neutral process is a special case of the neutral process discussed by Doksum (1974); see also Ferguson and Phadia (1979). A result of Hjort (1990) implies that beta-neutral $(\alpha;\beta)$ processes are conjugate priors when the data are right-censored. Right censored data are defined as follows.

Let F be a distribution on $[0,\infty)$, and $S = 1-F$ be the corresponding survival function. Suppose $T_1,\ldots,T_n|S$ are iid S. Let $\mathcal{L}\{C_i\} = G_i$ where C_i's are censoring variables; the C_i's and the T_i's are independent. Define $Y_i = \min\{T_i, C_i\}$, and $X_i = (Y_i, I_{\{T_i \le C_i\}})$ for $i = 1,\ldots,n$. Denote the Y_i's corresponding to $I_{\{T_i\le C_i\}} = 1$ by Y_u, the other Y_i's by Y_c.

According to Hjort (1990), (see, however, Remark 3.2 below)

$$\mathcal{L}\{S\} \text{ is a beta-neutral } (\alpha;\beta) \text{ process}$$
$$\text{implies } \mathcal{L}\{S|\boldsymbol{x}\} \text{ is a beta-neutral } (\alpha + \sum_u \delta_{y_u}; \beta + \sum_c \delta_{y_c}) \text{ process.}$$

Next we define a mixture of beta-neutral processes prior for S. Let $v(d\phi)$ be the distribution of $\phi = (k,\lambda)$. $\mathcal{L}\{S|\phi\}$ is a beta-neutral $(\alpha_k;\beta_\lambda)$ process, and $T_1,\ldots,T_n|(\phi,S)$ are iid S. Then the above arguments already showed that

$$\mathcal{L}\{S|\phi,\boldsymbol{x}\} \text{ is a beta-neutral } (\alpha_k + \sum_u \delta_{y_u}; \beta_\lambda + \sum_c \delta_{y_c}) \text{ process.} \tag{3.1}$$

Let $B_N(dS|\alpha;\beta)$ be the distribution of a beta-neutral $(\alpha;\beta)$ process. With obvious notation, averaging out the $\phi = (k,\lambda)$ in (3.1) results in a posterior distribution of S given $\boldsymbol{x}$

$$\pi(dS|\boldsymbol{x}) = \iint B_N(dS|\alpha_k + \sum_u \delta_{y_u}; \beta_\lambda + \sum_c \delta_{y_c}) v(d(k,\lambda)|\boldsymbol{x}),$$

where $v(d(k,\lambda)|\boldsymbol{x})$ is the conditional distribution of ϕ given $\boldsymbol{x}$. Denote the posterior mean $\int S(t)\pi(dS|\boldsymbol{x})$ by $\hat{S}(t)$.

We begin by establishing an equality similar to (2.1) in Section 2; this equality will then imply (1.3) in Lemma 1.1. First, we construct a probability space in which all "member posteriors" are defined. Suppose the data $\boldsymbol{x}$ are given. Let Z_i's be iid exponential random variables. Define independent gamma processes $\mu_{n,u} = \sum_u Z_u\delta_{x_u}$ and $\mu_{n,c} = \sum_c Z_c\delta_{x_c}$. Let $\mu_n = \mu_{n,u} + \mu_{n,c}$, and define

$$S_n(t) = \prod_{y:y\le t} \{1 - \Delta\mu_{n,u}(y)/\mu_n[y,\infty)\}.$$

Let $\hat{S}_n(t) = E_{\boldsymbol{x}} S_n(t)$ be the Kaplan-Meier estimator.

For each $\phi = (k, \lambda)$ and given $\boldsymbol{x}$, let μ_{α_k} and μ_{β_λ} be two independent gamma processes with shapes α_k and β_λ, respectively. The μ_{α_k} and μ_{β_λ} are assumed to be independent of the $Z_i's$. Let $\mu_\phi = \mu_{\alpha_k} + \mu_{\beta_\lambda}$, and $\mu_{\alpha_n} = \mu_{\alpha_k} + \mu_{n,u}$. For each ϕ, construct a "posterior beta-neutral process" by

$$S_{n,\phi}(t) = \prod_{y:y\le t} \{1 - \Delta\mu_{\alpha_n}(y)/[\mu_\phi[y,\infty) + \mu_n[y,\infty)]\}.$$

$S_{n,\phi}(\cdot)$ has a beta-neutral $(\alpha_k + \sum_u \delta_{x_u};\ \beta_\lambda + \sum_c \delta_{x_c})$ distribution.

Lemma 3.1. *Suppose*

$$\lim_{n\to\infty} \hat{S}_n(b) = S_0(b) > 0 \quad \textit{and} \quad \lim_{n\to\infty} n^{-1} \sum_i I_{\{y_i \ge b\}} = H_0[b,\infty)^{-1} > 0. \tag{3.2}$$

Then

$$\limsup_n \sup_\phi E_{\boldsymbol{x}} \sup_{t\le b} n|S_n(t) - S_{n,\phi}(t)|/\hat{S}_n(t)$$
$$\le [\sup_k \alpha_k(b) + \sup_\lambda \beta_\lambda(b)] \times S_0(b)^{-1} H_0[b,\infty)^{-1}[1 + H_0[b,\infty)^{-1}].$$

Proof. Use the elementary inequality $|\prod_i a_i - \prod_i b_i| \le \sum_i |a_i - b_i|$ for a_i's and b_i's with $|a_i| \le 1$ and $|b_i| \le 1$ to get,

$$\sup_{t\le b} |S_n(t) - S_{n,\phi}(t)|\hat{S}_n(t)$$
$$\le [\mu_{\alpha_k}(b) + \mu_{\beta_\lambda}(b)]\hat{S}_n(b)^{-1}\mu_n[b,\infty)^{-1}[1 + \mu_n(b)/\mu_n[b,\infty)].$$

Therefore, since μ_{α_k} and μ_{β_λ} are independent of μ_n,

$$nE_{\boldsymbol{x}}[\sup_{t\le b} |S_n(t) - S_{n,\phi}(t)|/\hat{S}_n(t)]$$
$$\le n\hat{S}_n(b)^{-1} E_{\boldsymbol{x}}[\mu_{\alpha_k}(b) + \mu_{\beta_\lambda}(b)] E_{\boldsymbol{x}}\{\mu_n[b,\infty)^{-1}[1 + \mu_n(b)\mu_n[b,\infty)^{-1}]\}$$
$$= [\alpha_k(b) + \beta_\lambda(b)]\hat{S}_n(b)^{-1} E_{\boldsymbol{x}}\{n\mu_n[b,\infty)^{-1}[1 + \mu_n(b)\mu_n[b,\infty)^{-1}]\}$$
$$\le [\alpha_k(b) + \beta_\lambda(b)]\hat{S}_n(b)^{-1} E_{\boldsymbol{x}}\{n\mu_n[b,\infty)^{-1}[1 + n\mu_n[b,\infty)^{-1}]\}.$$

Note that the assumption entails

$$\limsup_n \hat{S}_n(b)^{-1} E_{\boldsymbol{x}}\{n\mu_n[b,\infty)^{-1}[1 + n\mu_n[b,\infty)^{-1}]\}$$
$$\le S_0(b)^{-1} H_0[b,\infty)^{-1}[1 + H_0[b,\infty)^{-1}].$$

Let $D[0,b]$ be the space of cadlag functions equipped with the uniform norm and the projection σ-field. In this section, random functions are regarded as elements in $D[0,b]$.

Theorem 3.1. *Assume* (3.2). *Both* $\sup_k \alpha_k(b)$ *and* $\sup_\lambda \beta_\lambda(b)$ *are finite imply*

$$\mathcal{L}\{n^{1/2}[S(\cdot)/\hat{S}_n(\cdot)-1]|\boldsymbol{x}\} \text{ and } \mathcal{L}\{n^{1/2}[S_n(\cdot)/\hat{S}_n(\cdot)-1]|\boldsymbol{x}\} \text{ have the same limit.} \tag{3.3}$$

Furthermore, (3.3) *continues to hold if either one of* $\hat{S}_n$ *is replaced by the posterior mean* $\hat{S}$.

Since the simulation of $S_n(\cdot)$ is the basis for the censored data Baysian bootstrap (Lo (1993)), Theorem 3.1 implies that the censored data Bayesian bootstrap discussed in Lo (1993) is asymptotically correct in approximating posterior distributions of S with respect to a mixture of beta-neutral processes priors.

Recently Doss (1991) conjectured the consistency of the posterior distribution of the distribution function $F = 1 - S$ given right censored data with respect to a mixture of Dirichlet processes priors. Theorem 3.1 shows more: since a Dirichlet process with shape α_ϕ is a beta-neutral $(\alpha_\phi; 0)$ process (Hjort (1990) and Lo (1993)), Theorem 3.1 and the following (3.5) specialize to yield a limiting posterior distribution of S with respect to a mixture of Dirichlet processes priors, implying that the posterior distribution of S converges to a point mass at S_0 at the rate of $O(n^{-1/2})$.

Remark 3.1. The limiting distribution of $\mathcal{L}\{n^{1/2}[S_n(\cdot)/\hat{S}_n(\cdot)-1]|\boldsymbol{x}\}$ in Theorem 3.1 was derived in Lo (1993). The required assumption is that the data $\boldsymbol{x}$ obey the following laws of large numbers:

Assumption 3.1. There exist (sub)distribution functions $1 - S_0$ and H_0 such that

$$\sup_{t\le b}|\hat{S}_n(t) \to S_0(t)| \to 0 \quad \text{and} \quad \sup_{t\le b}\Big|n^{-1}\sum_i I_{\{y_i\ge t\}} - [1 - H_0(t^-)]\Big| \to 0, \tag{3.4}$$

where $b < \inf\{t : H_0(t) \ge 1\}$.

The first condition is the consistency of the Kaplan-Meier estimator, and the second one is the consistency of an empirical distribution function. Theorem 5.1 in Lo (1993) states that Assumption 3.1 implies

$$\mathcal{L}\{n^{1/2}[S_n(\cdot)/\hat{S}(\cdot)-1]|\boldsymbol{x}\} \to \mathcal{L}\{W(C_0(\cdot))\}, \tag{3.5}$$

where $C_0(t) = \int_0^t\{S_0(s)[1 - H_0(s^-)]\}^{-1}S_0(ds)$, and $W(s)$, $s \ge 0$ is a Brownian motion. The limit in (3.5) is the limiting distribution of the Kaplan-Meier estimator; (see Breslow and Crowley (1974)). Since Assumption 3.1 implies (3.2), conclusions of Theorem 3.1 are valid.

We have already noted in Remark 1.1 that if (3.4) in Assumption 3.1 is valid for almost all $\boldsymbol{x}$, the conclusions of Theorem 3.1 and (3.5) are valid for almost all $\boldsymbol{x}$.

Remark 3.2. Hjort (1990) proved the conjugate prior property of the beta-neutral processes under the conditions that (i) both $\alpha(t)$ and $\beta(t)$ are piecewise continuous, and (ii) $\alpha(t)$ jumps only a finite number of times.

4. Concluding Remarks

The proposed technique for finding a limiting posterior distribution with respect to mixtures of conjugate priors can also be applied to other Bayesian nonparametric problems. A list includes sampling from a nonhomogeneous Poisson process with or without censored data (Grenander (1981), Lo (1982, 1992)), and sampling without replacement from a finite population (Lo (1988)). Details are quite similar to those provided in Sections 2 and 3 and are omitted.

Acknowledgement

This paper is dedicated to Lucien LeCam on the occasion of his seventieth birthday. Research of the first author is supported in part by the Natural Sciences and Mathematics Research Council of Canada Grant OPGN 014. Research of the second author is supported in part by the NSF of USA Grant DMS-91-06460.

References

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

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. **2**, 437-453.

Dalal, S. R. and Hall, W. J. (1983). Approximating priors by mixtures of Natural conjugate priors. J. Roy. Statist. Soc. Ser.B **45**, 278-286.

DeGroot, M. H. (1970). Optimal statistical decisions. McGraw-Hill, Inc.

Diaconis, P. and Freedman, D. A. (1986). On inconsistent Bayes estimates of Location. Ann. Statist. **14**, 68-87.

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

Doss, H. (1985). Bayesian nonparametric estimation of the median; Part II: Asymptotic properties of the estimates. Ann. Statist. **13**, 1445-1464.

Doss, H. (1991). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. Statistics Research Report #850, Florida State University.

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. (1979). Bayesian nonparametric estimation based on censored data. Ann. Statist. **7**, 163-186.

Freedman, D. A. and Diaconis, P. (1983). On inconsistent Bayes estimates in the discrete case. Ann. Statist. **11**, 1109-1118.

Grenander, U. (1981). Abstract Inference. John Wiley, New York.

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

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for linear model (with discussion). J. Roy. Statist. Soc. Ser.B **34**, 1-41.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. Univ. Calif. Publ. in Statist. **1**, 277-330.

LeCam, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New York.

Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point process. Z. Wahrsch. Verw. Gebiete **59**, 55-66.

Lo, A. Y. (1987). A large sample study for the Bayesian bootstrap. Ann. Statist. **15**, 360-375.

Lo, A. Y. (1988). A Bayesian bootstrap for a finite population. Ann. Statist. **16**, 1684-1695.

Lo, A. Y. (1992). Bayesian inference for Poisson process models with censored data. J. Nonparametr. Statist. **2**, 71-80.

Lo, A. Y. (1993). A Bayesian bootstrap for censored data. Ann. Statist. **21**, 100-123.

Pollard, D. (1984). Convergence of stochastic processes. Springer-Verlag, New York.

Rubin, D. B. (1981). The Bayesian bootstrap. Ann. Statist. **9**, 130-134.

Susarla, V. and Van Ryzin, J. (1979). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. Ann. Statist. **6**, 755-768.

Department of Statistics, University of Toronto, Toronto M5S 1A, Canada.

Department of Information and Systems Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

(Received February 1994; accepted December 1994)