Statistica Sinica **19** (2009), 581-595

A BERNSTEIN-VON MISES THEOREM FOR DOUBLY CENSORED DATA

Yongdai Kim

Seoul National University

Abstract: We prove a Bernstein-von Mises theorem for the survival function based on doubly censored data. In particular, we develop a new technique for proving Bernstein-von Mises theorems for nonparametric problems. We consider two Bayesian approaches for doubly censored data: the direct approach, where we obtain the posterior of the distribution of the survival times by putting the Dirichlet process prior on the distribution of the survival times; an indirect approach, where we first obtain the posterior of the distribution of the observables with the Dirichlet process and from which we get the posterior of the distribution of the survival times. We show that the two posterior distributions from these two approaches are the same. Using this fact, we prove a Bernstein-von Mises theorem.

Key words and phrases: Bernstein-von Mises theorem, doubly censored data, survival model.

1. Introduction

Let X_i $i = 1, \ldots, n$, be independent identically distributed (i.i.d.) survival times with a common distribution function F_X. Under the doubly censoring mechanism, the survival times can be censored either from the right or the left. Let $Y_i \geq Z_i$, be i.i.d. pairs of right and left censoring times, independent of X_i, with marginal distribution functions F_Z and F_Y, that may have total mass less than 1. We observe only the pairs of (W_i, δ_i),

$$(W_i, \delta_i) = \begin{cases} (X_i, 1) & \text{if} \quad Z_i < X_i \leq Y_i \\ (Y_i, 2) & \text{if} \quad X_i > Y_i \\ (Z_i, 3) & \text{if} \quad X_i \leq Z_i \end{cases}.$$

Here δ_i is the censoring indicator whose value is 1 for the uncensored case, and 2 and 3 for the right and left censored cases, respectively. Based on the observations $\mathrm{D}_n = \{(W_1, \delta_1), \ldots, (W_n, \delta_n)\}$, we wish to estimate the distribution of the survival times F_X. Doubly censored data arise in many medical and reliability applications. For examples of doubly censored data, see Turnbull (1974) and Cai and Cheng (2004).

Hypothesis testing and estimation procedures under the doubly censoring were studied by Gehan (1965), Mantel (1967), Turnbull (1974) and Mykland and Ren (1996). Recently Cai and Cheng (2004) considered statistical analysis of doubly censored data when there are covariates. The estimation procedures are heavily based on the self-consistent equations and, the asymptotic properties of the self-consistent estimator (SCE) have been studied by Chang and Yang (1987), Chang (1990), and Gu and Zhang (1993).

In this paper, we consider the Bayesian nonparametric approach to the estimation of the distribution function based on doubly censored data. Suppose that a priori F_X is a Dirichlet process on $[0, \infty)$ with a base measure α whose support is $[0, \infty)$. For the definition and properties of the Dirichlet process, see Ferguson (1973) and Ghosh and Ramamoorthi (2003). The objective of this paper is to prove a Bernstein-von Mises theorem for F_X, that is, that

$$\mathcal{L}(\sqrt{n}(F_X - F_X^{(n)})|\mathrm{D}_n) \xrightarrow{d} W$$

on $\mathcal{D}[0, \infty)$ in probability, where $\mathcal{L}(\cdot|\mathrm{D}_n)$ is the posterior distribution of F_X given D_n, W is the limiting distribution of the sampling distribution of $\sqrt{n}(F_X^{(n)} - F_X^0)$ under regularity conditions, and $\mathcal{D}[0, \infty)$ is the space of right continuous functions on $[0, \infty)$ with left limits existing equipped with the Skorohod topology. Here, $F_X^{(n)}$ is the NPMLE and F_X^0 is the true distribution of X_i. For the NPMLE $F_X^{(n)}$ and the limit sampling distribution of $\sqrt{n}(F_X^{(n)} - F_X^0)$, see Chang and Yang (1987), Chang (1990), Gu and Zhang (1993), and Mykland and Ren (1996).

The Bernstein-von Mises theorems in parametric models have a long history, dating back to Laplace. In early studies of nonparametric models, there was doubt as to whether a Bernstein-von Mises theorem would hold, see Freedman (1999). Positive results include Lo (1983), Conti (1999), Shen (2002), and Kim and Lee (2004), among others. For a more detailed history of Bernstein-von Mises theorems, see Kim and Lee (2004), Shen (2002), and Ghosh and Ramamoorthi (2003). For the asymptotic properties of the posterior distribution of right censored data, see Kim and Lee (2001, 2004), and Kim (2006).

In general, there are two approaches for proving a Bernstein-von Mises theorem for nonparametric problems. The first approach considers a case where the prior mass is concentrated on the space of probability measures having a dominating σ-finite measure. Here densities exist, and so we can prove the result by calculating the size of the support of the prior in terms of various entropies and the degree of concentration of the prior mass around the true model. Shen (2002) took this route. The second approach considers the case where densities do not exist, but the closed form of the posterior distribution is available. In this case, we can prove a Bernstein-von Mises theorem by directly calculating the moments of the posteriors. See Kim and Lee (2004) and Kim (2006) for this approach. For doubly censored data, however, no density exists and no closed form of the posterior is available, and these two approaches are not directly applicable.

In this paper, we develop a new technique. We consider two Bayesian approaches for doubly censored data: the direct approach where we obtain the posterior of the distribution of the survival times by putting the Dirichlet process prior on the distribution of the survival times; an indirect approach where we first obtain the posterior of the distribution of the observables with the Dirichlet process, from which we get the posterior of the distribution of the survival times. We show that the two posterior distributions from these two approaches are the same. Using this, we prove the Bernstein-von Mises theorem.

The paper is organized as follows. In Section 2, we describe our two Bayesian approaches for doubly censored data. In Section 3, we prove that the posterior distributions of F_X from the two Bayesian approaches are the same. Using this equivalence, we prove the main result in Section 4. To illustrate our theoretical findings, we present simulation results in Section 5.

2. Direct and Indirect Bayesian Approaches

For inference on F_X, we consider two Bayesian approaches.

- Direct : We put a Dirichlet process prior with a base measure α on F_X and obtain the posterior, $\mathcal{L}_D(\cdot|\mathrm{D}_n)$, for F_X directly given the data.
- Indirect : We put a Dirichlet process prior with a base measure β on the distribution Q of the observables, (W_i, δ_i), and obtain the posterior of Q given the data. Using this , we obtain the posterior, $\mathcal{L}_I(\cdot|\mathrm{D}_n)$, of F_X by (2.1).

In the indirect approach, the posterior distribution of Q, denoted by $\mathcal{L}_Q(\cdot|\mathrm{D}_n)$, is the Dirichlet process with base measure $\beta^p(\cdot) = \beta(\cdot) + \sum_{i=1}^n I((W_i, \delta_i) \in \cdot)$. Here is how, we recover F_X, as well as F_Y and F_Z, from Q. Let $Q_k(t) = Q((t,\infty), \delta = k)$ for $k = 1,2,3$. Let $S_X = 1 - F_X$, $S_Y = 1 - F_Y$, and $S_Z = 1 - F_Z$. By Chang and Yang (1987), We can write

$$
\begin{aligned}
Q_1(t) &= -\int_t^\infty \Big(S_Y(u) - S_Z(u)\Big) dS_X(u), \\
Q_2(t) &= -\int_t^\infty S_X(u) dS_Y(u), \qquad\qquad (2.1)\\
Q_3(t) &= -\int_t^\infty \Big(1 - S_X(u)\Big) dS_Z(u).
\end{aligned}
$$

Let U be the map from $\mathbf{S} = (S_X, S_Y, S_Z)$ to $\mathbf{Q} = (Q_1, Q_2, Q_3)$ given at (2.1). And write $V = U^{-1}$. The explicit form of the V is not known, from (2.1), one can show

$$
\begin{aligned}
S_X(t) &= Q.(t) - \int_{u \le t} \frac{S_X(t)}{S_X(u)} dQ_2(u) \\
&\quad + \int_{t<u} \frac{1-S_X(t)}{1-S_X(u)} dQ_3(u), \qquad (2.2) \\
S_Y(t) &= 1 + \int_0^t \frac{dQ_2(u)}{S_X(u)}, \\
S_Z(t) &= -\int_t^\infty \frac{dQ_3(u)}{1-S_X(u)},
\end{aligned}
$$

where $Q.(t) = \sum_{k=1}^{3} Q_k(t)$. The SCE of S_X is computed by using (2.2) iteratively. In the indirect approach, once $\mathcal{L}_Q(\mathbf{Q}|\mathrm{D}_n)$ is obtained, theoretically the posterior of F_X can be obtained as $\mathcal{L}_I(\mathbf{S}|\mathrm{D}_n) = \mathcal{L}_Q(V(\mathbf{Q})|\mathrm{D}_n)$.

3. Equivalence of the Two Approaches

Let $\beta_1(\cdot) = \alpha(\cdot)$ and $\beta_2([0,\infty)) = \beta_3([0,\infty)) = 0$. In this section, we show that the posterior of S_X of the indirect approach is well-defined by proving that a measurable version of U^{-1} exists, and then prove that the posterior distributions of F_X of the two Bayesian approaches are the same. Throughout this section, we assume that

$$\{W_i : \delta_i = 1\} \cap \{W_i : \delta_i \neq 1\} = \emptyset. \qquad (3.1)$$

3.1. Existence of U^{-1}

Let $\mathcal{Q}$ be the set of probability measures on $[0,\infty)\times\{1,2,3\}$. Since any $Q \in \mathcal{Q}$ can be identified by $\mathbf{Q}$, without loss of generality, we let $\mathcal{Q} = \mathcal{Q}_1 \times \mathcal{Q}_2 \times \mathcal{Q}_3$ where $\mathcal{Q}_k = \{Q_k : Q \in \mathcal{Q}\}$ and consider $\mathcal{Q}$ as a subspace of $\mathcal{D}^3[0,\infty)$. Let $\mathcal{D}_{\mathcal{I}}$ be the set of nonincreasing nonnegative right continuous functions on $[0,\infty)$ which are bounded by 1 and have left limits. Note that $\mathcal{D}_I$ is also thought to be a subspace of $\mathcal{D}[0,\infty)$.

For any subspace $\mathcal{S}$ of $\mathcal{D}^3[0,\infty)$, we let $\mathcal{B}_{\mathcal{S}}$ be the Borel σ-field on $\mathcal{D}^3[0,\infty)$ (with respect to the Skorohod topology) restricted to $\mathcal{S}$. The question addressed in this subsection is whether U defines an inverse V, a measurable mapping from $(\mathcal{Q}, \mathcal{B}_{\mathcal{Q}})$ to $(\mathcal{D}_I^3, \mathcal{B}_{\mathcal{D}_I^3})$. There are three difficulties: it is not clear that (2.1) has a solution for all $Q \in \mathcal{Q}$; if exists, the solution may not be unique (see Gu and Zhang (1993) and Mykland and Ren (1996) for examples); the measurability is by no means obvious. Measurability is important for our purpose since we have to derive the posterior of S_X from the posterior of Q. In this subsection, we show that, for given data D_n, there exists a measurable mapping V from $(\mathcal{Q}, \mathcal{B}_{\mathcal{Q}})$ to $(\mathcal{D}_I^3, \mathcal{B}_{\mathcal{D}_I^3})$, that satisfies (2.1) with probability 1 with respect to $\mathcal{L}_Q(\cdot|\mathrm{D}_n)$.

Our strategy is to construct a measurable subset $\mathcal{Q}_0$ of $\mathcal{Q}$ with $\mathcal{L}_Q(\mathcal{Q}_0|\mathrm{D}_n) = 1$ such that (2.1) has a unique solution in $\mathcal{Q}_0$ and the induced mapping is measurable from $(\mathcal{Q}, \mathcal{B}_\mathcal{Q})$ to $(\mathcal{D}_I^3, \mathcal{B}_{\mathcal{D}_I^3})$. Then, since $\mathcal{Q}_0$ is measurable, we can easily extend this mapping from $\mathcal{Q}_0$ to $\mathcal{Q}$.

Let $\Psi_2 = \{u_1 < \cdots < u_{n_2}\} = \{W_i : \delta_i = 2\}$ and $\Psi_3 = \{v_1 < \cdots < v_{n_3}\} = \{W_i : \delta_i = 3\}$. Let $\Psi_{23} = \Psi_2 \cup \Psi_3 = \{0 = z_0 < z_1 < z_2 < \cdots < z_{n_{23}} < z_{n_{23}+1} = \infty\}$. Let $\mathcal{Q}_{01}$ be the set of nonincreasing nonnegative right continuous step functions Q_1 on $[0, \infty)$ such that $Q_1(z_k) - Q_1(z_{k-1}) > 0$ and $\Delta Q_1(z_k) = 0$ for $k = 1, \ldots, n_{23}+1$. Here we use the term step function to represent functions that can have countably many jumps and are constant between jumps. Let $\mathcal{Q}_{0k}$ be the sets of nonincreasing nonnegative right continuous step functions on $[0, \infty)$ such that they have jumps only at Ψ_k for $k = 2, 3$. Now we let $\mathcal{Q}_0 = \mathcal{Q}_{01} \times \mathcal{Q}_{02} \times \mathcal{Q}_{03}$. In Kim (2007), we proved that $\mathcal{Q}_0$ is a measurable subset of $\mathcal{Q}$. Since α has support on $[0, \infty)$, it is well known from the property of the Dirichlet process that $\mathcal{L}_Q(\mathcal{Q}_0|\mathrm{D}_n) = 1$. Theorem 1 proves that (2.1) has a unique solution on $\mathcal{Q}_0$. From now on, we use $\mathbf{S} = (S_X, S_Y, S_Z)$ and $\mathbf{S} = (S_1, S_2, S_3)$ interchangeably when there is no confusion.

Theorem 1. *For any $\mathbf{Q} \in \mathcal{Q}_0$, (2.1) has a unique solution in $\mathcal{D}_I^3$.*

Proof. First, we show that (2.1) has a solution in $\mathcal{D}_I^3$. For given $\mathbf{Q} \in \mathcal{Q}_0$, suppose that Q_1 has only finitely many jumps at $0 < x_1 < \cdots < x_k < \infty$. We can prove the existence of a solution of (2.1) by modifying the proof of Theorem 6 in Mykland and Ren (1996). See Kim (2007) for details.

For general $\mathbf{Q} \in \mathcal{Q}_0$, we can find a sequence of $\mathbf{Q}_n \in \mathcal{Q}_0$ such that each member has only finitely many jumps, and $\sup_{t\in[0,\infty)} |Q_{nk}(t) - Q_k(t)| \to 0$ as $n \to \infty$ for $k = 1, 2, 3$. For instance, set $Q_{n2} = Q_2$ and $Q_{n3} = Q_3$. As for Q_1, choose t_k in (z_{k-1}, z_k) for $k = 1, \ldots, n_{23} + 1$ such that $\Delta Q_1(t_k) > 0$, and let $A_n = \{t : \Delta Q_1(t) \geq 1/n\} \cup \{t_1, \ldots, t_{n_{23}+1}\}$. Since A_n has only finite number of elements, we write $A_n = \{0 = v_0 < v_1 < v_2 < \cdots < v_l < v_\infty\}$. Let $Q_{n1}(t) = 1 - \sum_{j=1}^{l} w_j I(v_j \leq t)$, where $w_j = Q_1(v_{j-1}) - Q_1(v_j)$ for $j = 1, \ldots, l-1$, and $w_l = Q_1(v_{l-1})$. Then it is easy to show that $\sup_{t\in[0,\infty)} |Q_{n1}(t) - Q_1(t)| \to 0$ as $n \to \infty$.

For given $\mathbf{Q}_n$, let $\mathbf{S}_n$ be a solution of (2.1). Since the $\mathbf{S}_n$ are uniformly bounded and nonincreasing functions, Helly's Selection Theorem implies that there is a subsequence $\mathbf{S}_{n_k}$ such that $\mathbf{S}_{nk}$ converges pointwise to $\mathbf{S} \in \mathcal{D}^3$. Since S_{n_k2} and S_{nk3} have jumps only at Ψ_2 and Ψ_3, respectively, they converge to S_2 and S_3 uniformly. Since Q_{n_k1} converges uniformly to Q_1, the first equation of (2.1) implies that S_{n_k1} converges to S_1 uniformly. Hence, S_1 is a solution of the first equation of (2.2) and $\mathbf{S}$ is a solution of (2.1).

For proving the uniqueness of the solution, we can use Theorem 3.2 of Chang and Yang (1987). Who assumed that S_2 and S_3 are continuous, the proof of their theorem 3.2 can be modified for our problem, see Kim (2007) for details.

Finally, in Kim (2007), we proved that the mapping from $(\mathcal{Q}, \mathcal{B}_{\mathcal{Q}})$ to $(\mathcal{D}_I^3, \mathcal{B}_{\mathcal{D}_I^3})$ induced by (2.1) is measurable.

3.2. Equivalence of the two posteriors

In this subsection, we show that the posterior distributions of S_X from the direct and indirect approaches are the same, i.e., $\mathcal{L}_I(S_X|\mathrm{D}_n) = \mathcal{L}_D(S_X|\mathrm{D}_n)$. Recall that $\beta_1 = \alpha$, and that β_2 and β_3 are null measures.

For our purpose, it suffices to show that, for any $0 < t_1 < \cdots < t_k < \infty$,

$$\mathcal{L}_D(S_X(t_1), S_X(t_2), \ldots, S_X(t_k)|\mathrm{D}_n) = \mathcal{L}_I(S_X(t_1), S_X(t_2), \ldots, S_X(t_k)|\mathrm{D}_n). \quad (3.2)$$

Let $\Psi_1 = \{t_1, \ldots, t_k\} \cup \Psi_2 \cup \Psi_3 \cup \{0, \infty\} = \{0 = x_0 < \cdots < x_{m+1} = \infty\}$. Let $F_i = S_X(x_{i-1}) - S_X(x_i)$ for $i = 1, \ldots, m+1$. Then, (3.2) will hold if

$$\mathcal{L}_D(\mathbf{F}|\mathrm{D}_n) = \mathcal{L}_I(\mathbf{F}|\mathrm{D}_n),$$

where $\mathbf{F} = (F_1, \ldots, F_{m+1})$.

First consider $\mathcal{L}_D(\mathbf{F}|\mathrm{D}_n)$. Let $\mathrm{D}_{n1} = \{(W_i, \delta_i) : \delta_i = 1\}$. Let $\theta_i = \alpha((x_{i-1}, x_i]) + \sum_{k=1}^n I(W_k \in (x_{i-1}, x_i], \delta_i = 1)$, $i = 1, \ldots, m$, $\theta_{m+1} = \alpha((x_m, x_{m+1})) + \sum_{k=1}^n I(W_k \in (x_i, x_{m+1}), \delta_i = 1)$, $\phi_i = \sum_{k=1}^n I(W_k = y_i, \delta_k = 2)$, $i = 1, \ldots, n_2$, and $\psi_i = \sum_{k=1}^n I(W_k = z_i, \delta_k = 3)$, $i = 1, \ldots, n_3$. Then we can write

$$\begin{aligned}\mathcal{L}_D(\mathbf{F}|\mathrm{D}_n) &\propto \prod_{i=1}^{n_2}\Big(1 - \sum_{k:x_k>y_i} F_k\Big)^{\phi_i} \prod_{i=1}^{n_3}\Big(\sum_{k:x_k\le z_i} F_k\Big)^{\psi_i} \times \mathcal{L}_D(\mathbf{F}|\mathrm{D}_{\mathbf{1}}) \\ &\propto \prod_{i=1}^{n_2}\Big(1 - \sum_{k:x_k>y_i} F_k\Big)^{\phi_i} \prod_{i=1}^{n_3}\Big(\sum_{k:x_k\le z_i} F_k\Big)^{\psi_i} \prod_{i=1}^{m+1} F_i^{\theta_i - 1} \\ &\quad \times I\Big(\sum_{i=1}^{m+1} F_i = 1\Big), \end{aligned} \quad (3.3)$$

because $\mathcal{L}_D(F_X|\mathrm{D}_{n1})$ is the Dirichlet process with the base measure $\alpha(\cdot) + \sum_{k=1}^n I(W_i \in \cdot, \delta_i = 1)$. The next theorem proves that $\mathcal{L}_I(\mathbf{F}|\mathrm{D}_n)$ is also proportional to (3.3).

Theorem 2. *Suppose* (3.1) *holds. Then*

$$\mathcal{L}_I(\mathbf{F}|\mathrm{D}_n) \propto \prod_{i=1}^{n_2}\Big(1 - \sum_{k:x_k>y_i} F_k\Big)^{\phi_i} \prod_{i=1}^{n_3}\Big(\sum_{k:x_k\le z_i} F_k\Big)^{\psi_i} \prod_{i=1}^{m+1} F_i^{\theta_i - 1} I\Big(\sum_{i=1}^{m+1} F_i = 1\Big).$$

Proof. Let $d_i = Q_1(x_{i-1}) - Q_1(x_i)$, $i = 1, \ldots, m+1$, $e_i = -\Delta Q_2(y_i)$, $i = 1, \ldots, n_2$, and $f_i = -\Delta Q_3(z_i)$, $i = 1, \ldots, n_3$. Then we have $\mathcal{L}_I(\mathbf{d}, \mathbf{e}, \mathbf{f}|\mathrm{D}_n) \sim$ *Dirichlet*$(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\psi})$, where $\mathbf{d} = (d_1, \ldots, d_{m+1})$, $\mathbf{e} = (e_1, \ldots, e_{n_2})$,$\mathbf{f} = (f_1, \ldots, f_{n_3})$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{m+1})$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{n_2})$, and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{n_3})$.

Let $a_i = S_X(x_{i-1}) - S_X(x_i)$, $i = 1, \ldots, m+1$, $b_i = -\Delta S_Y(y_i)$, $i = 1, \ldots, n_2$, and $c_i = -\Delta S_Z(z_i)$, $i = 1, \ldots, n_3$. Also, let $b_{n_2+1} = S_Y(\infty)$, $y_{n_2+1} = \infty$, and $c_0 = 1 - S_Z(0)$, $z_0 = 0$. Then (2.1) and (3.1) imply that

$$
\begin{aligned}
d_i &= \Big\{ \sum_{k:z_k > x_i} c_k - \sum_{k:y_k > x_i} b_k \Big\} a_i \quad i = 1, \ldots, m, \\
e_i &= \Big(1 - \sum_{k:x_k \le y_i} a_k \Big) b_i \quad i = 1, \ldots, n_2, \qquad (3.4) \\
f_i &= \Big(\sum_{k:x_k \le z_i} a_k \Big) c_i \quad i = 1, \ldots, n_3,
\end{aligned}
$$

with $a_{m+1} = 1 - \sum_{i=1}^{m} a_i$, $b_{n_2+1} = 1 - \sum_{i=1}^{n_2} b_i$ and $c_0 = 1 - \sum_{i=1}^{n_3} c_i$. The variable transformation technique yields that

$$
\begin{aligned}
\mathcal{L}_I(\mathbf{a}, \mathbf{b}, \mathbf{c}|\mathrm{D}_n) \propto & \prod_{i=1}^{m+1} a_i^{\theta_i - 1} \prod_{i=1}^{n_2} \Big(1 - \sum_{k:x_k \le y_i} a_k \Big)^{\phi_i - 1} \prod_{i=1}^{n_3} \Big(\sum_{k:x_k \le z_i} a_k \Big)^{\psi_i - 1} \\
& \times \prod_{i=1}^{m+1} \Big\{ \sum_{k:z_k > x_i} c_k - \sum_{k:y_k > x_i} b_k \Big\}^{\theta_i - 1} \prod_{i=1}^{n_2} b_i^{\phi_i - 1} \prod_{i=1}^{n_3} c_i^{\psi_i - 1} \\
& \times I\Big(\sum_{i=1}^{m+1} a_i = 1 \Big) |J|,
\end{aligned}
$$

where $\mathbf{a} = (a_1, \ldots, a_{m+1})$, $\mathbf{b} = (b_1, \ldots, b_{n_2})$, $\mathbf{c} = (c_1, \ldots, c_{n_3})$, and $|J|$ is the Jacobian. Note that $\mathbf{F} = \mathbf{a}$. By Lemma 1 in the Appendix,

$$
|J| = \prod_{i=1}^{n_2} \Big(1 - \sum_{k:x_k \le y_i} a_k \Big) \prod_{i=1}^{n_3} \Big(\sum_{k:x_k \le z_i} a_k \Big) \prod_{i=1}^{m} \Big\{ \sum_{k:z_k < x_i} c_k - \sum_{k:y_k < x_i} b_k \Big\}.
$$

Hence we conclude that $\mathbf{a}$ and $(\mathbf{b}, \mathbf{c})$ are independent and the distribution of $\mathcal{L}_I(\mathbf{a}|\mathrm{D}_n)$ is the same as that of $\mathcal{L}_D(\mathbf{F}|\mathrm{D}_n)$, as required.

4. The Main Theorem

In this section, we prove the Bernstein-von Mises theorem for the posterior distribution of S_X; that is, we show that the posterior distribution of S_X centered by the NPMLE and scaled by $\sqrt{n}$ is asymptotically equivalent to the asymptotic sampling distribution of the NPMLE.

Throughout the remainder of the paper, the following conditions are assumed to hold.

A1. The random variables X_i and (Y_i, Z_i) are independent for $i = 1, \ldots, n$, with true survival functions S_X^0 and (S_Y^0, S_Z^0), respectively.

A2. $\Pr(Z \leq Y) = 1$.

A3. $S_Y^0(t) - S_Z^0(t) > 0$ on $(0, \infty)$.

A4. S_X^0, S_Y^0 and S_Z^0 are continuous functions of t, on $t \geq 0$, and $0 < S_X^0(t) < 1$ for $t > 0$.

A5. $S_X^0(0) = S_Y^0(0) = 1$, $S_X^0(\infty) = S_Y^0(\infty) = S_Z^0(\infty) = 0$.

A6. There exist δ and T, $0 < \delta < T < \infty$, such that $S_Z^0(t) = \text{constant} < 1$ on $[0, \delta]$ and $S_Z^0(T) = 0$, i.e., $\Pr(Z = 0) > 0$, $\Pr(Z \in (0, \delta)) = 0$ and $\Pr(Z \leq T) = 1$.

These assumptions have been made by Chang (1990) to estabilish weak convergence of the SCE. Milder conditions for the weak convergence of the NPMLE, introduced by Gu and Zhang (1993), could also be used for our purpose; we choose the former for simplicity.

Let S_X^n be the NPMLE of S_X for given D_n. Under A1 to A6, Chang (1990) proved that

$$\sqrt{n}(S_X^n - S_X^0) \xrightarrow{d} W$$

for some Gaussian process W on $\mathcal{D}[0, T]$. The specifies of W can be found in Chang (1990). A Bernstein-von Mises theorem for doubly censored data is given now.

Theorem 3. *Suppose that* D_n *consists of i.i.d. samples of* (W_i, δ_i) *from* (S_X^0, S_Y^0, S_Z^0). *Then*

$$\mathcal{L}_D(\sqrt{n}(S_X - S_X^n)|\mathrm{D}_n) \xrightarrow{d} W$$

on $\mathcal{D}[0, T]$ *in probability with respect to* P^n*, where* P^n *is the probability measure of* D_n.

Before proving the main theorem, we clarify the definition of weak convergence in probability. Let $\mathcal{X}$ be a Polish space (complete separable metric space), let P_n be random probability measures on $\mathcal{X}$, and P be a probability measure on $\mathcal{X}$. By $P_n \xrightarrow{d} P$ in probability, we mean that for any bounded continuous function f on $\mathcal{X}$, $\int f dP_n$ converges to $\int f dP$ in probability. Since the space of probability measures on a Polish space is metrizable (see, for example, Stroock and Varadhan (1979)), we can define weak convergence in probability by asking that $d_w(P_n, P)$ converge to 0 in probability, where d_w is a metric on the space of probability measures induced by the weak convergence. Since the space of bounded continuous functions on a Polish space is separable, the two definitions are equivalent.

Proof of Theorem 3. Let $\mathbf{Q}^0 = (Q_1^0, Q_2^0, Q_3^0)$ be the true sampling distribution of (W_i, δ_i), and $\mathbf{Q}^n$ be the empirical version of $\mathbf{Q}^0$; that is, $Q_k^n(t) = \sum_{i=1}^n I(W_i > t, \delta_i = k)/n$ for $k = 1, 2, 3$. It is well known that $\sqrt{n}(\mathbf{Q}^n - \mathbf{Q}^0) \xrightarrow{d} \mathbf{B}$, where $\mathbf{B}$ is the corresponding Browninan bridge on $\mathcal{D}^3[0, T]$. For given $\mathbf{Q}^n$, let $\mathcal{L}_I(\cdot|\mathbf{Q}^n)$ be a probability measure on $\mathcal{D}^3$ induced by the Dirichlet process with base measure $\beta^p(\cdot) = \beta(\cdot) + n\mathbf{Q}^n(\cdot)$. It suffices to show that

$$\mathcal{L}_I(\sqrt{n}(S_X - S_X^n)|\mathbf{Q}^n) \xrightarrow{d} W$$

in probability with respect to P^n.

Let $u^n = \sqrt{n}(\mathbf{S} - \mathbf{S}^n)'$, $q^n = \sqrt{n}(\mathbf{Q} - \mathbf{Q}^n)'$,

$$\begin{aligned}
\theta^n &= \frac{1}{\sqrt{n}}\bigg(-\int_0^t \frac{u_2^n - u_3^n}{S_Y^n - S_Z^n} du_1^n, -\int_0^t \frac{u_1^n}{S_X^n} du_2^n, -\int_t^T \frac{u_1^n}{1 - S_X^n} du_3^n\bigg)', \\
\alpha^n &= \bigg(-\int_0^t \frac{dq_1^n}{S_Y^n - S_Z^n}, -\int_0^t \frac{dq_2^n}{S_X^n}, \int_t^T \frac{dq_3^n}{1 - S_X^n}\bigg)', \\
\mu^n(ds) &= -\mathrm{diag}(dS_X^n(s), dS_Y^n(s), dS_Z^n(s))
\end{aligned}$$

and $k^n(t, s)$ be a 3×3 matrix with elements $k_{11}^n = k_{22}^n = k_{23}^n = k_{32}^n = k_{33}^n = 0$, $k_{13}^n = -k_{12}^n$, And

$$\begin{aligned}
k_{12}^n(t, s) &= \frac{I(0 < s < t)}{S_Y^n(s) - S_Z^n(s)}, \\
k_{21}^n(t, s) &= \frac{I(0 < s < t)}{S_X^n(s)}, \\
k_{31}^n(t, s) &= \frac{I(t < s < T)}{1 - S_X^n(s)}.
\end{aligned}$$

Then, as at (12) in Chang (1990), we have

$$(I - K_n)u^n = \alpha^n + \theta^n, \tag{4.1}$$

where I is the identity operator, and the operator K_n is defined as

$$K_n u = \int_0^T \mu^n(ds) k^n(\cdot, s) u(s).$$

Let K be the operator defined similarly to K_n but with $\mathbf{S}^n$ replaced by $\mathbf{S}^0$. Then, (4.1) can be rewritten as

$$(I - K)u^n = \alpha^n + \theta^n + (K_n - K)u^n. \tag{4.2}$$

Chang (1990) proved that there exists a resolvent kernel matrix Γ such that each element is a bounded measurable functions on $[0,T]\times[0,T]$, and $u^n = (I+\Gamma)(\alpha^n+\theta^n+(K_n-K)u^n)$, where

$$\Gamma a = \int_0^T \mu(ds)\Gamma(\cdot,s)a(s)$$

and $\mu(s)$ is a matrix similar to μ^n but with $\mathbf{S}^n$ replaced by $\mathbf{S}^0$.

By a slight modification of Lo (1987), we have that

$$\mathcal{L}_I(\sqrt{n}(\mathbf{Q}-\mathbf{Q}^n)|\mathbf{Q}^n) \xrightarrow{d} \mathbf{B} \tag{4.3}$$

on $\mathcal{D}^3[0,T]$ with probability 1. Since $\mathbf{S}^n \to \mathbf{S}^0$ uniformly on $[0,T]^3$ with probability 1 (see Theorem 4.2 of Chang and Yang (1987)), (4.3) implies that $\mathcal{L}_I(\alpha^n|\mathbf{Q}^n)$ converges weakly to a Gaussian process with probability 1, the weak limit of the sampling distribution of α_0^n, where α_0^n is defined as was α^n, except that $\mathbf{Q}$ and $\mathbf{Q}^n$ are replaced by $\mathbf{Q}^n$ and $\mathbf{Q}^0$, respectively. Since the sampling distribution of the first component of $(I+\Gamma)\alpha_0^n$ converges weakly to W on $\mathcal{D}[0,T]$ (Chang (1990)), the proof would be complete if we could show that for any $\epsilon>0$, $\mathcal{L}_I(|\theta^n|>\epsilon|\mathbf{Q}^n)\to 0$ and $\mathcal{L}_I(|(K-K_n)u^n|>\epsilon|\mathbf{Q}^n)\to 0$ in probability with respect to P^n.

By the Skorohod Representation Theorem (Pollard (1987)), without loss of generality we can assume that

$$\sqrt{n}(\mathbf{Q}^n-\mathbf{Q}^0)\to\mathbf{B} \tag{4.4}$$

on $\mathcal{D}[0,T]^3$ with Probability 1. From now on, we assume that a sequence of $\{\mathbf{Q}^n\}$, for which $\sqrt{n}(\mathbf{Q}^n-\mathbf{Q}^0)$ converges uniformly to a continuous function on $[0,T]^3$ and (4.3) holds, is given. By the application of the Skorohod Representation Theorem, we can assume that there exist a sequence of random functions $\{\mathbf{Q}^{n*}\}$ such that $\sqrt{n}(\mathbf{Q}^{n*}-\mathbf{Q}^n)$ converges to $\mathbf{B}$ with Probability 1, and $\mathbf{Q}^{n*}\sim\mathcal{L}_I(\cdot|\mathbf{Q}^n)$. Then, it suffices to show that $(K-K_n)u^n$ and θ^n obtained from $\mathbf{Q}^{n*}$, instead of from $\mathbf{Q}$, converge to 0 with probability 1. Note that $\mathbf{Q}^{n*}-\mathbf{Q}^n$ converges to 0 uniformly on $[0,T]$ with Probability 1, and so that $\mathbf{S}^{n*}-\mathbf{S}^n$ also converges to 0 uniformly on $[0,T]$ with Probability 1.

First, consider $(K-K_n)u^n=(Z_1^n,Z_2^n,Z_3^n)'$, say. We prove that Z_1^n converges uniformly to 0 on $[0,T]$. The convergence of Z_2^n and Z_3^n can be proved similarly. We can write

$$\begin{aligned} Z_1^n(t) &= \int_0^t (u_2^n-u_3^n)\left(\frac{dS_X^n}{S_Y^n-S_Z^n}-\frac{dS_X^0}{S_Y^0-S_Z^0}\right)\\ &= \int_0^t \Big[(S_Y^{n*}-S_Y^{n*})-(S_Z^{n*}-S_Z^n)\Big]\sqrt{n}\left(\frac{1}{S_Y^n-S_Z^n}-\frac{1}{S_Y^0-S_Z^0}\right)dS_X^n \qquad (4.5)\\ &\quad+\int_0^t (u_2^n-u_3^n)\frac{d(S_X^n-S_X^0)}{S_Y^0-S_Z^0}. \qquad (4.6)\end{aligned}$$

Since $\sqrt{n}(\mathbf{S}^n - \mathbf{S}^0)$ converges to a continuous function, so does

$$\sqrt{n}\left(\frac{1}{S_Y^n - S_Z^n} - \frac{1}{S_Y^0 - S_Z^0}\right).$$

Since $(S_Y^{n*} - S_Y^n) - (S_Z^{n*} - S_Z^n)$ converges to 0 uniformly with Probability 1 and S_X^n converges to S_X^0 uniformly, we conclude that (4.5) converges to 0 uniformly with Probability 1. For (4.6), integration by part yields

$$\begin{aligned}(4.6) = &\frac{(S_Y^{n*}(t) - S_Y^n(t)) - (S_Z^{n*}(t) - S_Z^n(t))}{S_Y^0(t) - S_Z^0(t)}\sqrt{n}(S_X^n(t) - S_X^0(t)) \\ &+ \int_0^t \frac{(S_Y^{n*}(t) - S_Y^n(t)) - (S_Z^{n*}(t) - S_Z^n(t))}{(S_Y^0(t) - S_Z^0(t))^2}\sqrt{n}(S_X^n - S_X^0)d(S_Y^0 - S_Z^0) \\ &- \int_0^t \frac{\sqrt{n}(S_X^n - S_X^0)}{S_Y^0 - S_Z^0}\Big[d(S_Y^{n*} - S_Y^n) - d(S_Z^{n*} - S_Z^n)\Big].\end{aligned}$$

The first term on the right side of the above equation clearly converges uniformly to 0, and the second term converges uniformly to 0 because the integrand does. The third term also converges uniformly to 0 since the integrand converges uniformly to a continuous function and the integrator converges to 0 uniformly with Probability 1.

The uniform convergence of θ^n to 0 in probability can be proved similarly as was Lemma 3.3 of Chang (1990) with $\mathbf{Q}^{n*}$, and so the proof is done.

5. Simulation

In this section, we present simulation results to evaluate the true coverage probability of the Bayesian probability interval. Survival times X were generated from $Exp(100)$ — the exponential distribution with mean 100. The left and right censoring variables (Z, Y) were generated by $(Z, Y) = (Z, Z + W)$, where $Z \sim Exp(10)$ and $W \sim Exp(140)$ and Z and W are independent. Under this model, the censoring probability is about 48%, of which 38% is due to right censoring and 10% due to left censoring. For each of four sample sizes $n = 20, 50, 100$ and 200, we generated 1,000 data sets, and calculated the empirical coverage probabilities of the Bayesian probability interval of F at times $t = 50, 100, 150, 200$. The empirical coverage probability is the proportion of the data sets having the probability intervals including the true parameter value. The posteriors are calculated by the MCMC algorithm of Doss (1994) with 10,000 iterations, of which the first 1,000 iterations are discarded as burn-in. For the base measure of the Dirichlet process prior, we set $\alpha[0, t] = 1 - \exp(-t)$.

Simulation results are presented in Figure 5.1. The thee solid lines represent the nominal coverage probability 0.9 and two standard errors, $2\sqrt{0.9 \cdot 0.1/1,000} =$

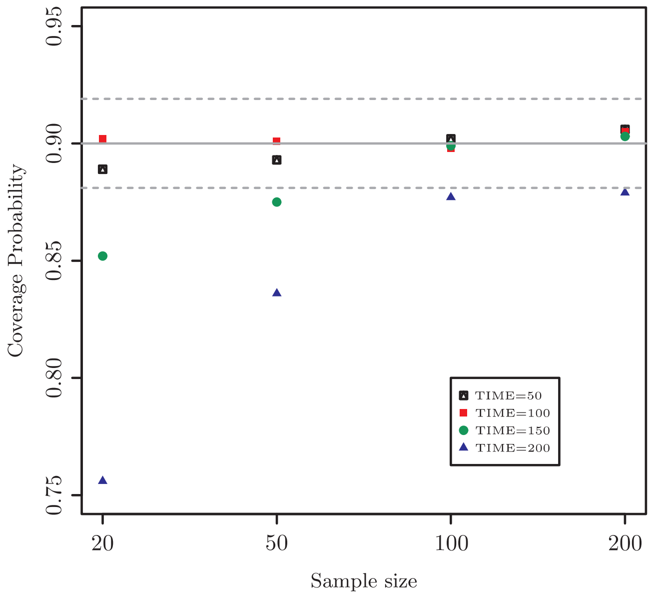

Figure 5.1. Empirical coverage probabilities of the Bayesian credible sets for $S_X(t)$ at $t = 50, 100, 150, 200$ with nominal level 90%. The three solid lines represent the nominal level and two standard errors from it. The dots are the empirical coverage probabilities.

0.0190, away from it. For $t = 100$ (the mean survival time), the coverage probability of the probability interval is very close to the nominal level when the sample size is small (i.e., $n = 20$). In contrast, for $t = 200$, the coverage probability is not close to the nominal level when $n = 200$. Based on these results, we conclude that the posterior distribution is a good approximation of the sampling distribution of the MLE for most time points, unless the sample size is too small and a time point is too large.

Acknowledgement

This research was supported (in part) by the SRC/ERC program of MOST/ KOSEF (R11-2000-073-00000) and grant R01-2004-000-10284-0 from the Basic Research Program of the KOSEF.

A. Appendix

Lemma 1.

$$|J| = \prod_{i=1}^{n_2}\Big(1 - \sum_{k:x_k\le y_i} a_k\Big)\prod_{i=1}^{n_3}\Big(\sum_{k:x_k\le z_i} a_k\Big)\prod_{i=1}^{m}\Big\{\sum_{k:z_k<x_i} c_k - \sum_{k:y_k x_i} b_k\Big\}.$$

Proof. Write

$$J = \begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix}$$

where

$$J_{11} = \begin{bmatrix} \frac{dd_1}{da_1} & \cdots & \frac{dd_1}{da_m} \\ \vdots & \ddots & \vdots \\ \frac{dd_m}{da_1} & \cdots & \frac{dd_m}{da_m} \end{bmatrix},$$

$$J_{22} = \begin{bmatrix} \frac{de_1}{db_1} & \cdots & \frac{de_1}{db_{n_2}} \\ \vdots & \ddots & \vdots \\ \frac{de_{n_2}}{db_1} & \cdots & \frac{de_{n_2}}{db_{n_2}} \end{bmatrix},$$

$$J_{33} = \begin{bmatrix} \frac{df_1}{dc_1} & \cdots & \frac{df_1}{dc_{n_3}} \\ \vdots & \ddots & \vdots \\ \frac{df_{n_3}}{dc_1} & \cdots & \frac{df_{n_3}}{dc_{n_3}} \end{bmatrix},$$

and J_{ij}, $i \neq j$ are defined accordingly. Note that $J_{ii}, i = 1, 2, 3$ are diagonal matrices.

Let g_{ij} be the (i, j)th element of J, and let $N = m + n_2 + n_3$. Let $I_1 = \{1, \ldots, m\}$, $I_2 = \{m + 1, \ldots, m + n_2\}$, and $I_3 = \{m + n_2 + 1, \ldots, N\}$. Let Π be the set of all permutations of $\{1, \ldots, N\}$. Now

$$|J| = \sum_{\pi \in \Pi} (\pm) g_{1\pi(1)} g_{2\pi(2)} \cdots g_{N\pi(N)},$$

where $(\pm)$ is either $+1$ or -1 depending on the permutation π.

Note that

$$\prod_{i=1}^{N} g_{ii} = \prod_{i=1}^{n_2} \Big(1 - \sum_{k: x_k \le y_i} a_k\Big) \prod_{i=1}^{n_3} \Big(\sum_{k: x_k \le z_i} a_k\Big) \prod_{i=1}^{m} \Big\{\sum_{k: z_k < x_i} c_k - \sum_{k: y_k < x_i} b_k\Big\}.$$

We prove that for any permutation $\pi \in \Pi$, $\prod_{i=1}^{N} g_{i\pi(i)} = 0$ unless $\pi(i) = i$ for all $i = 1, \ldots, N$.

For a given permutation π, suppose $\prod_{i=1}^{N} g_{i\pi(i)} \neq 0$. Let $B = \{i : \pi(i) \neq i\}$. And k be the smallest element in $B \cap I_1$. Then, in order that $\prod_{i=1}^{N} g_{i\pi(i)} \neq 0$, $\pi(k)$ should be either in I_2 or I_3 because J_{11} is a diagonal matrix. Suppose further that $\pi(k) \in I_2$. Then, direct calculation with the first equation of (3.4) yields that $g_{k,\pi(k)} \neq 0$ only if

$$x_k > y_{\pi(k)-m}. \tag{A.1}$$

Next, note that $\pi(k) \in B$. Consider $\pi(\pi(k))$. Since J_{23} and J_{32} are zero matrices and J_{22} is a diagonal matrix, $\pi(\pi(k))$ should be in I_1. However, $g_{\pi(k),\pi(\pi(k))} \neq 0$ only if

$$x_{\pi(\pi(k))} \leq y_{\pi(k)-m} \tag{A.2}$$

from the second equation of (3.4). Hence, from (A.1) and (A.2), we conclude that $\pi(\pi(k)) < k$, which is impossible since $\pi(\pi(k)) \in B \cap I_1$ and k is assumed to be the smallest element in $B \cap I_1$. Similarly, we can prove that it is impossible that $\pi(k) \in I_3$. Hence, we conclude that $I_1 \cap B = \emptyset$.

Next, suppose that $k \in I_2$. Then $g_{k,\pi(k)} = 0$ unless $\pi(k) \in I_1$. However, $\pi(k) \notin I_1$ since $\pi(k)$ is also in B. Hence, $I_2 \cap B = \emptyset$. Similarly, we can show that $I_3 \cap B = \emptyset$. Therefore, we conclude that B is the empty set and the proof is complete.

References

Cai, T and Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika* **91**, 277-290.

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **81**, 391-404.

Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15**, 1536-1547.

Conti, P. L. (1999). Large sample Bayesian analysis for Geo/G/1 discrete-time queueing models. *Ann. Statist.* **27**, 1785-1807.

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

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

Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27**, 1119-1140.

Gehan, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52**, 650-653.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparanetrics.* Springer, New York.

Gu, M. G. and Zhang, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, **21**, 611-624.

Kim, Y. (2006). The Bernstein-von Mises theorem for the proportional hazard model. *Ann. Statist.* **34**, 1678-1700.

Kim, Y. (2007). Supplements for A Bernstein-von Mises theorem in the non-parametric right-censoring model. Preprint.

Kim, Y. and Lee, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29**, 666-686.

Kim, Y. and Lee, J. (2004). A Bernstein-von Mises theorem in the non-parametric right-censoring model. *Ann. Statist.*, **32**, 1492-1512.

Lo, A. Y. (1983). Weak convergence for Dirichlet processes. *Sankhyā* **45**, 105-111.

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

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observations. *Biometrics* **23**, 65-78.

Mykland, P. A. and Ren, J. J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data . *Ann. Statist.* **24**, 1740-1764.

Pollard, D. (1987). *Convergence of Stochastic Processes.* Springer, New York.

Shen, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97**, 222-235.

Stroock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Prcesses.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Berlin.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69**, 169-173.

Department of Statistics, Seoul National University, Seoul, Korea.

E-mail: ydkim@stats.snu.ac.kr

(Received April 2007; accepted January 2008)