The Annals of Statistics
1993, Vol. 21, No. 1, 100–123

A BAYESIAN BOOTSTRAP FOR CENSORED DATA[1]

BY ALBERT Y. LO

SUNY at Buffalo

A Bayesian bootstrap for a censored data model is introduced. Its small sample distributional properties are discussed and found to be similar to Efron's bootstrap for censored data. In the absence of censoring, the Bayesian bootstrap for censored data reduces to Rubin's Bayesian bootstrap for complete data. A first-order large-sample theory is developed. This theory shows that both censored data bootstraps are consistent bootstraps for approximating the sampling distribution of the Kaplan–Meier estimator. It also shows that both bootstraps are consistent bootstraps for approximating a posterior distribution of the survival function with respect to each member of the class of conjugate beta-neutral process priors.

1. Introduction. The bootstrap method for censored data (CDB) was suggested by Efron (1981). The CDB is derived from a frequentist viewpoint and, in the absence of censoring, the CDB reduces to Efron's (1979) complete data bootstrap. Akritas (1986) showed that Efron's CDB is consistent in the sense that the conditional limiting distribution of the bootstrapped Kaplan–Meier function is identical to the limiting distribution of the Kaplan–Meier estimator [Breslow and Crowley (1974) and Gill (1983)]. Reid (1981) discussed another resampling method for censored data; Akritas (1986) showed that Reid's method and Efron's CDB are not asymptotically equivalent.

This paper introduces a Bayesian analogue of Efron's CDB, called the Bayesian bootstrap for censored data (CDBB). The CDBB is defined by replacing the 1's in the Kaplan–Meier estimator by standard i.i.d. exponential random variables (Section 2). This definition of a CDBB is analogous to a definition of Rubin's (1981) Bayesian bootstrap, in which the 1's in the empirical distribution function are replaced by standard i.i.d. exponential random variables [see, e.g., Weng (1989)]. Furthermore, in the absence of censoring, the CDBB reduces to Rubin's (1981) Bayesian bootstrap for complete data. Section 3 discusses the small-sample similarities between the CDB and the CDBB.

Section 4 is concerned with the case of categorical censored data. It is shown that the CDBB distribution is a posterior distribution with respect to a "flat" prior. It is also shown that the CDBB is a consistent bootstrap approximation to posterior distributions with respect to smooth prior densities.

Received November 1990; revised February 1992.

[1]Partially supported by NSF Grant DMS-89-06382.

AMS 1991 *subject classifications*. Primary 62G09; secondary 62G20, 62G99.

Key words and phrases. Censored data, Bayesian bootstrap methods, beta-neutral prior and posteriors distributions, Kaplan–Meier estimator, Bayesian Kaplan–Meier function.

Section 5 shows that the CDBB and Efron's CDB are asymptotically equivalent. In both cases, the limiting conditional distribution of the bootstrapped survival function is identical to that of the Kaplan–Meier estimator obtained by Breslow and Crowley (1974) and Gill (1983).

Section 6 shows that the CDBB is consistent in approximating a posterior distribution of the survival function for Bayesian nonparametric problems. A solution to this problem requires a Bayesian version of Breslow and Crowley's (1974) functional central limit theorem. A class of conjugate neutral process priors [Doksum (1974) and Ferguson and Phadia (1979)], called the beta-neutral process prior [Hjort (1990)], is discussed. A path-wise definition of a beta-neutral process prior using two gamma processes is introduced. Posterior distributions of a survival function with respect to beta-neutral survival process priors are shown to obey the Breslow–Crowley limit (Corollary 6.1).

Section 7 discusses a Markov chain property of the survival function, the censored data and the censoring distributions.

REMARK 1.1. A referee pointed out that recently Hjort (1991) discussed results for the cumulative hazard similar to Lemma 5.1 and Corollary 6.2, as well as other interesting nonparametric bootstrap methods for censored data.

2. Censored data and the bootstraps. Suppose $T_1, \dots, T_n$ are i.i.d. survival times which are censored on the right by n follow-up times, $C_1, \dots, C_n$. The survival function of T_1 is $S(t) = 1 - F(t)$. The C_i's are independent, and C_i has a distribution function $G_i(c)$. We say that the observations $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$ are sample data from a random censoring model (with parameters F and G_i's) if

$$Y_i = \min\{T_i, C_i\},\qquad \delta_i = 1 \text{ if } T_i \le C_i \text{ and } \delta_i = 0 \text{ if } T_i > C_i. \tag{2.1}$$

Suppose we observe $(Y_i, \delta_i) = (y_i, \delta_i)$, $i = 1, \dots, n$ [denoted concisely by $(Y, \delta) = (y, \delta)$] from a random censoring model. Let $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ be the distinct ordered values for times to death (uncensored data). For $j = 1, \dots, k$, $D(j) = \{i\colon y_i = t(j),\ \delta_i = 1\}$ and $R(j) = \{i\colon y_i$ is "alive" just before time $t(j)\}$. The number of elements of the sets $D(j)$ and $R(j)$ are denoted by d_j and r_j, respectively.

The Kaplan–Meier estimator of $S(t) = 1 - F(t)$ based on the (y, δ) is defined by

$$\hat{S}(t) = \prod_{j:\, t(j) \le t} \left(1 - \sum_{q \in D(j)} 1 \Big/ \sum_{q \in R(j)} 1\right). \tag{2.2}$$

Suppose the sampling distributional quantities of $\theta(\hat{S}, S)$, given S, is of interest. Efron's (1981) CDB algorithm produces an approximation to the sampling distributional quantities of $\theta(\hat{S}, S)$ (given S), as follows: One puts the censored data $(y_1, \delta_1), \dots, (y_n, \delta_n)$ in an urn, and then takes an i.i.d. sample from the urn to obtain a bootstrap sample $(Y_1^*, \delta_1^*), \dots, (Y_n^*, \delta_n^*)$.

Construct a Kaplan–Meier estimator $K^*(t)$ based on $(Y_1^*, \delta_1^*), \dots, (Y_n^*, \delta_n^*)$ and evaluate $\theta^* = \theta(K^*, \hat{S})$. Repeat this process B times to get $\theta_1^*, \dots, \theta_B^*$, and use the empirical distribution function of $\theta_1^*, \dots, \theta_B^*$ to approximate the (sampling) distribution $\mathscr{L}\{\theta(\hat{S}, S)|S\}$, the sample variance of $\theta_1^*, \dots, \theta_B^*$ to approximate $\mathrm{Var}\{\theta(\hat{S}, S)|S\}$ and so on.

We next turn to the Bayesian bootstrap. For uncensored data, Rubin (1981) proposed the use of the (infinite) Bayesian bootstrap, and compared it with Efron's (1979) bootstrap. Lo (1987, 1988) suggested the use of the Bayesian bootstraps to approximate posterior distributions; Weng (1989) [see also Weng (1988)] showed that the Bayesian bootstrap approximation is superior to the standard normal approximation to a posterior distribution of the unknown population mean with respect to a Dirichlet prior. For the random censoring model (2.1), the primary interest for a Bayesian is the posterior distribution of $\theta(S, \hat{S})$ given (y, δ), and the CDBB provides an approximation to the posterior quantities of $\theta(S, \hat{S})$ given (y, δ).

Rubin's infinite Bayesian bootstrap is based on simulation rather than resampling [see, however, Lo (1988)]; the CDBB is also based on simulation. Let $Z_1, \dots, Z_n$ be i.i.d. standard exponential random variables. Replace the "1's" in the Kaplan–Meier estimator (2.2) by "Z_q's" to obtain

$$S^*(t) = \prod_{j:\, t(j) \le t} \left(1 - \sum_{q \in D(j)} Z_q \Big/ \sum_{q \in R(j)} Z_q\right). \tag{2.3}$$

Given (y, δ), $S^*(t)$ is a random survival function. For a Bayesian, S^* plays the role of Efron's bootstrapped Kaplan–Meier estimator. We call $S^*(t)$ the CDBB survival function. In the absence of censoring, $S^*(t)$ reduces to $1 - D^*(t)$ where

$$D^*(t) = \sum_{1 \le i \le n} [Z_i/(Z_1 + \cdots + Z_n)] I_{\{y_i \le t\}}. \tag{2.4}$$

Note that $Z_i/(Z_1 + \cdots + Z_n)$ is distributed as the gaps of $n - 1$ i.i.d. $U(0, 1)$ random variables. The simulation of these gaps in $D^*(t)$ is the basis of Rubin's (1981) Bayesian bootstrap. Likewise, the simulation of $S^*(t)$ is the backbone of the CDBB method. Next is the CDBB algorithm:

(2.5)

Simulation step: Simulate n i.i.d. standard exponential random variables $Z_1, \dots, Z_n$.

Construction step: Replace the "1's" in the Kaplan–Meier estimator (2.2) by "Z_q's" to obtain $S^*(t)$.

Evaluation step: Evaluate $\theta^* = \theta(S^*, \hat{S})$.

Repeat the previous three steps a large number of times, say B times, to obtain $\theta_1^* = \theta(S_1^*, \hat{S}), \dots, \theta_B^* = \theta_B(S_B^*, \hat{S})$, and use the empirical distribution of $\theta_1^*, \dots, \theta_B^*$ to approximate the posterior distributions $\mathscr{L}\{\theta(S^*, \hat{S})|(y, \delta)\}$ with respect to smooth priors.

Consequently one can use the sample median, the sample average and the sample variance of $\theta_1^*, \dots, \theta_B^*$ to approximate the posterior median of $\theta(S^*, \hat{S})$, the posterior mean of $\theta(S^*, \hat{S})$ and the posterior variance of $\theta(S^*, \hat{S})$, respectively.

The following example illustrates the CDBB algorithm.

EXAMPLE 2.1. The following are times (weeks) of remission (i.e., freedom from symptoms) of leukemia patients [Gehan (1965)]. Some patients are treated with the drug 6-mercaptopurine, the others serving as a control. Treatment allocation was randomized.

The treatment group data is

$$6, 6, 6, 6+, 7, 9+, 10, 10+, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+,$$
$$32+, 32+, 34+, 35+.$$

(Times in "+" are censored data.)

To construct a $(1-\alpha)$-CDBB band for the survival function $S(t)$ based on these data, we choose the functional

$$\theta^* = \theta(S^*, \hat{S}) = n^{1/2} \max_{0 \le t \le y(n)} |S^*(t)/\hat{S}(t) - 1| \hat{\tau}(t), \tag{2.6}$$

where

$$\hat{\tau}(t) = \hat{C}(T)^{1/2} \Big/ \left[\hat{C}(T) + \hat{C}(t)\right],$$

$$\hat{C}(t) = n \sum_{j:\, t(j) \le t} d_j \Big/ \left[(r_j - d_j) r_j\right]$$

and $y(n) = \max\{y_i\}$. The $(1-\alpha)$-CDBB (uniform) band is given by

$$\hat{S}(t) \pm z_{1-\alpha}\left[\hat{S}(t)/\hat{\tau}(t)\right] n^{-1/2},$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-percentile point of the functional $\theta^* = \theta(S^*, \hat{S})$.

The CDBB band obtained will only be uniform between two adjacent uncensored data. Note that this functional (2.6) is an asymptotic pivotal quantity if the "true" survival function is continuous (see Corollary 5.1 in Section 5). That is, the limiting distribution of θ^* is independent of the "true" survival function. The CDBB algorithm (2.5) is executed for the above functional (2.6) based on Gehan's (1965) data. Percentile points of the (finite sample) conditional distribution for the functional (2.6) are given in Table 1 ($B = 1000$).

TABLE 1

	50%	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**
CDBB	0.535	0.77	0.932	1.07	1.21	1.34	1.39
CDB	0.558	0.81	0.98	1.09	1.23	1.339	1.51
$G(4;1)$	0.53	0.80	0.955	1.08	1.22	1.36	1.52

In the third row, percentile points of Efron's CDB distribution are displayed. The CDB results in a slightly wider band.

In the fourth row, we give results when gamma (4; 1) Z_i's are used instead of the standard exponential Z_i's in the CDBB algorithm. In this case the correct functional is $\rho \times \theta^*$, where $\rho = E(Z_1)/\sigma(Z_1)$. (See Remark 2.1.) The band based on the gamma (4; 1) CDBB appears to be a compromise between the CDBB and CDB.

It is also noted that the histograms for the CDBB distributions are smoother than that for the CDB; this point was also noted by Rubin (1981) in the complete data case.

The construction of a CDBB-approximate HPD (highest posterior density) region for the survival function is similar. The posterior distribution of the survival function is an infinite dimensional distribution, and there are technical difficulties in evaluating and in defining the mode of an infinite dimensional distribution. Here, the CDBB approximation is again useful. Note that S^* [conditional on (y, δ)] is a finite dimensional vector. The mode of S^* is $\hat{S}$ [see Ferguson and Phadia (1979), page 180]. A CDBB approximate HPD region for the survival function can be obtained by simulating $\theta^* = \max_{0 \le t \le y(n)} |S^*(t) - \hat{S}(t)|$.

REMARK 2.1. It is natural to ask whether we can use nonexponential random variables Z_i to play the role of the standard exponential Z_i. In the complete data situation, this method, called Bayesian bootstrap clones, has been developed in Lo (1991). This theory is supported by a study of Weng (1989) [see also Weng (1988)], Remark 2.3, who suggests the use of gamma (4, 1) Z_i's instead. A corresponding theory for the censored data model (2.1) based on simulating other independent Z_i's will be developed elsewhere.

REMARK 2.2. This section is concluded with a discussion of a Bayesian bootstrap which has the ability to incorporate prior information. This has been discussed by Lo (1988), Remark 3.1, for a simple random sampling model [Lo (1986)]. In the present case, suppose the prior information is summarized by a set of p prior censored data $\{(x_i, \delta_i), \dots, (x_p, \delta_p)\}$. Combine these prior data with the current data $\{(y_i, \delta_i),\ i = 1, \dots, n\}$ to get an updated data urn. Carry out the Bayesian bootstrap algorithm using this updated urn instead of $\{(y_i, \delta_i),\ i = 1, \dots, n\}$. Here it is necessary to simulate $n + p$ i.i.d. exponential random variables Z_i in each execution of the Bayesian bootstrap algorithm. It is also preferable to simulate gamma $(\alpha_i; 1)$ Z_i corresponding to the prior data x_i (α_i could be any nonnegative positive number). [The Z_i's corresponding to the current data (y_i, δ_i)'s continue to be standard exponential random variables.] These gamma random variables reflect one's prior belief of the prior data. The limiting case of this construction takes us into the domain of Bayesian nonparametric inference; see Section 6. In a typical Bayesian approach, the posterior distribution is the distribution of a stochastic survival process with infinitely many random jumps, which is another reason why a Bayesian bootstrap approximation is useful.

3. Small sample distributional properties of the CDBB. The CDBB survival function $S^*(t)$ is a cumulative product of independent beta $(r_j - d_j; d_j)$ random variables. This property follows from Theorem 1.2.3 in Bickel and Doksum (1977). As a consequence, the conditional mean of $S^*(t)$ is

$$E[S^*(t)|(y,\delta)] = \hat{S}(t), \tag{3.1}$$

which is defined on $[0, \max\{y_i\}]$.

Efron's bootstrapped Kaplan–Meier function is a cumulative random product of rescaled binomial random variables. Using the conditional arguments as in Kaplan and Meier (1958), one sees that

$$E[K^*(t)|(y,\delta)] = \hat{S}(t) \tag{3.2}$$

defined on $[0, \max\{y_i^*\}]$. It is possible that $\max\{y_i^*\}$ is less than $\max\{y_i\}$.

Next, we turn to the conditional variance of $S^*(t)$. Since given (y, δ), $S^*(t)$ is a product of independent beta $(r_j - d_j; d_j)$ random variables, the conditional variance of $S^*(t)$ is given by

$$\begin{aligned} \operatorname{Var}\{S^*(t)|(y,\delta)\} &= E\left[(S^*(t))^2|(y,\delta)\right] - E[S^*(t)|(y,\delta)]^2 \\ &\approx \left[\hat{S}(t)\right]^2 \times \sum_{j:\, t(j)\le t} d_j \Big/ \left[(r_j - d_j)(r_j+1)\right], \end{aligned} \tag{3.3}$$

which is approximately equal to Greenwood's formula [Kaplan and Meier (1958)]

$$\left[\hat{S}(t)\right]^2 \times \sum_{j:\, t(j)\le t} d_j \Big/ \left[(r_j - d_j) r_j\right]. \tag{3.4}$$

[We retain the $r_j + 1$ instead of an r_j in the denominator of (3.3) to show the effect of the posterior variance of a beta random variable.]

Note that Efron (1981) also showed that the variance of his bootstrapped Kaplan–Meier function, given (y, δ), is approximately given by Greenwood's formula (3.4). Since both $S^*(t)$ and $K^*(t)$ have Greenwood's formula as their asymptotic (conditional) variance, it is to be expected that the two are first-order large-sample equivalent. This theory will be developed in the next two sections.

There is an alternative description of the CDBB, employing the gaps of $n - 1$ i.i.d. uniform random variables; this is perhaps more in the spirit of Rubin's (1981) description of the Bayesian bootstrap. Suppose

$$0 = U(0) < U(1) < U(2) < \cdots < U(n-1) < U(n) = 1$$

are the order statistics of $n - 1$ i.i.d. $U(0,1)$ random variables with gaps $\Delta_j = U_j - U_{j-1}$ for $j = 1, \dots, n$. Define $U^*(t)$ based on the data $\{(y_i, \delta_i)$: $i = 1, \dots, n\}$ by

$$\begin{aligned} U^*(t) &= \prod_{i:\, y_i \le t} \{1 - \Delta_i/[\Delta_i + \Delta_{i+1} + \cdots + \Delta_n]\}^{\delta_i} \\ &= \prod_{i:\, U(i)\le t} \{[1 - U(i)]/[1 - U(i-1)]\}^{\delta_i}, \end{aligned} \tag{3.5}$$

where the last equality is in distribution (given the data) equality. Note that $\mathscr{L}\{S^*(\cdot)|\text{data}\} = \mathscr{L}\{U^*(\cdot)|\text{data}\}$. In the case of no censoring ($\delta_i = 1$ for all i), the numerator of a term in the product cancels with the denominator of the next term, and if $U(j) \le t < U(j+1)$, $U^*(t)$ reduces to the numerator of the last term, which is $1 - U(j)$.

REMARK 3.1. The CDBB survival function $S^*(t)$ can be represented alternatively via the CDBB cumulative hazard $\Lambda^*(t)$ as follows:

$$S^*(t) = \prod_{s \le t} [1 - \Delta\Lambda^*(s)], \tag{3.6}$$

where $\Lambda^*(s)$ is a cumulative sum of independent beta $(d_j, r_j - d_j)$ random variables. Hjort (1991) suggested simulating these independent betas as the basis of CDBB. Note that replacing the 1's by i.i.d. exponential Z_q's in (2.3) automatically produces these independent betas. It must also be pointed out that simulating independent betas directly may not be as efficient as simulating i.i.d. exponentials.

4. CDBB in categorical models with smooth priors. Suppose both F and the G_i's are concentrated on a finite set $\{a_1, \dots, a_{b+1}\}$, where $a_1 < a_2 < \cdots < a_{b+1}$ and $p_j = P\{T = a_j\}$, $j = 1, \dots, b+1$ are the cell probabilities; the p_j's are in (0, 1) with $\Sigma_{1 \le j \le b+1} p_j = 1$. For censored data on the line that has an intrinsic ordered structure, it is convenient to reparametrize this $p = (p_1, \dots, p_{b+1})$ by the discrete hazard rates

$$\lambda_j = P\{T = a_j\}/P\{T \ge a_j\}. \tag{4.1}$$

The likelihood function based on the data (y, δ), that is, $\text{Lik}(\lambda|\text{data})$, is proportional to

$$\prod_{1 \le j \le b} (\lambda_j)^{d_j} (1 - \lambda_j)^{r_j - d_j}, \tag{4.2}$$

which is essentially a likelihood function of b independent binomial $(r_j; \lambda_j)$ random variables. (Recall that d_j equals the number of deaths at a_j and r_j equals the number of units at risk just before time a_j.) Assuming an independent beta prior for the λ_j's results in independent beta posteriors; see Cox and Oakes (1984) and Hjort (1990). In particular, if one assigns a Dirichlet prior on the cell probabilities p_j, the map $p \to \lambda = (\lambda_1, \dots, \lambda_b)$ induces an independent beta prior on the λ_j's.

Assume a "flat" prior density $\pi_0'(\lambda)$ (with respect to Lebesgue measure) for the λ_j's, that is, $\pi_0'(\lambda) = 1/\prod_{1 \le j \le b} \lambda_j(1 - \lambda_j)$. [Note that $\pi_0(d\lambda)$ is σ finite.] Then the corresponding posterior distribution of the λ_j's is given by the following:

(i) λ_j's$|(y, \delta)$ are independent.
(ii) $\mathscr{L}\{\lambda_j|(y, \delta), \pi_0\} = \text{beta}\,(d_j; r_j - d_j)$, $j = 1, \dots, b$.

If the distribution of λ_j's is specified by (i) and (ii), $\prod_{j:\, t(j) \le t}(1 - \lambda_j)$ has the same distribution as the CDBB survival function $S^*(t)$. In this sense, the CDBB is a Bayesian method based on a "flat" prior. Note that if the λ_j's are independent, $S(t) = \prod_{j:\, t(j) \le t}(1 - \lambda_j)$ is a "beta-neutral process" in time t [Doksum (1974)]. This point will be generalized to a nonparametric setting in Section 6. Meeden, Ghosh, Srinivasan and Vardeman (1989) also provide an interesting discussion of the prior and posterior analysis for this model from a decision theoretic viewpoint.

The next result states that for categorical models, the CDBB has the same limiting distribution as the posterior distribution of the survival functions with respect to smooth and fixed priors. In this sense, the CDBB is consistent in approximating the posterior distribution for categorical models. In the following, (y, δ) is the initial segment of the sequence $(y_1, \delta_1), (y_2, \delta_2), \dots$. Let $\lambda_0 = (\lambda_{01}, \dots, \lambda_{0b})$.

THEOREM 4.1. *Suppose the following*:

(i) *For each* $j = 1, \dots, b$, $\hat{\lambda}_j = d_j/r_j \to \lambda_{0j} \in (0, 1)$ *and* $n^{-1}r_j \to 1 - H_{0j} \in (0, 1)$.

(ii) *The* (σ *finite*) *prior density* $\pi'(\lambda)$ *satisfies* (a) $\pi'(\lambda)$ *is continuous and positive in a neighborhood of* λ_0 *and* (b) $\pi'(\lambda) \times \prod_{1 \le j \le b}\lambda_j(1 - \lambda_j)$ *is bounded*.

Then

$$\mathscr{L}\left\{\sqrt{n}\left[S(a_j)/\hat{S}(a_j) - 1\right]: j = 1, \dots, b \mid (y, \delta), \pi\right\} \to \mathscr{L}\{Z_j: j = 1, \dots, b\},$$

where $Z_j = V_1 + \cdots + V_j$, *and* V_j*'s are independent* $N(0, \lambda_{0j}/[(1 - H_{0j})(1 - \lambda_{0j})])$ *random variables*.

REMARK 4.1. (i) Theorem 4.1 is formulated in a way that can be applied to other sampling plans. Suppose the model (2.1) is actually incorrect (say the T_i's are not actually i.i.d.), and a Bayesian computes the posterior distribution based on the incorrect model (2.1). Theorem 4.1 still gives a limit theory for this incorrect modelling as long as $\hat{\lambda}_j = d_j/r_j \to \lambda_{0j} \in (0, 1)$ and $n^{-1}r_j \to 1 - H_{0j} \in (0, 1)$.

(ii) Suppose $\hat{\lambda}_j = d_j/r_j \to \lambda_{0j} \in (0, 1)$ and $n^{-1}r_j \to 1 - H_{0j} \in (0, 1)$ for almost all $(y_1, \delta_1), (y_2, \delta_2), \dots, (y_n, \delta_n), \dots$. Then the conclusion of Theorem 4.1 holds for almost all $(y_1, \delta_1), (y_2, \delta_2), \dots, (y_n, \delta_n), \dots$.

Theorem 4.1 is the consequence of the following two lemmas. In particular, the first lemma, Lemma 4.1, already identifies the limiting distribution of the CDBB.

LEMMA 4.1. $\hat{\lambda}_j = d_j/r_j \to \lambda_{0j} \in (0, 1)$ *and* $n^{-1}r_j \to 1 - H_{0j} \in (0, 1)$, *for each* $j = 1, \dots, b$, *imply*

$$\mathscr{L}\left\{\sqrt{n}\left[S(a_j)/\hat{S}(a_j) - 1\right]: j = 1, \dots, b \mid (y, \delta), \pi_0\right\} \to \mathscr{L}\{Z_j: j = 1, \dots, b\}.$$

PROOF. We first show that, as $\hat{\lambda}_j \to \lambda_{0j}$ and $n^{-1}r_j \to 1 - H_{0j}$

$$\mathscr{L}\left\{\left(\lambda_j - \hat{\lambda}_j\right)/\hat{\sigma}_j|(y,\delta),\pi_0\right\} \to N(0,1),$$

where $\hat{\sigma}_j^2 = (1-\hat{\lambda}_j)\hat{\lambda}_j/(r_j+1)$ for each $j = 1,\dots,b$.

Suppressing the subscript j, let $g_n(t) = \hat{\lambda} + t\hat{\sigma}$. Using

$$P\left\{(\lambda - \hat{\lambda})/\hat{\sigma} \le t\right\} = P\{N \le r - 1 - d\},$$

where N is a binomial $(r-1; 1-g_n(t))$ random variable with mean $E(N)$ and standard deviation $\sigma(N)$. Let $h_n(t) = [r-1-d-E(N)]/\sigma(N)$,

$$P\left\{(\lambda - \hat{\lambda})/\hat{\sigma} \le t\right\} = P\{[N - E(N)]/\sigma(N) \le h_n(t)\}.$$

According to the Berry–Esseen theorem,

$$\sup|P\{[N-E(N)]/\sigma(N) \le h_n(t)\} - \Phi(h_n(t))| = O(1/\sqrt{n}).$$

It remains to show that

$$|\Phi(h_n(t)) - \Phi(t)| \le (2\pi)^{-1/2}|h_n(t) - t| \to 0.$$

Note that $|h_n(t) - t| \to 0$ follows from $\hat{\lambda}_j \to \lambda_{0j}$ and $n^{-1}r_j \to 1 - H_{0j}$.

The usual delta method can then be applied to conclude the proof. □

The following Lemma 4.2 states that for general Bayesian inference, a posterior limiting distribution obtained for a prior π_0 implies that the limiting posterior distribution is shared by other priors "close" to π_0.

LEMMA 4.2. *Suppose a statistical model is defined by* $\mathscr{L}\{\text{data}|\theta\}$, *where* $\theta \in R^k$ *is the parameter. Let* $\pi_0(d\theta)$ *be a* σ*-finite prior for* θ. *The corresponding posterior distribution of* θ *is denoted by* $\mathscr{L}\{\theta|\text{data},\pi_0\}$. *Suppose, as* $n \to \infty$, (a) $\hat{\theta} \to \theta_0$ *and* (b) $\mathscr{L}\{b_n[\theta - \hat{\theta}]|\text{data},\pi_0\} \to \mathscr{L}\{Z\}$ *for some increasing constant* b_n. *Then for any* σ*-finite prior* $\pi(d\theta)$ *such that* $R(\theta) = [d\pi/d\pi_0](\theta)$ (*the derivative of* π *with respect to* π_0) *is bounded, and is continuous and positive in a neighborhood of* θ_0,

$$\text{(i)} \qquad \mathscr{L}\left\{b_n[\theta - \hat{\theta}]|\text{data},\pi\right\} \to \mathscr{L}\{Z\}.$$

If $\theta R(\theta)$ *is also bounded,*

$$\text{(ii)} \qquad \mathscr{L}\left\{b_n\left[\theta - \hat{\theta}_\pi\right]|\text{data},\pi\right\} \to \mathscr{L}\{Z\},$$

where $\hat{\theta}_\pi$ *is the posterior mean of* θ *with respect to the prior* π.

PROOF. Let g be any bounded and continuous function of θ. We have

$$E\left[g\left(b_n[\theta - \hat{\theta}]\right)|\text{data},\pi\right] = N_n/D_n \quad (= 0 \text{ if } D_n = 0),$$

where $N_n = \int g(b_n\varphi)R(\varphi + \hat{\theta})Q_n(d\varphi)$, $D_n = \int R(\varphi + \hat{\theta})Q_n(d\varphi)$ and Q_n is a posterior distribution of $\varphi = [\theta - \hat{\theta}]$ with respect to the prior π_0.

Note that (b) implies that Q_n converges weakly to a point mass at zero. Use a multivariate extension of the analytic Lemma 4.1 in Doksum and Lo (1990) to get

$$\left| N_n - R(\hat{\theta}) \int g(b_n \varphi) Q_n(d\varphi) \right| \to 0$$

and

$$\left| D_n - R(\hat{\theta}) \int Q_n(d\varphi) \right| \to 0.$$

Note that

$$R(\hat{\theta}) \int g(b_n \varphi) Q_n(d\varphi) \to R(\theta_0) E[g(Z)]$$

and

$$R(\hat{\theta}) \int Q_n(d\varphi) = R(\hat{\theta}) \to R(\theta_0) > 0.$$

Hence $N_n/D_n \to E[g(Z)]$. This completes the proof of (i); (ii) can be proved similarly. □

Note that Lemma 4.2 does not require the existence of the likelihood function for the statistical model $\mathscr{L}\{\text{data}|\theta\}$. In applications, we find the posterior limit with respect to a convenient prior π_0 (a conjugate prior, or a σ-finite "flat" prior), and then use Lemma 4.2 to obtain the posterior limit with respect to other smooth priors. Theorem 4.1 is a case in point.

5. A large sample theory for the CDBB. The approximate conditional moments of $S^*(t)$ in Section 3 suggest that the (conditional) limiting distribution for $S^*(t)$ should be identical to that of the Kaplan–Meier estimator [Breslow and Crowley (1974)], and that of Efron's (1981) bootstrap Kaplan–Meier function [Akritas (1986)]. We discuss this phenomenon in this section.

Next is the main central limit theorem. The key condition for (conditional) central limit theorems to hold for Bayesians and bootstrappers is the convergence of some averages of the sample. That is, some laws of large numbers for the sample have to hold. In the case of a functional central limit theorem, the condition is the convergence of some sample empirical functions. In the following, we suppose the observation (y, δ) is given, and is an initial segment of the sequence (y_1, δ_1), $(y_2, \delta_2), \dots$. Let $\hat{Y}(t) = \sum_{1 \le i \le n} \delta_{y_i}[t, \infty)$ denote the number of observations (censored or not) in $[t, \infty)$. For any (sub)distribution function K on the line, let $\tau_K = \sup\{t\colon K(t) < 1\} \le \infty$.

ASSUMPTION 5.1. There exist (sub)distribution functions $F_0 = 1 - S_0$ and H_0 on $[0, \infty)$ such that for each $b < \tau_{H_0}$,

(i) $$\sup_{0 \le t \le b} |\hat{S}(t) - S_0(t)| \to 0,$$

and

(ii) $$\sup_{0 \le t \le b} |n^{-1}\hat{Y}(t) - (1 - H_0(t^-))| \to 0.$$

These conditions are essentially functional forms for condition (i) in Theorem 4.1. A typical situation is that Assumption 5.1 is valid for almost all sample sequences $(y_1, \delta_1), (y_2, \delta_2), \dots$. In this case, (i) is the strong consistency of the Kaplan–Meier estimator; (ii) is a strong law of large numbers.

The functions in (i) and (ii) are cadlag (right continuous with left limits) functions. Let $D[0, b]$ be the space of cadlag functions with uniform metric. The convergence in (i) and (ii) is just convergence of random elements in $D[0, b]$. In this paper, convergence of distributions of random functions means convergence in $D[0, b]$ equipped with the uniform metric and the projection σ-field [see Chapter V in Pollard (1984)]. Let

$$C_0(t) = \int_0^t \{S_0(s)[1 - H_0(s^-)]\}^{-1} S_0(ds).$$

Next is the main limit result. Let $\{W(s);\ s \ge 0\}$ be a standard Brownian motion, and $\{B(s);\ 0 \le s \le 1\}$ be a Brownian bridge.

THEOREM 5.1. *Assumption* 5.1 *implies that*

$$\mathscr{L}\left\{n^{1/2}\left[S^*(\cdot)/\hat{S}(\cdot) - 1\right]|(y, \delta)\right\} \to \mathscr{L}\{W(C_0(\cdot))\}$$

or equivalently,

$$\mathscr{L}\left\{n^{1/2}\left[S^*(\cdot) - \hat{S}(\cdot)\right]|(y, \delta)\right\} \to \mathscr{L}\{S_0(\cdot) \times W(C_0(\cdot))\}.$$

$(S^*(t)/\hat{S}(t) = 1$ *if* $\hat{S}(t) = 0.)$

REMARK 5.1. Suppose the conditions in Assumption 5.1 are valid with $P(\cdot|S_0, H_0)$-probability 1, where $P(\cdot|S_0, H_0)$ is the "true" joint distribution of the sequence $\{(Y_i, \delta_i)\}$. Then the conclusion of Theorem 5.1 is also valid with $P(\cdot|S_0, H_0)$-probability 1. This implication is also valid in all the limit results for the rest of the paper. Akritas (1986) obtained a functional central limit theorem for the CDB in this $P(\cdot|S_0, H_0)$-probability 1 setting.

The above Theorem 5.1, Theorem 2.1 in Akritas (1986) and the Breslow and Crowley limit theorem (1974) [see also Gill (1983)] state that the sampling distribution $\mathscr{L}\{n^{1/2}[\hat{S}(\cdot)/S_0(\cdot) - 1]|S_0\}$, Efron's (1981) CDB distribution $\mathscr{L}\{n^{1/2}[K^*(\cdot)/\hat{S}(\cdot) - 1]|(y, \delta)\}$ and $\mathscr{L}\{n^{1/2}[S^*(\cdot)/\hat{S}(\cdot) - 1]|(y, \delta)\}$ have the same limiting distribution. Therefore, the CDB and the CDBB are first-order asymptotically equivalent, and both CDB and CDBB are consistent bootstrap methods in approximating the sampling distribution of the Kaplan–Meier estimator.

The next corollary states a limit for an asymptotic pivotal quantity, and results in a large-sample approximated band for the CDBB survival function.

This type of large-sample band estimate was first studied by Hall and Wellner (1980); see also Nair (1984) and Andersen and Borgan (1985). For $t \le b < \tau_{H_0}$, let

$$\hat{C}(t) = n\int_0^t \{\hat{S}(s)\hat{Y}(s)\}^{-1}\hat{S}(ds),$$

and

$$\hat{\tau}(t) = \hat{C}(b)^{1/2}\big/\big[\hat{C}(b) + \hat{C}(t)\big].$$

It follows from Theorem 5.1 that:

COROLLARY 5.1. *Assumption* 5.1 *implies* (i) *if* F_0 *is continuous*:

$$\mathscr{L}\Big\{\sup_{0\le t\le b} \hat{\tau}(t)\big|n^{1/2}\big[S^*(t)/\hat{S}(t) - 1\big]\big|(y,\delta)\Big\} \to \mathscr{L}\Big\{\sup_{0\le s\le 1/2} |B(s)|\Big\};$$

(ii)
$$\mathscr{L}\Big\{\int_0^b n^{1/2}\big[S^*(s) - \hat{S}(s)\big]\,ds|(y,\delta)\Big\} \to \mathscr{L}\Big\{\int_0^b [S_0(s) \times W(C_0(s))]\,ds\Big\}.$$

The passage from Theorem 5.1 to Corollary 6.1 suggests that if the functional of interest $\theta(S^*, \hat{S})$ can be written as a functional of $S^* - \hat{S}$, say $g(S^* - \hat{S})$ where g may depend on $\hat{S}$, then the CDBB distribution $\mathscr{L}\{[\theta(S^* - \hat{S})]|(y,\delta)\}$ approximates the posterior distribution

$$\mathscr{L}\big\{\big[\theta(S - \hat{S})\big]|(y,\delta)\big\},$$

provided that $\mathscr{L}\{n^{1/2}[S(\cdot) - \hat{S}(,)]|(y,\delta)\}$ also converges to $\mathscr{L}\{S_0(\cdot) \times W(C_0(\cdot))\}$. It follows from Theorem 4.1 that the CDBB is consistent for categorical models with respect to smooth priors. For nonparametric models, Corollary 6.1 in the next section states that the CDBB is consistent if the prior of $S(\cdot)$ is a member of the beta-neutral process priors.

REMARK 5.2. If the functional of interest $\theta(S^*, \hat{S})$ can be written as

$$\theta(S^*, \hat{S}) = \theta(S^*) - \theta(\hat{S}),$$

where $\theta(\cdot)$ is a "differentiable" function of its argument [Chapter 6 in Serfling (1980), Filippova (1961) and Gill (1989)]. Recent work of Arcones and Giné (1991) [see also Babu (1984)] on (frequentist) bootstrapping U- and V-statistics, and Theorem 5.1 suggest that, quite often $\mathscr{L}\{n^{k/2}g_k(\hat{S}; S^* - \hat{S})|(y,\delta)\}$ and $\mathscr{L}\{n^{k/2}[\theta(S^*) - \theta(\hat{S})] - \Sigma_{1\le j\le k-1} g_j(\hat{S}; S^* - \hat{S})|(y,\delta)\}$ have the same limiting distribution, where $\mathscr{L}\{n^{j/2}g_j(\hat{S}; S^* - \hat{S})|(y,\delta)\}$ has a degenerate (at zero) limiting distribution for $j = 1,\dots,k-1$. Usually

$$n^{j/2}g_j(\hat{S}; S^* - \hat{S}) = \int\cdots\int h_j(s)\prod_{1\le i\le j} n^{1/2}(S^* - \hat{S})(ds_i)$$

for a suitable function $h_j(s)$ (depending on $\hat{S}$), $j = 1, \ldots k$. In view of Corollary 6.1 in Section 6, these arguments are also valid with S replacing S^*, and the posterior distributions

$$\mathscr{L}\left\{n^{k/2}\left[\theta(S) - \theta(\hat{S})\right] - \sum_{1 \le j \le k-1} g_j(\hat{S}; S - \hat{S}) | (y, \delta)\right\} \quad \text{and}$$

$$\mathscr{L}\left\{n^{k/2} g_k(\hat{S}; S - \hat{S}) | (y, \delta)\right\}$$

have the same limiting distribution.

We next turn to the proof of Theorem 5.1. We will first prove a central limit theorem for the CDBB cumulative hazard

$$\Lambda^*(t) = \sum_{j:\, t(j) \le t} \sum_{q \in D(j)} Z_q \Big/ \sum_{q \in R(j)} Z_q. \tag{5.1}$$

Since Λ^* is a cumulative sum of independent beta $(d_j, r_j - d_j)$ random variables,

$$\begin{aligned} E[\Lambda^*(t) | (y, \delta)] &= \sum_{j:\, t(j) \le t} d_j / r_j \\ &= \int_0^t \hat{Y}(s)^{-1} N_u(ds), \end{aligned} \tag{5.2}$$

where $N_u(t) = \Sigma_u \delta_{y_u}[0, t]$. Furthermore, since

$$\hat{S}(t) = \prod_{s \le t} \left[1 - \Delta N_u(s) / \hat{Y}(s)\right] = \prod_{s \le t} \left[1 - \Delta \hat{\Lambda}(s)\right],$$

$$\Delta \hat{\Lambda}(s) = \Delta \hat{S}(s) / \hat{S}(s^-)$$

and

$$\hat{\Lambda}(t) = \int_0^t \hat{S}(ds) / \hat{S}(s^-).$$

Also,

$$\begin{aligned} \operatorname{Var}[\Lambda^*(t) | (y, \delta)] &= \sum_{j:\, t(j) \le t} (r_j + 1)^{-1} (1 - d_j / r_j) d_j / r_j \\ &= \int_0^t \{\hat{Y}(s) + 1\}^{-1} \left[1 - \Delta \hat{\Lambda}(s)\right] \hat{\Lambda}(ds). \end{aligned}$$

Let

$$\Lambda_{H_0}(t) = \int_0^t \{1 - H_0(s^-)\}^{-1} [1 - \Delta \Lambda_0(s)] \Delta_0(ds).$$

The next Lemma 5.1 gives a functional central limit theorem for the CDBB cumulative hazard process Λ^*, and prepares for the proof of Theorem 5.1.

LEMMA 5.1. *Assumption* 5.1 *implies that,*

$$\mathscr{L}\left\{n^{1/2}\left[\Lambda^*(\cdot) - \hat{\Lambda}(\cdot)\right] \mid (y,\delta)\right\} \to \mathscr{L}\left\{W\left(\Lambda_{H_0}(\cdot)\right)\right\}. \tag{5.3}$$

PROOF. Suppose the uncensored data y_u's in the sequence (y_1, δ_1), $(y_2, \delta_2), \dots$ are all distinct, and S_0 is continuous. We first show convergence in $D[0, b]$ equipped with Skorohod metric. The convergence of finite-dimensional distributions follows from Assumption 5.1 and Lindeberg's theorem. To show tightness, we let

$$X_n^*(t) = n^{1/2}\left[\Lambda^*(t) - \hat{\Lambda}(t)\right].$$

Note that for $t_1 \le t_2$,

$$\begin{aligned} P\{|X_n^*(t_2) - X_n^*(t_1)| \ge \varepsilon \mid (y,\delta)\} &\le \varepsilon^{-2} E\left[|X_n^*(t_2) - X_n^*(t_1)|^2 \mid (y,\delta)\right] \\ &\le \varepsilon^{-2} n \int_{t_1}^{t_2} \{\hat{Y}(s) + 1\}^{-1}\left[1 - \Delta\hat{\Lambda}(s)\right]\hat{\Lambda}(ds) \\ &= \varepsilon^{-2} n \int_{t_1}^{t_2} \{\hat{Y}(s) + 1\}^{-1}\left[\hat{S}(s)/\hat{S}^2(s^-)\right]\hat{S}(ds). \end{aligned}$$

Therefore

$$\begin{aligned} &\limsup_n P\{|X_n^*(t_2) - X_n^*(t_1)| \ge \varepsilon \mid (y,\delta)\} \\ &\quad\le \varepsilon^{-2} \limsup_n n \int_{t_1}^{t_2} \{\hat{Y}(s) + 1\}^{-1}\left[\hat{S}(s)/\hat{S}^2(s^-)\right]\hat{S}(ds) \\ &\quad= \varepsilon^{-2} \int_{t_1}^{t_2} \{1 - H_0(s^-)\}^{-1}\left[S_0(s)/S_0^2(s^-)\right] S_0(ds) \end{aligned}$$

by Assumption 5.1. According to the limiting form of the fluctuation inequality of Billingsley (1968; Theorem 15.6, extended to $D[0, b]$), the sequence $X_n^*(\cdot)$ is tight. Hence, $\mathscr{L}\{X_n^*(\cdot) \mid (y,\delta)\} \to \mathscr{L}\{W(\Lambda_{H_0}(\cdot))\}$ in $D[0, b]$ equipped with the Skorohod metric. Since $W(\Lambda_{H_0}(\cdot))$ has continuous paths, the convergence is also valid in $D[0, b]$ equipped with the uniform metric.

Next, according to Skorohod's a.s. representation, there are $D[0, b]$-valued random functions $\tilde{X}_n(\cdot)$ and $\tilde{W}(\cdot)$ such that

$$\mathscr{L}\left\{\tilde{X}_n(\cdot) \mid (y,\delta)\right\} = \mathscr{L}\{X_n^*(\cdot) \mid (y,\delta)\},$$

$$\mathscr{L}\left\{\tilde{W}\left(\Lambda_{H_0}(\cdot)\right)\right\} = \mathscr{L}\left\{W\left(\Lambda_{H_0}(\cdot)\right)\right\}$$

and

$$\sup_{0 \le s \le b} \left|\tilde{X}_n(s) - \tilde{W}\left(\Lambda_{H_0}(s)\right)\right| \to 0 \quad \text{almost surely } P(\cdot \mid y, \delta).$$

By Assumption 5.1,

$$\sup_{0\le s\le b}\left|\tilde W(\Lambda_{H_0}(s)) - \tilde W(\hat\Lambda(s))\right| \to 0 \quad \text{almost surely } P(\cdot\,|y,\delta).$$

Therefore, for any subset of the y_u's, say $t(1) < \cdots < t(k)$, we have

$$\begin{aligned}(5.4)\qquad &\max_{1\le j\le k}\left|\tilde X_n(t(j)) - \tilde W(\hat\Lambda(t(j)))\right|\\ &\le \sup_{0\le s\le b}\left|\tilde X_n(s) - \tilde W(\hat\Lambda(s))\right| \to 0 \quad \text{almost surely } P(\cdot\,|y,\delta).\end{aligned}$$

In the general case (the y_u's may have ties, and S_0 may be discontinuous), let $t(1) < \cdots < t(k)$ be the distinct values of the y_u's. The corresponding $X_n^*(\cdot)$ is constant between adjacent $t(j)$'s, and at $t(j)$, $X_n^*(\cdot)$ is equal to the value of an $X_n^*(\cdot)$ corresponding to a sequence $(y_1, \delta_1), (y_2, \delta_2), \dots$ with distinct y_u's. The behaviour of $\hat\Lambda(\cdot)$ is similar. By (5.4), a version of $X_n^*(\cdot)$, also denoted by $\tilde X_n(s)$, also obeys

$$\sup_{0\le s\le b}\left|\tilde X_n(s) - \tilde W(\hat\Lambda(s))\right| \to 0 \quad \text{almost surely } P(\cdot\,|y,\delta).$$

Therefore, $\mathscr{L}\{\tilde X_n(\cdot)|(y,\delta)\}$ and $\mathscr{L}\{\tilde W(\tilde\Lambda(\cdot))\}$ have the same limit. It remains to note that $\mathscr{L}\{\tilde W(\hat\Lambda(\cdot))\} \to \mathscr{L}\{\tilde W(\Lambda_{H_0}(\cdot))\}$ under Assumption 5.1. □

REMARK 5.3. The preceding method of proof avoids the explicit construction of quantile transforms [Akritas (1986) and Lo (1988)]. This method is essentially an example of a technique mentioned by Le Cam [(1986), page 531].

PROOF OF THEOREM 5.1. Put $\Lambda^* = A$, and $\hat\Lambda = B$ in Proposition A.4.1. in Gill [(1980), page 153]. Then, $Z_n^* = n^{1/2}[S^*(\cdot)/\hat S(\cdot) - 1]$ can be represented as

$$(5.5)\qquad Z_n^*(t) = -\int_0^t M^*(s)\sqrt{n}\left\{\Lambda^*(ds) - \hat\Lambda(ds)\right\},$$

where

$$M^*(s) = \left[S^*(s^-)/\hat S(s^-)\right] \times \left[1 - \Delta\hat\Lambda(s)\right]^{-1} = S^*(s^-)/\hat S(s).$$

Define a map h from $D_L[0,b] \times D[0,b]$ to $D[0,b]$ by $h(u(\cdot), v(\cdot)) = z(\cdot)$, where

$$(5.6)\qquad z(t) = \int_0^t u(s)v(ds),$$

and $D_L[0,b]$ is the space of caglad (left continuous with right limits) functions equipped with uniform metric. Convergence of order pairs is defined to be coordinatewise convergence. The map $h(\cdot,\cdot)$ is continuous.

Let $U^*(s) = S^*(s^-)/\hat S(s^-) = M^*(s)\hat S(s)/\hat S(s^-)$, and $V^*(t) = \int_0^t[\hat S(s^-)/\hat S(s)]X_n^*(ds)$. Then $h(U^*, V^*) = Z_n^*$.

Since $|\prod_j a_j - \prod_j b_j| \le \sum_j |a_j - b_j|$ for a's and b's with norm 1,

$$\sup_{t\le b}|S^*(t) - \hat{S}(t)| \le \sum_{j:\, t(j)\le b} |\lambda_j - d_j/r_j|, \tag{5.7}$$

where λ_j's are independent beta $(d_j; r_j - d_j)$ random variables. Hence,

$$\begin{aligned} P\Big\{\sup_{t\le b}|S^*(t) - \hat{S}(t)| \ge \varepsilon|(y,\delta)\Big\} & \\ \le \varepsilon^{-2}\int_0^b \{\hat{Y}(s) + 1\}^{-1}\big[1 - \Delta\hat{\Lambda}(s)\big]\hat{\Lambda}(ds) & \\ = O(n^{-1}) & \end{aligned} \tag{5.8}$$

by Assumption 5.1. Therefore,

$$\sup_{t\le b}|U^*(t) - 1| \to 0 \quad \text{in } P(\cdot|y,\delta)\text{-probability}. \tag{5.9}$$

Furthermore, Lemma 5.1 implies that

$$\mathscr{L}\{V^*(\cdot)|(y,\delta)\} \to \mathscr{L}\{W(C_0(\cdot))\}. \tag{5.10}$$

An application of the continuous mapping theorem [Pollard (1984), page 70] concludes the proof. □

REMARK 5.4. In a previous version, a direct proof of a probability 1 version of Theorem 5.1 is given based on an application of Rebolledo's (1980) central limit theorem for martingale process, and the fact that Z_n^* in (5.5) is a martingale process with compensator

$$\begin{aligned} \langle Z_n^*, Z_n^*\rangle(t) &= n\int_0^t [M^*(s)]^2 \langle X_n^*, X_n^*\rangle(ds) = \operatorname{Var}\{X_n^*(t)|(y,\delta)\} \\ &= n\int_0^t [M^*(s)]^2 \{\hat{Y}(s) + 1\}^{-1}\big[1 - \Delta\hat{\Lambda}(s)\big]\hat{\Lambda}(ds) \\ &= n\int_0^t \big[S^*(s^-)/\hat{S}(s^-)\big]^2 \big\{\hat{S}(s)\big[\hat{Y}(s) + 1\big]\big\}^{-1}\hat{S}(ds). \end{aligned} \tag{5.11}$$

This martingale-method proof is parallel to the work of Gill (1980) on weak convergence of the Kaplan–Meier estimator, and the work of Akritas (1986) on the frequentist bootstrap version of it.

6. A limiting posterior distribution for a survival function.

Suppose the survival function may not be categorical, and $\pi(dS(\cdot))$ is a prior on the space of survival functions. Will the posterior distribution with respect to this prior π, that is, $\mathscr{L}\{n^{1/2}[S(\cdot)/\hat{S}(\cdot) - 1]|(y,\delta),\pi\}$, and its CDBB analogue $\mathscr{L}\{n^{1/2}[S^*(\cdot)/\hat{S} - 1]|(y,\delta)\}$ have the same limiting distribution? An affirmative answer to this question will indicate that the CDBB is a consistent bootstrap method for a Bayesian equipped with the prior π. Bayesians had been remarkably silent on this topic: The limiting posterior distribution of $\mathscr{L}\{n^{1/2}[S(\cdot)/\hat{S}(\cdot) - 1]|(y,\delta),\pi\}$ had not yet been found for any nonparametric prior $\pi(dS(\cdot))$. The following Corollary 6.1 is a result in this direction (see also Corollary 6.2 and Remark 1.1).

The lack of such a Bayesian version of the Breslow–Crowley limit theorem stemmed from the complexities in the description of the posterior distribution of the survival function. Furthermore, different researchers seem to favor different model assumptions. For example, Susarla and Van Ryzin (1976) assume that the G_i's are known and identical, whereas Ferguson and Phadia (1979) assume that G_i is a point mass distribution (a deterministic model). It is not clear that whether the results obtained for one model of (2.1) also apply to another model of (2.1). This point was also mentioned by Ferguson and Phadia [(1979), page 178]. Lemma 7.1 in Section 7 clarifies this situation. Essentially it states that, if F and the G_i's are independent according to the prior, once the censored data are given, one can operate as if one is working with a deterministic model (with the G_i's being point mass at the incomplete data). That is, with such an independent prior, knowing G_i's or not does not affect the statistical inference of $S(\cdot)$.

Next, we turn to discuss the posterior distributions. The first approach [Susarla and Van Ryzin (1976)] assumes a Dirichlet prior [Ferguson (1973)] for the survival function $S(t)$. However, Blum and Susarla (1977) identified the posterior distribution as a mixture of Dirichlet processes [Antoniak (1974)], which is difficult to handle. In a far-reaching study, Ferguson and Phadia (1979) showed that neutral processes priors [Doksum (1974)] on $S(t)$ are conjugate priors, and that the posterior moment generating function of $S(t)$ can be evaluated. In view of Lemma 7.1, Ferguson and Phadia's result implies that the mixture of Dirichlet processes encountered by Susarla and Van Ryzin (1976) is a neutral process. An important technical point for the Doksum–Ferguson–Phadia approach is that they reparametrized the model by the cumulative hazard $Y(t) = -\log(S(t))$, and the prior to posterior analysis is essentially carried out for the independent increment process $Y(t)$.

When this paper was written, Hjort (1990) appeared. This section is rewritten to relate to his approach. Hjort (1990) reparametrized the survival function by the so-called "infinitesimal" cumulative hazard $A(t)$. He gave a full description of a posterior distribution of $A(t)$ with respect to an independent increment process prior for $A(t)$. An independent increment cumulative hazard rate $A(t)$ gives rise to a neutral process survival function $S(t)$, and Ferguson and Phadia's (1979) results on the posterior distribution of a neutral survival process can be applied to obtain Hjort's theory. This was also noted by Hjort [(1990), page 1274]. On the other hand, Hjort's (1990) method produces a class of beta cumulative hazard processes which is a conjugate family of priors for the cumulative hazard [see Corollary 4.1 in Hjort (1990)], and the following Corollary 6.1 states that the posterior distribution of the corresponding beta-neutral survival process has a Breslow–Crowley limit.

We first discuss a simple beta-neutral survival process which does not require an "infinitesimal" definition. An advantage of the simple beta-neutral process is that it can be understood without infinitesimal calculus. For any measure ν on $[0, \infty)$, let $\nu(t)$ be the corresponding cumulative function $\nu(t) =$

$\nu\{[0, t]\}$. Let α be a finite measure on $[0, \infty)$, let β be another finite measure with atoms at the c_j's,

$$0^- = c_0 < c_1 < \cdots < c_m < c_{m+1} = \infty.$$

Let $\mu_\alpha(t)$ be a gamma process with shape measure α [i.e., (a) $\mu_\alpha(t)$ is a gamma $(\alpha(t); 1)$ random variable for each t, and (b) $\mu_\alpha(\cdot)$ is an independent increment process]; independently let $\mu_\beta(t)$ be a gamma process with shape measure β. Define a neutral process as follows: for $t \in (c_k, c_{k+1}]$,

$$\tilde{S}_{\alpha,\beta}(t) = \prod_{j \le k} \frac{\mu_\alpha(c_j, \infty) + \mu_\beta(c_{j-1}, \infty)}{\mu_\alpha(c_{j-1}, \infty) + \mu_\beta(c_{j-1}, \infty)} \times \frac{\mu_\alpha(t, \infty) + \mu_\beta(c_k, \infty)}{\mu_\alpha(c_k, \infty) + \mu_\beta(c_k, \infty)}. \tag{6.1}$$

Note that $\tilde{S}_{\alpha,\beta}$ relativized to the interval $(c_k, c_{k+1}]$ eliminates the effect of β, and is a Dirichlet process on $(c_k, c_{k+1}]$ with shape measure α.

It is tempting to let the step function β tend to a limit, also denoted by β, and the resulting beta-neutral process will be a natural extension to (6.1). Indeed, this limiting procedure has been carried out by Hjort (1990) in the context of the beta cumulative hazard model, which requires perhaps unnecessary restrictions on α and β (see Remark 6.1). A direct approach, pioneered by Ferguson (1973) and Doksum (1974), is more desirable. This leads to the following path-wise definition of a beta-neutral process for general α and β.

To do this, note that for any collection of a_i's such that $\{a_i\}$ contains $\{c_i\}$, $\tilde{S}_{\alpha,\beta}(t)$ in (6.1) can be rewritten as a finer cumulative product with the a_i's playing the same role as the c_i's. This "self-consistency" property suggests the following construction of a beta-neutral survival process. This construction is analogous to Ferguson's (1973) second definition of a Dirichlet process based on a gamma process. Thinking as a Bayesian, one sees that Ferguson (1973) uses one gamma process to summarize the information conveyed by prior complete data. It is therefore natural to expect that another gamma process is required to summarize the prior information carried by the incomplete data. Let α and β be finite measures that summarize the prior information carried by prior complete and incomplete data, respectively. Let μ_α and μ_β be independent gamma processes with shape measures α and β, respectively. Note that μ_α and μ_β have discrete paths (see Remark 6.2). Therefore we can define, path-wise, a Bayesian copy of the Kaplan–Meier function

$$S_{\alpha,\beta}(t) = \prod_{y: y \le t} \frac{\mu_\alpha(y, \infty) + \mu_\beta[y, \infty)}{\mu_\alpha[y, \infty) + \mu_\beta[y, \infty)} = \prod_{y: y \le t} \left\{1 - \frac{\Delta\mu_\alpha(y)}{\mu_\alpha[y, \infty) + \mu_\beta[y, \infty)}\right\}. \tag{6.2}$$

The product is overall (random) y such that $\Delta\mu_\alpha(y) = \mu_\alpha(\{y\}) > 0$.

DEFINITION 6.1. $S_{\alpha,\beta}(t)$ is called a beta-neutral process with parameters α and β. Notation: $S_{\alpha,\beta}(\cdot) \sim \mathrm{BN}(\alpha;\beta)$.

If β is atomic, then (6.2) reduces to (6.1) (note that $(0^-, c_1]$ is $[0, c_1]$). If $\beta = 0$, (6.2) reduces to $\mu_\alpha(t)/\mu_\alpha[0,\infty)$, which is Ferguson's (1973) second definition of a Dirichlet process.

Recall $\{y_u\}$ is the subset of uncensored data, and let $\{y_c\}$ be the subset of incomplete data. Note that $\{y_u\} \cup \{y_c\} = \{y_i: i = 1,\dots,n\}$. We will study the limiting distribution of a $\mathrm{BN}(\alpha + \Sigma_u \delta_{y_u}; \beta + \Sigma_c \delta_{y_c})$ process conditional on (y,δ). Note that Hjort (1990), Corollary 4.1, can be used to deduce that a $\mathrm{BN}(\alpha + \Sigma_u \delta_{y_u}; \beta + \Sigma_c \delta_{y_c})$ distribution is the posterior distribution of a survival function if the prior is a $\mathrm{BN}(\alpha;\beta)$ survival process. Hjort's result requires both $\alpha(t)$ and $\beta(t)$ to be piece-wise continuous, and that α has at most finitely many jumps (see Remark 6.1). These conditions are perhaps unnecessary. Hjort's result (extended to an arbitrary and finite α) also implies that if one assumes a simple beta-neutral process prior (6.1), the posterior survival function is a $\mathrm{BN}(\alpha + \Sigma_u \delta_{y_u}; \beta + \Sigma_c \delta_{y_c})$ process. Since a $\mathrm{BN}(\alpha, 0)$ process is a Dirichlet process, the mixture of Dirichlet processes encountered by Susarla and Van Ryzin (1976) is in fact a $\mathrm{BN}(\alpha + \Sigma_u \delta_{y_u}; \beta + \Sigma_c \delta_{y_c})$ process. While these results can also be deduced from Ferguson and Phadia's (1979) work on neutral processes with censored data, it is expected that a careful execution of Doksum's (1974) method of proof will yield the result for arbitrary (finite) α and β. [Both Ferguson and Phadia (1979) and Hjort (1990) use the Lévy representation for an independent increment process as their main technique.] We now turn to the asymptotic theory.

The path-wise definition (6.2) for a beta-neutral process also paves the way for an easy proof of a large-sample theory for the posterior distribution of the survival function. This method, however, requires that we "tailor-make" a posterior $\mathrm{BN}(\alpha + \Sigma_u \delta_{y_u}; \beta + \Sigma_c \delta_{y_c})$ process. Let μ_α and μ_β be two independent gamma processes and let $Z_1, Z_2, \dots$ be i.i.d. standard exponential random variables; the μ's and the Z's are assumed to be independent. Given the data (y,δ), let $\mu_{\alpha_n}(t) = \mu_\alpha(t) + \Sigma_u Z_u \delta_{y_u}(t)$ and $\mu_{\beta_n}(t) = \mu_\beta(t) + \Sigma_c Z_c \delta_{y_c}(t)$. Routine computation shows that $\mu_{\alpha_n}(t)$ and $\mu_{\beta_n}(t)$ are gamma processes with shape measures $\alpha_n = \alpha + \Sigma_u \delta_{y_u}$ and $\beta_n = \beta + \Sigma_c \delta_{y_c}$, respectively. Following the path-wise definition of a "prior" beta-neutral process given in (6.2), we define, path-wise, a "posterior" beta-neutral process as follows:

$$S_{\alpha_n,\beta_n}(t) = \prod_{y: y \le t} \left\{ 1 - \frac{\Delta \mu_{\alpha_n}(y)}{\mu_{\alpha_n}[y,\infty) + \mu_{\beta_n}[y,\infty)} \right\}. \tag{6.3}$$

Note that $S_{\alpha_n,\beta_n}(t)$ is a beta-neutral $(\alpha_n;\beta_n)$ process. Letting $\alpha \to 0$ and $\beta \to 0$ results in a $\mathrm{BN}(\Sigma_u \delta_{y_u}; \Sigma_c \delta_{y_c})$ distribution for a survival function. This is the CDBB distribution of $S^*(\cdot)$, as can be seen from (6.3) with $\alpha = 0$ and $\beta = 0$. In this sense, the CDBB distribution for $S^*(\cdot)$ is a posterior distribution with respect to a "flat" beta-neutral prior. The next result states that the (uniform) distance between $S_{\alpha_n,\beta_n}(t)$ and $S^*(t)$ goes to zero as fast as n^{-1}.

THEOREM 6.1. *Assumption* 5.1 *implies that*

$$\sup_{t \le b} |S_{\alpha_n, \beta_n}(t) - S^*(t)| = O(n^{-1}), \quad a.s.\ P(\cdot | y, \delta).$$

PROOF. We shall apply the inequality $|\prod_j a_j - \prod_j b_j| \le \sum_j |a_j - b_j|$ for a's and b's with norm 1. Let $Y^*(s) = \sum_i Z_i \delta_{y_i}[s, \infty)$ and $N_u^*(t) = \sum_u Z_u \delta_{y_u}[0, t]$:

$$(6.4)\qquad \begin{aligned} |S_{\alpha_n, \beta_n}(t) - S^*(t)| &= \sum_{y: y \le t} \left| \frac{\Delta \mu_{\alpha_n}(y)}{\mu_{\alpha_n}[y, \infty) + \mu_{\beta_n}[y, \infty)} - \frac{\Delta N_u^*(y)}{Y^*(y)} \right| \\ &= \sum_{y: y \le t} \{A_y + B_y\}, \end{aligned}$$

where

$$A_y = \frac{\Delta \mu_\alpha(y)}{\mu_{\alpha_n}[y, \infty) + \mu_{\beta_n}[y, \infty)}$$

and

$$B_y = \frac{\Delta N_u^*(y)}{Y^*(y)} + \frac{\mu_\alpha[y, \infty) + \mu_\alpha[y, \infty)}{\mu_{\alpha_n}[y, \infty) + \mu_{\alpha_n}[y, \infty)}.$$

It remains to bound $\sum_{y: y \le t} A_y$ and $\sum_{y: y \le t} B_y$. Let $y(n) = \max\{y_i: i = 1, \dots, n\}$, and $b_n = \min\{b, y(n)\}$:

$$(6.5)\qquad \begin{aligned} \sum_{y: y \le t} A_y &\le \mu_\alpha(t)/\mu_{\alpha_n + \beta_n}[t, \infty) \\ &\le \mu_\alpha(b_n)/\mu_{\alpha_n + \beta_n}[b_n, \infty) \\ &= O(1/n) \quad \text{a.s. } P(\cdot | y, \delta) \text{ by Assumption 5.1.} \end{aligned}$$

[This is a conditional big "O", i.e., the constant in the big "O" is a finite random variable.] Next,

$$(6.6)\qquad \begin{aligned} \sum_{y: y \le t} B_y &\le \{\mu_{\alpha+\beta}[0, \infty)/\mu_{\alpha_n + \beta_n}[t, \infty)\} \times \{N_u^*(t)/Y^*(t)\} \\ &\le \{\mu_{\alpha+\beta}[0, \infty)/\mu_{\alpha_n + \beta_n}[b_n, \infty)\} \times \{N_u^*(b_n)/Y^*(b_n)\} \\ &= O(1/n) \quad \text{a.s. } P(\cdot | y, \delta) \text{ by Assumption 5.1.} \end{aligned}$$

□

Theorem 5.1 and Theorem 6.1 imply the following Bayesian version of the Breslow and Crowley limit theorem (see also Remark 6.3). Define

$$Z_n(t) = n^{1/2}\left[S_{\alpha_n, \beta_n}(t)/\hat{S}(t) - 1\right].$$

COROLLARY 6.1. *Assumption* 5.1 *implies*

(i) $\mathscr{L}\left\{n^{1/2}\left[S_{\alpha_n, \beta_n}(\cdot)/\hat{S}(\cdot) - 1\right] | (y, \delta)\right\} \to \mathscr{L}\{W(C_0(\cdot))\}$, *or equivalently*

$$\mathscr{L}\left\{n^{1/2}\left[S_{\alpha_n,\beta_n}(\cdot) - \hat{S}(\cdot)\right]|(y,\delta)\right\} \to \mathscr{L}\{S_0(\cdot) \times W(C_0(\cdot))\};$$

(ii) *if* F_0 *is continuous,*

$$\mathscr{L}\left\{\sup_{0\le t\le b} \hat{\tau}(t)|n^{1/2}\left[S_{\alpha_n,\beta_n}(t)/\hat{S}(t) - 1\right]|\,|(y,\delta)\right\} \to \mathscr{L}\left\{\sup_{0\le s\le 1/2} |B(s)|\right\};$$

(iii)

$$\mathscr{L}\left\{\int_0^b n^{1/2}\left[S_{\alpha_n,\beta_n}(s) - \hat{S}(s)\right] ds|(y,\delta)\right\} \to \mathscr{L}\left\{\int_0^b [S_0(s) \times W(C_0(s))]\, ds\right\}.$$

We conclude this study by a discussion of the large-sample behavior of Hjort's (1990) beta cumulative hazard which is not the emphasis in this paper. An inspection of (6.2) suggests the following alternative path-wise definition of a "prior" beta cumulative hazards

$$\Lambda_{\alpha,\beta}(t) = \int I_{\{0\le s\le t\}}\mu_\alpha(ds)/\left\{\mu_\alpha[s,\infty) + \mu_\beta[s,\infty)\right\}. \tag{6.7}$$

A moment of reflection, using Theorem 1.2.3 in Bickel and Doksum (1977), indicates that $\Lambda_{\alpha,\beta}(t)$ is an independent increment process. A "posterior" beta cumulative hazard is given by $\Lambda_{\alpha_n,\beta_n}(t)$. From the proof of Theorem 6.1, we note that:

THEOREM 6.2. *Assumption* 5.1 *implies that*

$$\sup_{t\le b}|\Lambda_{\alpha_n,\beta_n}(t) - \Lambda^*(t)| = O(n^{-1}), \quad a.s.\ P(\cdot|y,\delta).$$

The following result then follows from Lemma 5.1.

COROLLARY 6.2. *Assumption* 5.1 *implies that*

$$\mathscr{L}\left\{\sqrt{n}\left[\Lambda_{\alpha_n,\beta_n}(\cdot) - \hat{\Lambda}(\cdot)\right]\middle|(y,\delta)\right\} \to \mathscr{L}\left\{W(\Lambda_{H_0}(\cdot))\right\}.$$

REMARK 6.1. Corollary 4.1 in Hjort (1990) requires that the prior parameters for his "infinitesimal" hazard rate model be (i) $c(t)$ is piecewise continuous and (ii) $A_0(t)$ jumps only a finite number of times. Note that $c(t) = \alpha[t,\infty) + \beta[t,\infty)$ and $A_0(t) = \int_0^t\{\alpha[s,\infty) + \beta[s,\infty)\}^{-1}\alpha(ds)$. For the survival function model considered here, these conditions become (i) both α and β are piecewise continuous and (ii) $\alpha(t)$ jumps only a finite number of times.

REMARK 6.2. For this path-wise construction to be well defined, it is required that μ_α (and μ_β), rather than the usual a "version" of it, to have discrete paths. A simple proof of this phenomenon is given in Corollary 3.1 in Lo and Weng (1989). See also Kingman (1975) which discusses the fine points of the path properties of a random distribution, and that of the path property of its "versions."

REMARK 6.3. Several authors suggest the investigation of the sampling distribution of the posterior mean [Susarla and Van Ryzin (1978) and Hjort (1990)]. Inequalities (6.4), (6.5) and (6.6) yield that the uniform distance between $E[S_{\alpha_n,\beta_n}(t)|(y,\delta)]$ and $\hat{S}(t)$ (and between $E[\Lambda_{\alpha_n,\beta_n}(t)|(y,\delta)]$ and $\hat{\Lambda}(t)$) is $O(n^{-1})$. Hence, Susarla and Van Ryzin's (1978) result on the limiting sampling distribution of $E[S_{\alpha_n,\beta_n}(t)|(y,\delta)]$ follows from that of $\hat{S}(t)$, which was obtained by Breslow and Crowley (1974). However, the study of the posterior distribution, rather than its mean, is of primary interest to Bayesians.

7. Conjugate independent priors for sampling from a random censoring model. In this section, we discuss a conjugate prior property for sampling from a random censoring model. This result states essentially that if the model is defined by (2.1), and if the survival function F and the censoring distributions G are independent under the prior distribution, $F,(Y,\delta),G$ is a three-term Markov chain. As an application, the posterior means of the neutral survival process derived by Ferguson and Phadia (1979) under a deterministic censoring model apply also to random censoring models.

Recall model (2.1): $T=\{T_1,\dots,T_n\}$, $C=\{C_1,\dots,C_n\}$ and $G=\{G_1,\dots,G_n\}$. For $i=1,\dots,n$, $Y_i=\min\{T_i,C_i\}$; $\delta_i=1$ if $T_i\le C_i$ and $\delta_i=0$ if $T_i>C_i$. Suppose $(Y,\delta)=(y,\delta)$, then (y,δ) carry the same amount of information as $\{y_u\colon u\in U\}\cup\{y_c\colon c\in C\}$ where $U=\{i\colon \delta_i=1\}$ and $C=\{i\colon \delta_i=0\}$.

LEMMA 7.1. *Suppose F and G are independent according to the prior distribution, and given F and G, (Y,δ) is a sample from model* (2.1). *Then, given $(Y,\delta)=(y,\delta)$,* (i) *F and G remain to be independent, and* (*ii*) *the posterior distribution of F is given by, for all* (*measurable*) $h\ge 0$,

$$E[h(F)|(Y,\delta)=(y,\delta),G]$$
$$=\int\int h(F)\pi(df|T)P\{d(T_c)|T_u=y_u,\ u\in U;\ T_c>y_c,\ c\in C\},$$

provided that $P\{T_c>y_c,\ c\in C|T_u=y_u,\ u\in U\}>0$.

PROOF. Note that

$$\pi(dF|Y,\delta,G)=\int\pi(dF|T,Y,\delta,G)P(dT|Y,\delta,G)$$
$$=\int\pi(dF|T)P\{dT|Y,\delta,G\}$$

since $F,T,(Y,\delta,G)$ is a three-term Markov chain.

It remains to identify $P(dT|Y, \delta, G)$.

(i) Suppose $\delta = 1$, then $Y = T$, and $P\{dT|Y, 1, G\}$ is a point mass at T.

(ii) Suppose $\delta = 0$, then $Y = C$, and $P\{dT|Y, 0, G\} = P\{dT|C, 0\}$ is the conditional distribution of T given $T > C$, defined for almost all C.

To complete the proof of the Lemma, we first use (i) to update the prior based on the complete data to get $\pi(dF|y_u, u \in U)$, and then use (ii) to update this posterior based on the incomplete data $\{y_c, c \in C\}$ to conclude the proof of Lemma 7.1. □

Acknowledgments. This paper is dedicated to the late Betty Scott. The author also would like to acknowledge that a stimulating conversation with Michael Woodroofe precipitated this work on the bootstrap theory. The author is grateful to a referee for a very careful review and for remarks and insights that have greatly improved the paper.

REFERENCES

AKRITAS, M. G. (1986). Bootstrapping the Kaplan–Meier estimator. *J. Amer. Statist. Assoc.* **81** 1032–1037.

ANDERSEN, P. K. and BORGAN, Ø. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.* **12** 97–158.

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

ARCONES, M. A. and GINÉ, E. (1991). On the bootstrap of U and V statistics. Preprint, Univ. Connecticut.

BABU, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā Ser. A* **46** 86–93.

BICKEL, P. J. and DOKSUM, K. A. (1977). *Mathematical Statistics*. Holden-Day, San Francisco.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BLUM, J. and SUSARLA, V. (1977). On the posterior distribution of a Dirichlet process given randomly right censored data. *Stochastic Process. Appl.* **5** 207–211.

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.

COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

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

DOKSUM, K. A. and LO, A. Y. (1990). Consistent and robust Bayes procedures for location based on partial information. *Ann. Statist.* **18** 443–453.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76** 312–319.

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.

FILIPPOVA, A. A. (1961). Mises' Theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theor. Probab. Appl.* **7** 24–57.

GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52** 203–223.

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Math. Centre Tract* **124**. Math. Centrum, Amsterdam.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

GILL, R. D. (1989). Non- and semiparametric maximum likelihood estimation and the von Mises method (part I, with discussion). *Scand. J. Statist.* **16** 97–128.

HALL, W. L. and WELLNER, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67** 133–143.

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

HJORT, N. L. (1991). Bayesian and empirical Bayesian bootstrapping. Preprint, Univ. Oslo.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KINGMAN, J. F. K. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B* **37** 1–22.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LO, A. Y. (1986). Bayesian statistical inference for sampling a finite population. *Ann. Statist.* **14** 1226–1233.

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. (1991). Bayesian bootstrap clones and a biometry function. *Sanhkyā Ser. A* **53** 320–333.

LO, A. Y. and WENG, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41** 227–245.

MEEDEN, G., GHOSH, M., SRINIVASAN, C. and VARDEMAN, S. (1989). The admissibility of the Kaplan–Meier and other maximum likelihood estimators in the presence of censoring. *Ann. Statist.* **17** 1509–1531.

NAIR, V. N. (1984). Confidence bands for survival functions with censored data: A comparative study. *Technometrics* **26** 265–275.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286.

REID, N. (1981). Estimating the median survival time. *Biometrika* **68** 601–608.

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

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

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

SUSARLA, V. and VAN RYZIN, J. (1978). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. *Ann. Statist.* **6** 755–768.

WENG, C. S. (1988). On some second order asymptotic properties of the Bayesian bootstrap. Ph.D. dissertation, Dept. Statistics, SUNY, Buffalo.

WENG, C. S. (1989). On a second order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17** 705–710.

DEPARTMENT OF STATISTICS
SCHOOL OF MEDICINE AND
BIOMEDICAL SCIENCES
SUNY AT BUFFALO
BUFFALO, NEW YORK 14214