The Annals of Statistics
1997, Vol. 25, No. 4, 1595–1621

A STUDY OF A CLASS OF WEIGHTED BOOTSTRAPS FOR CENSORED DATA

BY LANCELOT F. JAMES

Johns Hopkins University

Edgeworth expansions are derived for a class of weighted bootstrap methods for the Kaplan–Meier and Nelson–Aalen estimates using the methods contained in the monograph by Barbe and Bertail. Von Mises representations up to the third order are established for the weighted bootstrap versions of these estimators. It is shown that there exists weights which outperform Efron's bootstrap method in terms of coverage accuracy. Moreover, it is shown that this holds for a particular choice of gamma weights which are very easy to use in practice. The general weighting schemes are also useful in approximating the posterior distribution of a survival function with respect to mixtures of beta-neutral process priors.

1. Introduction. Lo (1993a) introduced the idea of a censored data Bayesian bootstrap. This is the direct analogue to Rubin's (1981) Bayesian bootstrap (BB) in the complete data case and provides an alternative to Efron's (1981) bootstrap method for the Kaplan–Meier estimator. James (1993) studied a generalization of this idea to the case of arbitrary i.i.d. weights. He proved the consistency result, under a general random censoring model, for this class of weights using martingale techniques in the spirit of Gill (1980). From the work of Mason and Newton (1992) [see also Praestgaard and Wellner (1993)] it is perhaps not difficult to see that one may further define the notion of a generalized exchangeably weighted Kaplan–Meier bootstrap estimate; we will mention this briefly in the coming paragraphs.

Perhaps a more interesting question, and the main focus of this paper, is to determine the proper choice of weights for a class of weighted bootstrapped Kaplan–Meier estimators and related functionals. The recent work of Barbe and Bertail (1995) involving Edgeworth expansions for a class of weighted bootstrap versions of general von Mises differentiable functionals provides the basis for this study. Prior to their work, most of the results on Edgeworth expansions for the weighted bootstrap has been relegated to the case of the sample mean. Related works include papers by Weng (1989), Haeusler, Mason and Newton (1992), Lo (1993b), Hall and Mammen (1994) and Guillou (1995). Their results indicated that one could indeed choose weights which

Received September 1995; revised October 1996.

[1]This work was initiated while the author was visiting the INRA and CREST in Paris, France, during the latter half of 1994.

AMS 1991 *subject classifications*. Primary 62G09; secondary 62G20, 62G30, 60F17.

Key words and phrases. Edgeworth expansions, bootstrap, weighted bootstrap, Kaplan–Meier, von Mises differentiable functions, mixtures of beta-neutral priors and posterior distributions.

are as accurate as the classical bootstrap in approximating the sampling distribution of the sample mean. Weng (1989) and Lo (1933b) also looked explicitly at the case for approximating the posterior distribution of the mean. Shao and Tu (1995) give a survey of weighted bootstrap methods in other settings as well. However, it seems that Haeusler, in an unpublished work, was the first to point out that in the case of the mean one could choose weights which were superior to the “classical bootstrap” in terms of coverage accuracy.

Barbe and Bertail’s (1995) recent monograph investigates the properties of the weighted bootstrap for general von Mises functionals. Among other things, they extend the results of Hauesler to the case of nonlinear functionals $T(P)$. This, along with Mason and Newton (1991) and Praestgaard and Wellner (1993) provides a basis for study of the weighted bootstrap. In general their work compares the performance of the weighted bootstrap against the bootstrap in the regular case using several different criteria. To illustrate we present an excerpt from their monograph.

> However, the choice of weights depends essentially on what one considers to be important. Accuracy of the estimation of the entire distribution of the statistic, accuracy of a confidence interval (related to coverage accuracy), accuracy in a large deviation sense, accuracy for a finite sample size? Some of the criteria may not be satisfied at the same time. [Barbe and Bertail (1995)]

Their results seem to indicate that the area where the weighted bootstrap clearly performs better than the bootstrap is in terms of coverage accuracy. The authors outline methods that must be tailored to each specific situation. One needs to investigate individual functionals of interest to ascertain what are the best choice of weights for that problem. The task involves an investigation into some aspects of the underlying structure of the functional of interest, independent of their work.

Here, we apply these techniques to the case of the Kaplan–Meier and Nelson–Aalen estimators. We shall mainly concentrate on performance in terms of coverage accuracy, although we also address accuracy in terms of approximating the entire distribution. In order to use their framework one needs to verify the validity of the third-order Edgeworth expansion of the sampling distribution of the statistics of interest, in this case the Nelson–Aalen and Kaplan–Meier estimators. We verify these in both of the cases where the variance is assumed known and unknown. We also look at the estimator $e^{-\hat{\Lambda}(t)}$. To establish this and justify the analogous results for the weighted bootstrap, one develops a third-order von Mises expansion for the functional of interest $T(P)$, where P is the underlying distribution. We compute the first three canonical gradients of the functionals associated with the Kaplan–Meier and Nelson–Aalen estimators. These just correspond to the first three Gateaux derivatives of the functional $T(P)$ in the respective cases, and yield the proper form of the expansions. The control of the remainder in the von Mises expansion which holds uniformly over the weighted bootstrap empirical measures, P_w, and the usual empirical measure, P_n, is facilitated by employing a distance proposed in Barbe and Bertail (1995). This allows for Frechet differentiability at a fixed point P for a wider class of functionals than the Kolmogorov distance, while having many of the nice features of the Kolmogorov distance. We do not, however, require the full notion of Frechet differentiability since we only need it to hold at a suitable P (which we may perhaps choose to be continuous) uniformly over a suitable class of empirical measures, P_w, and P_n. That is, the expansions need to be valid for particular functionals of the form $T(P) - T(P_w)$ and $T(P) - T(P_n)$, with proper control on the remainder terms depending upon the underlying structure of the functional in question and the choice of weights.

Once the von Mises expansions are established we are able to modify the arguments used in Lai and Wang (1993) [see also Gross and Lai (1996)] to obtain the validity of the Edgeworth expansions up to the third order. The work involves a careful orchestration of ideas from martingale theory, Edgeworth expansion, empirical process theory and other ideas, and is as much an investigation about the underlying structure of the Kaplan–Meier estimator as it is about the weighted bootstrap. Most importantly, we are able to identify a class of weights which are very simple to use, familiar and superior in terms of coverage accuracy to Efron's (1981) censored data bootstrap. That is, we identify weights which in terms of coverage are accurate up to $o(n^{-1})$ instead of $O(n^{-1})$ in the classical bootstrap theory. In the case where the variance is assumed known, a second-order expansion for the Kaplan–Meier estimator and hazard estimator was done by M. N. Chang (1991) and a third-order expansion for the Nelson–Aalen estimator was shown to be valid by Lai and Wang (1993). Lai and Wang (1993) also developed a second-order expansion for the bootstrap analogue of the Nelson–Aalen estimator. Recently, Chen and Lo (1996) examined the second-order properties of the studentized version of the Kaplan–Meier, Nelson–Aalen, and moment estimators, and their corresponding classical bootstrap statistics. Other related results are given in Burr (1994) and Babu (1991). In an effort not to inundate the reader with technical details, the main results are presented in Section 2. Nevertheless, it is believed that the material in the other sections may be of independent interest to the reader. In Section 1 a description of the model is presented and some remarks on consistency are given. Section 2.1 gives a brief discussion on the third-order Edgeworth expansions for the sampling distribution of the functionals considered. In Section 2.2, coverage probabilities associated with the weighted bootstraps are presented. In Section 2.3, an easily generated class of weights is suggested which provides better coverage accuracy than Efron's classical bootstrap. This should be of particular interest to the potential user. In Section 2.4, second-order Edgeworth expansions are given for the weighted bootstrap; here one can identify weights which are as accurate as the classical bootstrap in terms of approximating the entire sampling distribution. One can also identify weights which are second-order accurate in terms of approximating posterior quantities if one is able to obtain the corresponding expansions for the posterior distributions. The works of Lo (1993a), James (1993) and Brunner and Lo (1996) show that these bootstrap schemes are consistent methods in approximating the posterior distribution of the survival function with respect to mixtures of beta-neutral process priors. This suggests that the weighted bootstrap in this setting is a viable alternative to the sampling schemes suggested in Doss (1994) and Damien, Laud and Smith (1996). In Section 3, third-order von Mises expansions are derived. Weights are identified which give the proper rate for the remainder. These terms can then be ignored in the Edgeworth expansions. In Section 4, validity of the third-order Edgeworth expansions for the sampling distribution of the functionals considered is discussed. In Section 5, canonical gradients associated with the functionals considered here are given. The first canonical gradient $T^{(1)}(x, P)$ is just the influence function associated with $T(P)$. Here we shall need also second- and third-order canonical gradients. These play a crucial role at every level in obtaining the main results in this paper. These appear throughout the presentation and we ask the reader to refer to Section 5 for the explicit forms. In Section 6, explicit forms of the third-order Edgeworth expansions of the sampling distributions are given.

1.1. *Preliminaries.* Let $T_1, T_2, \dots, T_n$ be i.i.d. survival times with continuous survival function $S = 1 - F$ and $C_1, C_2, \dots, C_n$ be independent censoring times with d.f. $G_i(c)$. In the censoring set-up, we observe only the pair $Y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$, which denotes whether an observation has been censored or not. Let $(y_1, \delta_1), (y_2, \delta_2), \dots, (y_n, \delta_n)$ denote the observed data points and $t(1) < t(2) < \cdots < t(k)$ be the k distinct death times. Now define the death set and risk set as follows for $j = 1, \dots, k$:

$$D(j) = \{i: y_i = t(j), \delta_i = 1\} \tag{1.1}$$

and

$$R(j) = \{i: y_i \ge t(j)\}. \tag{1.2}$$

The Kaplan–Meier estimator is often expressed as

$$\hat{S}(t) = \prod_{s \le t} \left(1 - \Delta\hat{\Lambda}(s)\right), \tag{1.3}$$

where $\hat{\Lambda}$ is the well-known Nelson–Aalen estimate for the cumulative hazard Λ. Alternatively, since

$$\hat{\Lambda}(t) = \int_0^t \frac{d\hat{N}(s)}{\hat{Y}(s)} = \sum_{j:\, t(j) \le t} \frac{\sum_{q \in D(j)} 1}{\sum_{q \in R(j)} 1}, \tag{1.4}$$

this indicates that the Kaplan–Meier estimator may be written as follows:

$$\hat{S}(t) = \prod_{j:\, t(j) \le t} \left(1 - \frac{\sum_{q \in D(j)} 1}{\sum_{q \in R(j)} 1}\right). \tag{1.5}$$

Typically one uses the notation, d_j and r_j, where $\sum_{q \in D(j)} 1 = d_j$ and $\sum_{q \in R(j)} 1 = r_j$. [See Andersen, Borgan, Gill and Keiding (1993) and Fleming and Harrington (1991) for more details.] Now, one may express the bootstrap version of these functionals in varying degrees of generality as follows. First, Efron's (1981) censored data bootstrap (CDB),

$$K^*(t) = \prod_{j:\, t(j) \le t} \left(1 - \frac{\sum_{q \in D^*(j)} 1}{\sum_{q \in R^*(j)} 1} \right), \tag{1.6}$$

where D^* and R^* are the death and risk sets for the resampled data points. Now let Z_i, $i = 1, \dots, n$, be n i.i.d. nonnegative r.v.'s with $E[Z] = \mu$, and $\mathrm{Var}(Z) = \sigma^2$, and replace the 1's in the above expression to obtain

$$S^*(t) = \prod_{j:\, t(j) \le t} \left(1 - \frac{\sum_{q \in D(j)} Z_q}{\sum_{q \in R(j)} Z_q} \right). \tag{1.7}$$

Lo (1993a), using exp(1) Z_i's, defined a censored data Bayesian bootstrap (CDBB) which is the direct analog to Rubin's bootstrap scheme in the complete data situation. James (1993), using the arbitrary nonnegative Z_i's above, examined the first-order properties of the class of random weighted bootstraps for the Kaplan–Meier censored data Bayesian bootstrap clones (CDBBC) which may be thought of as the direct analogue to Lo's (1991) Bayesian bootstrap clones (BBC). Finally, let $W_{i:n}$, $i = 1, \dots, n$, be general exchangeable weights, where $\sum_{j=1}^n W_{j:n} = 1$, as in Mason and Newton (1992) [see also Praestgaard and Wellner (1993)]. One may define a generally exchangeable weighted bootstrap scheme for the Kaplan–Meier estimator and related functionals as follows:

$$S_w(t) = \prod_{j:\, t(j) \le t} \left(1 - \frac{\sum_{q \in D(j)} W_{q:n}}{\sum_{q \in R(j)} W_{q:n}} \right). \tag{1.8}$$

The functionals K^* and S^* are special cases of S_w. The weighted bootstrap cumulative hazard $\Lambda_w(t)$ is defined analogously. Now let $\{\mathscr{Y}_n,\ n \ge 1\}$ be a sequence of distribution functions which may depend on the sample $(y_1, \delta_1), \dots, (y_n, \delta_n)$. Then, define a triangular array $\{Y_{i:n}\colon 1 \le i \le n,\ n \ge 1\}$ of r.v.'s such that each row is an i.i.d. vector with distribution $\mathscr{Y}_n$. Further, let $n\overline{Y}_n = Y_{1:n} + \cdots + Y_{n:n}$ and define BBC-type weights [see Lo (1991)],

$$W_{i:n} = \frac{Y_{i:n}}{\sum_{j=1}^n Y_{j:n}}. \tag{1.9}$$

If we choose $Y_{i:n} = Z_i$, then we have S^*. The BBC-type weights will be the main focus of discussion in the coming sections unless otherwise specified. First, however, the question of whether S_w enjoys the same property of consistency as K^* and S^* will be briefly addressed. That is, does the process have the same (conditional) limiting distribution as the sampling distribution of the Kaplan–Meier estimator? [See Breslow and Crowley (1974); see also Gill (1980).] That question was answered for K^* by Akritas (1986), using martingale techniques à la Gill (1980), and for S^* by James (1993), who showed that S^* also enjoys a martingale property which is induced by a sigma-field generated by the weights. We present, for continuity, the following result from that work. Let

$$Z_n^*(t) = \sqrt{n}\left(\frac{S^*(t)}{\hat{S}(t)} - 1\right),$$

where $Z_n^*(t) \equiv 0$ if $\hat{S}(t) = 0$. Define $\Delta K(s) = K(s) - K(s-)$ and let $D[0, b]$ be the space of cadlag (right continuous with left-hand limits) functions equipped with the uniform metric and projection sigma field. Let $\{W(s); s \geq 0\}$ be a standard Brownian motion, and for any (sub)distribution K on the line, let $\tau_K = \sup\{t: K(t) < 1\} \leq \infty$.

ASSUMPTION 1.1. 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) $$\sup_{0 \leq t \leq b} \left|\hat{S}(t) - S_0(t)\right| \to 0$$

and

(ii) $$\sup_{0 \leq t \leq b} \left|\frac{\hat{Y}(t)}{n} - \{1 - H_0(t)\}\right| \to 0.$$

Now we define the following increasing sigma fields, for $t(j-1) \leq t \leq t(j)$, $j = 1, \dots, k+1$, where $t(0) = 0$ and $t(k+1) = b$. Let

$$\mathscr{F}_t = \sigma\{Z_q: q \in R(1) - R(j)\}$$

$$\mathscr{O}_t = \sigma\{\text{order statistics generated by } Z_q\text{'s: } q \in R(j)\}.$$

Define

$$M^*(t) = N^*(t) - \int_0^t Y^*(s)\, d\hat{\Lambda}(s),$$

where $N^*(t) = \sum_{i=1}^n Z_i I(y_i \leq s,\ \delta_i = 1)$ and $Y^*(s) = \sum_{i=1}^n Z_i I(y_i \geq s)$. Then under Assumption 1.1, we have the theorem.

THEOREM 1.1 [James (1993)]. J1. *$\{Z_n^*(t), \mathscr{F}_t \vee \mathscr{O}_t\colon\ t \in [0, b]\}$ is a right continuous martingale with compensator*

$$\langle Z_n^*, Z_n^*\rangle(t) = n\int_0^t \left(\frac{S^*(t-)}{\hat{S}(t-)\left[1 - \Delta\hat{\Lambda}(s)\right]Y^*(s)}\right)^2 \langle M^*, M^*\rangle(ds),$$

where

$$\langle M^*, M^*\rangle(t) = \sum_{j: t(j) \leq t} d_j\left(1 - \frac{d_j}{r_j}\right)\left\{\frac{r_j}{r_j - 1}\right\}\left[\frac{\sum_{q \in R(j)} Z_q^2}{r_j} - \left(\frac{\sum_{q \in R(j)} Z_q}{r_j}\right)^2\right]$$

J2. *Let* $\rho = \mu/\sigma$ *then*

$$\mathscr{L}\{\rho Z_n^*(\cdot)\ (\mathbf{y}, \delta)\} \to \mathscr{L}\{W(C_0(\cdot))\} \quad in\, D[0, b],$$

where

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

REMARK 1.1. The martingale property in statement J1 is induced by the weights conditional on the data. The compensator above may be used as a bootstrapped variance estimator. The statement in J2 shows that BBC weights with finite second moments, subject to a scale factor modification ρ, are equivalent in the first-order sense to Efron's classical bootstrap scheme for the Kaplan–Meier estimator under the general random censorship model.

Gill (1994) [see also van der Laan (1993)], under the more restrictive assumption that the C_i's are i.i.d. (which will be the case throughout the rest of this paper), showed that one can get a rather elegant and essentially two-line proof of the consistency result for Efron's K^* using modern empirical process techniques [see Giné and Zinn (1990)] and employing an identity by van der Laan (1993) for linear convex models applied to the Kaplan–Meier and its "classical" bootstrap analogue. We show how one may get the consistency result for S_w by mimicking those arguments. Let $X = (Y, \delta)$ with distribution P_{FG} (hereafter denoted as P) and define the following empirical measures:

$$P_n = \sum_{i=1}^{n} \frac{1}{n} \delta_{X_i} \tag{1.10}$$

and

$$P_w = \sum_{i=1}^{n} W_{i,n} \delta_{X_i}, \tag{1.11}$$

where δ_x is the Dirac measure. Now, similarly to van der Laan [(1993), page 30], define

$$\begin{aligned} \tilde{I}_1(x) &= \tilde{I}(\hat{S}, G_w, t)(y, \delta) \\ &= -\hat{S}(t) \int_0^t \frac{I(y \in ds,\ \delta = 1) - I(y \geq s)\, d\hat{\Lambda}(s)}{\hat{S}(s)\{1 - G_w(s-)\}}. \end{aligned} \tag{1.12}$$

Alternatively one may also define

$$\begin{aligned} \tilde{I}_2(x) &= \tilde{I}(S_w, \hat{G}, t)(y, \delta) \\ &= S_w(t) \int_0^t \frac{I(y \in ds,\ \delta = 1) - I(y \geq s)\, d\Lambda_w(s)}{S_w(s)\{1 - \hat{G}(s-)\}}, \end{aligned} \tag{1.13}$$

where $1 - \hat{G}(t-) = \hat{Y}(t)/\hat{S}(t)$ and $1 - G_w(t-) = Y_w(t)/S_w(t)$ is its weighted bootstrap analogue. Here $\hat{Y}(t) = P_n(I(y \geq t))$ and $Y_w(t) = P_w(I(y \geq t))$. We then have that

$$\hat{S}(t) - S_w(t) = (P_n - P_w)(\tilde{I}_1) \tag{1.14}$$

with $P_n(\tilde{I}_1) = 0$ and

$$\hat{S}(t) - S_w(t) = (P_n - P_w)(-\tilde{I}_2), \tag{1.15}$$

with $P_w(\tilde{I}_2) = 0$. The idea is then to verify that a class of functions containing $\tilde{I}_2$ or $\tilde{I}_2$ (where t is allowed to vary) is Donsker. The consistency result is then obtained by applying the work of Praestgaard and Wellner (1993) in the place of Giné and Zinn (1990), that is, for weights which satisfy conditions A1–A5 in Praestgaard and Wellner (1993). We shall assume these conditions hold throughout our presentation.

2. Edgeworth expansions and the choice of weights.

Define

$$H_n(x, T(P_n), T(P)) = P\{n^{1/2}(T(P_n) - T(P))/S(P) \leq x\} \tag{2.1}$$

and

$$K_n(x, T(P), T(P_n)) = P\{n^{1/2}(T(P_n) - T(P))/S(P_n) \leq x\}, \tag{2.2}$$

where $T(P)$ is the functional of interest and $S^2(P) = E_P[T^{(1)}(X, P)^2]$. Thus we may then write the sampling distributions associated with the Kaplan–Meier, Nelson–Aalen and $e^{-\hat{\Lambda}(t)}$ as follows:

$$\begin{aligned}
H_n(x, \hat{S}(t), S(t)) &= P\left\{n^{1/2}\frac{\hat{S}(t) - S(t)}{S(t)C^{1/2}(t)} \leq x\right\},\\
K_n(x, \hat{S}(t), S(t)) &= P\left\{n^{1/2}\frac{\hat{S}(t) - S(t)}{\hat{S}(t)\hat{C}^{1/2}(t)} \leq x\right\},\\
H_n(x, \hat{\Lambda}(t), \Lambda(t)) &= P\left\{n^{1/2}\frac{\hat{\Lambda}(t) - \Lambda(t)}{C^{1/2}(t)} \leq x\right\},\\
K_n(x, \hat{\Lambda}(t), \Lambda(t)) &= P\left\{n^{1/2}\frac{\hat{\Lambda}(t) - \Lambda(t)}{\hat{C}^{1/2}(t)} \leq x\right\}
\end{aligned} \tag{2.3}$$

and

$$H_n(x, e^{-\hat{\Lambda}(t)}, S(t)) = P\left\{n^{1/2}\frac{e^{-\hat{\Lambda}(t)} - S(t)}{S(t)C^{1/2}(t)} \leq x\right\}.$$

Here $S^2(P) = S^2(t)C(t)$, for the Kaplan–Meier estimator, where $C(t) = \int_0^t (d\Lambda(s)/Y(s))$. For the Nelson–Aalen $S^2(P)$ is just $C(t) = \int_0^t (d\Lambda(s)/Y(s))$ and for $e^{-\hat{\Lambda}(t)}$ we use the same as that for the Kaplan–Meier estimator. We also define $A_j(t) = \int_0^t (d\Lambda(s)/Y^{j-1}(s))$ for $j = 3, 4$. The weighted bootstrap analogues may be defined simply as

$$H_n(x, T(P_w), T(P_n)) = P^*\{n^{1/2}(T(P_w) - T(P_n))/(\sigma_{W,n} S(P_n)) \le x\} \tag{2.4}$$

and

$$K_n(x, T(P_w), T(P_n)) = P^*\{n^{1/2}(T(P_w) - T(P_n))/(\sigma_{W,n} S(P_w)) \le x\}, \tag{2.5}$$

where P^* denotes the conditional distribution with respect to the data and $\sigma^2_{W,n} = \mathrm{var}(nW_{1:n})$. Note that in (2.5) we are able to use $S(P_w)$ as the analogue to $S(P_n)$ because in the case of the functionals considered here $S(P)$ has an explicit functional form. In the case of the Kaplan–Meier estimator, we use $S^2(P_w) = S^2_w(t)C_w(t)$, where $C_w(t) = \int_0^t (d\Lambda_w(s)/Y_w(s))$. We derive the form of the expansions using the work of Bertail (1992) [see also Barbe and Bertail (1995)] and justify these expansions by results contained in Lai and Wang (1993) which we discuss in Section 4. We give exact forms of the Edgeworth expansions for the sampling distributions of the statistics in Section 6. We assume throughout the following conditions as in Lai and Wang (1993). That is,

$$F(t) > 0 \quad \text{and} \quad 1 - H(t-) = E\left[I_{(Y_1 \ge t)}\right] > 0. \tag{2.6}$$

PROPOSITION 2.1. *Under* (2.6), *the Edgeworth expansions for* $H_n(x, \hat{S}(t), S(t))$, $K_n(x, \hat{S}(t), S(t))$, $H_n(x, \hat{\Lambda}(t), \Lambda(t))$, $K_n(x, \hat{\Lambda}(t), \Lambda(t))$ *and* $H_n(x, e^{-\hat{\Lambda}(t)}, S(t))$ *are valid up to* $o_p(n^{-1})$ *and hold uniformly in* x.

2.1. *Coverage probability.* We now turn to the case of coverage probability where it will be seen upon further reading that one may achieve better coverage accuracy than one can obtain using the "classical bootstrap." We assume that the weights considered here are continuous and of the form in (1.9).

In the general setup, let $\gamma_{\alpha,w,n}$ and $\zeta_{\alpha,w,n}$ be the αth quantile of the weighted bootstrap analogues in the case where the variance is assumed known and the studentized case, respectively. That is,

$$P^*\{n^{1/2}(T(P_w) - T(P_n))/(\sigma_{W,n} S(P_n)) \le \gamma_{\alpha,w,n}\} = \alpha \tag{2.7}$$

and

$$P^*\{n^{1/2}(T(P_w) - T(P_n))/(\sigma_{W,n} S(P_w)) \le \zeta_{\alpha,w,n}\} = \alpha. \tag{2.8}$$

The idea is now to choose weights such that

$$H(\gamma_{\alpha,w,n}, T(P_n), T(P)) = P\{n^{1/2}(T(P_n) - T(P))/S(P) \le \gamma_{\alpha,w,n}\} \tag{2.9}$$

and

$$K(\zeta_{\alpha,w,n}, T(P_n), T(P)) = P\{n^{1/2}(T(P_n) - T(P))/S(P_n) \le \zeta_{\alpha,w,n}\} \tag{2.10}$$

are as close to α as possible. In order to do this, one needs to derive the

Cornish–Fisher expansions for the weighted distributions. These expansions will be valid under the following conditions.

W1. Valid third-order Edgeworth expansions for the sampling distributions of the original. See Section 4.

W2. Valid third-order von Mises expansions of $T(P_w) - T(P)$ with remainders of $o_p(n^{-1})$ for the particular weights considered (see Section 3 for more details).

W3. $\limsup_{n\to\infty} \limsup_{t\to\infty} |E \exp(itY_{1,n})| < 1$.

W4. Moment conditions to be specified below in Proposition 2.2.

REMARK 2.1. Since we consider only continuous only weights in this section, condition W3 is automatically satisfied. For the studentized cases, one also needs to be able to satisfy W2 for $f(Q) = (T(Q) - T(P))/S(Q)$ for the appropriate measure Q about $f(P) = 0$. This requires knowledge of the form of the first two gradients of $S^2(P)$ (see Section 5). Here "f" plays exactly the role of "T" throughout. In general this presents an added difficulty but for the functionals considered here it follows directly from W2 and we omit the details.

Let

$$V_3(t) = \frac{\left[\int_0^t (d\Lambda(s)/Y^2(s)) - (3/2)C^2(t)\right]}{C^{3/2}(t)}$$

and

$$V_4(t) = \frac{\left[\int_0^t (d\Lambda(s)/Y^3(s)) - 4\int_0^t C(s)(d\Lambda(s)/Y^2(s)) + 2C^3(t)\right]}{C^2(t)}.$$

PROPOSITION 2.2. *Let the moments of $Y_{i:n}$ be such that*

$$\begin{aligned} \sigma_{Y,n} &= 1 + gn^{-1/2}, \\ \beta_{Y,n} &= 1 + hn^{-1/2} \end{aligned} \tag{2.11}$$

where $\sigma^2_{Y,n} = \mathrm{var}(Y_{1,n})$ and $\beta_{Y,n}$ denotes the skewness. We denote the kurtosis by $K_{Y,n}$. As explained below, "h" and "g" will serve as corrective factors when determining coverage. The αth quantile of the standard normal distribution is written as z_α. Under conditions W1–W4 the coverage accuracy for the case of the weights $Y_{i:n}$ above are given for the Kaplan–Meier and Nelson–Aalen estimators as follows:

$$\begin{aligned} H_n\big((\gamma_{\alpha,w,n}, \hat{S}(t), S(t)\big) &= \alpha - \frac{h}{6n} V_3(t)\phi(z_\alpha)(z_\alpha^2 - 1) \\ &\quad + \frac{(K_{Y,n} - 1)}{24n} V_4(t)\phi(z_\alpha)(z_\alpha^3 - 3z_\alpha) + o(n^{-1}). \end{aligned}$$

It is interesting to note that in this case the terms above only depend upon the first gradient (influence function):

$$
\begin{aligned}
K\big(\zeta_{\alpha,w,n}, \hat{S}(t), S(t)\big) = \alpha &- \frac{h}{6n} V_3(t)\phi(z_\alpha)\big(z_\alpha^2 - 1\big) \\
&+ \frac{g}{n} V_3(t)\phi(z_\alpha)\big(1/2 + (1/2)\big(z_\alpha^2 - 1\big)\big) \\
&+ \frac{(K_{Y,n} - 1)}{24n} V_4(t)\phi(z_\alpha)\big(z_\alpha^3 - 3z_\alpha\big) + o(n^{-1}), \\
H\big(\gamma_{\alpha,w,n}, \hat{\Lambda}(t), \Lambda(t)\big) = \alpha &+ \frac{h}{6n} V_3(t)\phi(z_\alpha)\big(z_\alpha^2 - 1\big) \\
&+ \frac{3g}{6n} C^{1/2}(t)\phi(z_\alpha)\big(z_\alpha^2 - 1\big) \\
&+ \frac{(K_{Y,n} - 1)}{24n} V_4(t)\phi(z_\alpha)\big(z_\alpha^3 - 3z_\alpha\big) + o(n^{-1}), \\
K\big(\zeta_{\alpha,w,n}, \hat{\Lambda}(t), \Lambda(t)\big) = \alpha &+ \frac{h}{6n} V_3(t)\phi(z_\alpha)\big(z_\alpha^2 - 1\big) \\
&- \frac{g}{2n} \frac{\big[A_3(t) + (1/2)C^2(t)\big]}{C^{3/2}(t)} \phi(z_\alpha) \\
&- \frac{3g}{6n} \frac{\big[A_3(t) - (1/2)C^2(t)\big]}{C^{3/2}(t)} \phi(z_\alpha)\big(z_\alpha^2 - 1\big) \\
&+ \frac{(K_{Y,n} - 1)}{24n} V_4(t)\phi(z_\alpha)\big(z_\alpha^3 - 3z_\alpha\big) + o(n^{-1}).
\end{aligned}
$$

2.2. *Choice of weights*. We can see from the above results concerning coverage probability that one can achieve greater accuracy, up to $o(n^{-1})$, by choosing "h" and "g" such that the n^{-1} terms in the above quantities are cancelled. This implies that "h" and "g" will depend upon the particular α-level as well as the cumulants of the respective functionals in an obvious way. The task then remains to generate appropriate weights which satisfy the moment conditions mentioned previously. We propose a simple class of weights which satisfy those conditions and may be generated quite easily using the standard computational packages; for other choices of weights see Barbe and Bertail [(1995), page 32].

Let $Y_{i:n}$ $i = 1, \ldots, n$ be i.i.d. gamma r.v.'s with parameters λ and ν [denoted $G(\lambda, v)$]. Now choose λ and v such that

$$
\begin{aligned}
\lambda &= 4/\big(1 + hn^{-1/2}\big)^2, \\
v^2 &= 4/\big(1 + gn^{-1/2}\big)^2\big(1 + hn^{-1/2}\big)^2
\end{aligned}
\tag{2.12}
$$

and for further convenience we may choose "g" to be zero. By choosing g to be zero we erase the effect of the nonlinear terms in the expansion. This implies that one can get coverage with error of rate smaller than n^{-1} by choosing "h" which only depends upon the influence function in the following general sense, provided that an expansion exists and $E(T^{(1)}(X,P)^3)$ is nonzero,

$$(2.13)\qquad h(\alpha) = -\frac{7}{8}\frac{E\big(T^{(1)}(X,P)^4\big)}{E\big(T^{(1)}(X,P)^3\big)\big(E\big(T^{(1)}(X,P)^2\big)\big)^{1/2}}\frac{z_\alpha^3 - 3z_\alpha}{z_\alpha^2 - 1};$$

for general weights we replace $-\frac{7}{8}$ by $-\frac{1}{4}(K_{y,n} - 1)$. Here we approximate $K_{y,n}$ by $\frac{9}{2}$. In practice we must replace the unknown quantities in the above expression by suitable estimates (this should not change the level of accuracy). We denote h by $h(\alpha)$, which indicates that one needs to generate different weights for different α-levels. This amounts to consulting a normal table and then generating n gamma weights, adjusting $h(\alpha)$ accordingly. In the case of the functionals discussed in this paper, we may choose for the Kaplan–Meier estimator (whether studentized or not),

$$(2.14)\qquad h_S(\alpha) = \frac{7}{8}\frac{\hat{V}_4(t)}{\hat{V}_3(t)}\frac{z_\alpha^3 - 3z_\alpha}{z_\alpha^2 - 1}$$

and for the case of the Nelson–Aalen estimate (whether studentized or not),

$$(2.15)\qquad h_\Lambda(\alpha) = -h_S(\alpha).$$

However, we point out that for large values of the ratio $(z_\alpha^3 - 3z_\alpha)/(z_\alpha^2 - 1)$ relative to n, there may be large departures from (2.11). An ad hoc method may be to modify the estimators in (2.14) to further dampen the effect of $(z_\alpha^3 - 3z_\alpha)/(z_\alpha^2 - 1)$, or one may try to find another choice of weights which works well in the area where this ratio is large. Another option, in the nonstudentized case, is to use the variance estimator (S_n^2) suggested in Barbe and Bertail (1995) in place of the unknown $S^2(P)$. Here for the studentized estimators, we find a more satisfactory result by choosing "g" different from zero. That is, for the studentized Kaplan–Meier estimator, we choose

$$(2.16)\qquad h = 3g \quad\text{and}\quad g = \frac{(K_{y,n} - 1)}{12}\frac{\hat{V}_4(t)}{\hat{V}_3(t)}\big(z_\alpha^3 - 3z_\alpha\big),$$

which for the weights considered here results in choosing

$$(2.17)\qquad h_S^{\text{student}}(\alpha) = \frac{7}{8}\frac{\hat{V}_4(t)}{\hat{V}_3(t)}\big(z_\alpha^3 - 3z_\alpha\big).$$

Similarly, for the studentized Nelson–Aalen estimator, we can choose h and g such that

$$(2.18)\qquad h_{\Lambda}^{\text{student}}(\alpha) = \frac{7}{8}\frac{\hat{V}_4(t)}{\hat{V}_3(t)}\frac{\left[A_3(t) - (1/2)C^2(t)\right]}{\left[A_3(t) + (1/2)C^2(t)\right]}\left(z_{\alpha}^3 - 3z_{\alpha}\right).$$

As a general guideline in the studentized cases, one should use $h_S^{\text{student}}(\alpha)$, $h_{\Lambda}^{\text{student}}(\alpha)$ for $z_{\alpha}^2 < 2$ and $h_S(\alpha)$, $h_{\Lambda}(\alpha)$ otherwise.

Further, it is known that the $W_{i:n}$ defined by the $Y_{i:n}$ in this section have beta $(\lambda, \lambda(n-1))$ distributions; of course it may be a bit troublesome to generate directly exchangeable weights of this type. However, upon closer inspection, one may generate directly k (corresponding to the number of distinct "death" times) independent r.v.'s V_j distributed as beta $(\lambda d_j, \lambda(r_j - d_j))$, where $d_j = \sum_{q \in D(j)} 1$ and $r_j = \sum_{q \in R(j)} 1$ for $j = 1, \dots, k$, since in this case,

$$(2.19)\qquad \mathscr{L}\{\Lambda_w(t)\} = \mathscr{L}\left\{\sum_{j:t(j)\le t} V_j\right\}.$$

Then $\operatorname{var}(\sqrt{n}\,\Lambda_w(t)) = n\sum_{j:t(j)\le t}(d_j/r_j)(1-(d_j/r_j))(1/(\lambda r_j + 1))$ and may be used to play the role of $\sigma_{W,n}S(P_n)$; an appropriate weighted studentized version is defined analogously. The fact that gamma weights used in the framework of the survival models discussed here induce independent increment bootstrapped hazard processes was essentially pointed out by Lo (1993a) [see also Hjort (1990)].

REMARK 2.2. The statements above imply that the choice of the scale parameter v in (2.12) in practice (i.e., computationally) is irrelevant provided that one has incorporated the effect of "g" into "h" and adjusts the entire distribution by the scale factor $\sigma_{W,n}$. This is analogous to Lo's (1993b) results in the case of the BBC approximations for the mean, which states that second-order accuracy depends only upon the skewness of the weights and a rescaling of the distribution by multiplication of $\rho = \mu/\sigma$. In this case $\rho = \sqrt{\lambda}$ and $1/(\sigma_{W,n}) = \sqrt{(n\lambda + 1)/(n-1)}$.

2.3. *Weighted bootstrap expansions.* In this section we give the second-order results for the weighted bootstrap expansions. Of interest is which choice of weights approximate the entire sampling distribution of the estimators considered here as well as the bootstrap. We note that under this criterion we only find weights that work as well as Efron's classical scheme, although there are some real advantages in choosing continuous weights. It seems, however, that one can do better if one is interested in approximating posterior quantities. Here the results hold under slightly weaker conditions; that is, we can replace W1 and W2 by W1′ and W2′.

W1′. Valid second-order Edgeworth expansions for the sampling distributions of the original statistics.
W2′. Valid second-order von Mises expansions of $T(P_w) - T(P)$.
W3′. $\limsup_{n\to\infty} \limsup_{t\to\infty} |E\exp(itY_{1,n}| < 1$.
W4′. $\{Y_{i,n}^4: n \geq 1\}$ is uniformly integrable under $\mathscr{Y}_n$.

PROPOSITION 2.3. *Under* W1′–W4′ *we have, with probability* 1 *uniformly in* x, *the validity of the following weighted bootstrap expansions*:

$$H_n\big(x, \hat{S}(t), S_w(t)\big) = \Phi(x) + n^{-1/2}\frac{\beta_{Y,n}}{6}V_3(t)(x^2-1)\phi(x) + o(n^{-1/2}),$$

$$K_n\big(x, \hat{S}(t), S_w(t)\big) = \Phi(x) + n^{-1/2}\frac{\beta_{Y,n}}{6}V_3(t)(x^2-1)\phi(x) - n^{-1/2}\frac{\sigma_{Y,n}}{2}V_3(t)x^2\phi(x) + o(n^{-1/2}),$$

$$H_n\big(x, \hat{\Lambda}(t), \Lambda_w(t)\big) = \Phi(x) - n^{-1/2}\frac{\beta_{Y,n}}{6}V_3(t)(x^2-1)\phi(x) - n^{-1/2}\frac{\sigma_{Y,n}C^{1/2}(t)}{2}(x^2-1)\phi(x) + o(n^{-1/2}),$$

$$\begin{aligned} K_n\big(x, \hat{\Lambda}(t), \Lambda_w(t)\big) = \Phi(x) &+ n^{-1/2}\frac{\sigma_{Y,n}\big[A_3(t) + 1/2C^2(t)\big]}{2C^{3/2}(t)}\phi(x) \\ &-n^{-1/2}\frac{\beta_{Y,n}}{6}V_3(t)(x^2-1)\phi(x) \\ &+n^{-1/2}\frac{\sigma_{Y,n}\big[A_3(t) - 1/2C^2(t)\big]}{2C^{3/2}(t)}(x^2-1)\phi(x) \\ &+o(n^{-1/2}). \end{aligned}$$

Furthermore, upon choosing weights with

$$\sigma_{Y,n} = 1 + O_p(n^{-1/2}) \tag{2.20}$$

and

$$\beta_{Y,n} = 1 + O_p(n^{-1/2}), \tag{2.21}$$

one achieves greater accuracy than the normal approximation and achieves the same level of accuracy as one gets for the case of the classical bootstrap [*the order of accuracy being* $O_p(n^{-1})$]. *That is, for the case of the functionals considered here we have* (*uniformly in* x, *a.s.*)

$$(2.22)\quad \sup_x \left| H_n(x, T(P_n), T(P)) - H_n(x, T(P_n), T(P_w)) \right| = O_p(n^{-1})$$

and

$$(2.23)\quad \sup_x \left| K_n(x, T(P_n), T(P)) - K_n(x, T(P_n), T(P_w)) \right| = O_p(n^{-1}).$$

REMARK 2.3. Curiously, in the case of the Kaplan–Meier estimator (variance known), one only needs to adjust the skewness, $\beta_{Y,n}$, of the weights which is exactly what happens in the case of the mean [see Hauesler, Mason and Newton (1992), Lo (1993b)]. Nevertheless, if one is interested in approximating posterior quantities, as in Lo (1993a) and Wells and Tiwari (1994), then one may need to choose different weights. See Weng (1989), and Lo (1993b) for some ideas on the proper choice of weights. We point out that these bootstraps are direct competitors with other simulation techniques for evaluating posterior distributions such as those in Damien, Laud and Smith (1996). One needs to develop appropriate expansions for the posterior distributions in question and then if possible choose the moments for the weights to match the terms in those expansions. We conjecture that at least for the posterior distribution of the survival estimator with respect to Beta-neutral priors [see, for instance, Lo (1993a), Hjort (1990)] choosing weights with skewness $2 + o_p(1)$ will give the best order of approximation. We note further that, although we have not explicitly shown the validity of a second-order Edgeworth expansion for the classical bootstrap version of the functionals considered here, this is easily obtained using the results in the thesis of Bertail (1992) [see also Helmers (1991), Chen and Lo (1996)].

3. Von Mises representations. One of the key ideas is to obtain a von Mises representation of the functional of interest, say $T(P) - T(Q)$, which holds uniformly, at a suitable fixed P, over the class of measures $Q \in \mathscr{P}$ where $\mathscr{P}$ is required to be a convex class of probability measures containing the Dirac measures. We now establish the following third-order von Mises representation for the Kaplan–Meier and cumulative hazard estimators. The aim is then to show that these representations are valid up to $o_p(n^{-1})$ for the specific measures P_w, P_n, which correspond to the weighted bootstraps and the usual empirical measure. To facilitate this, we employ a distance indexed by a class of functions suggested by Barbe and Bertail (1995) which admits a larger class of Frechet differentiable functionals than the usual Kolmogorov norm. The idea is similar to Dudley (1990). Define, for the functionals considered here,

$$\mathscr{H}_T = \left\{ \left(\left| T^{(1)}(x, P) \right| \vee 1 \right) I_{\{x \in C\}};\, t \le \tau,\, C \in \mathscr{D} \right\},$$

where $T^{(1)}(x,P)$ (we suppress the dependence on t) is the influence function corresponding to $T(P)$ and $\mathscr{D}$ is a sufficiently rich class of V–C subsets of $R^+ \times \{0,1\}$, and define the distance

$$d_{\mathscr{H}}(Q,P) = \sup_{h \in \mathscr{H}} \left| \int h d(Q-P) \right|.$$

REMARK 3.1. In general, to get von Mises expansions valid to the third order, it may be necessary to include $T^{(2)}$ and $T^{(3)}$ in $\mathscr{H}_T$ as well. The notion of using another (pseudo)metric instead of the Kolmogorov distance to gain the Frechet differentiability of a wider class of statistical functionals has been recently discussed in numerous articles by Dudley and others. We cite Dudley (1992), (1994) and Arcones and Giné (1992) as further references.

For the sake of ease of notation, we define, whenever they exist, for any $Q \neq P$ the following:

$$\begin{aligned} \tilde{\Lambda}(t) &= \int I\{s \leq t,\, \delta = 1\} \frac{dQ(s,\delta)}{Q(A_s)}, \\ \tilde{S}(t) &= \prod_{s \leq t} \bigl(1 - \Delta\tilde{\Lambda}(s)\bigr), \end{aligned} \tag{3.1}$$

where $A_s = \{x = (y, \delta'): y \geq s,\ \delta' \in \{0,1\}\}$ [see Pons and Turckheim (1991)].

PROPOSITION 3.1. *Choose τ such that $P(A_\tau) > \eta > 0$; then for all Q such that $Q(A_\tau) > \eta$ and $t < \tau$ we have the following*:

$$\begin{aligned} &\hat{\Lambda}(t) - \Lambda(t) \\ &\quad = \int T^{(1)}(x, \Lambda(t))\, d(Q-P)(x) \\ &\qquad + \tfrac{1}{2} \int T^{(2)}(x, y, \Lambda(t))\, d(Q-P)(y)\, d(Q-P)(x) \\ &\qquad + \tfrac{1}{6} \int T^{(3)}(x, y, z, \Lambda(t))\, d(Q-P)(y)\, d(Q-P)(z)\, d(Q-P)(x) \\ &\qquad + O\bigl(d_{\mathscr{H}_\Lambda}(Q,P)\bigr)^3 \end{aligned}$$

and

$$\begin{aligned} \tilde{S}(t) - S(t) = &\int T^{(1)}(x, K(P)(t))\, d(Q-P)(x) \\ &+ \tfrac{1}{2} \int T^{(2)}(x, y, K(P)(t))\, d(Q-P)(y)\, d(Q-P)(x) \\ &+ \tfrac{1}{6} \int T^{(3)}(x, y, z, K(P)(t))\, d(Q-P)(y)\, d(Q-P)(z) \\ &\quad \times d(Q-P)(x) + O\bigl(d_{\mathscr{H}_\Pi}(Q,P)\bigr)^3 \end{aligned}$$

where $\mathscr{H}_\Lambda$, $\mathscr{H}_\Pi$ *are defined as* $\mathscr{H}_T$ *with* $T^{(1)}(x, P)$ *replaced by* $T^{(1)}(x, \Lambda(t))$ *and* $T^{(1)}(x, K(P)(t))$, *respectively. It is sufficient to use* $\mathscr{H}_\Lambda$ *throughout.*

PROOF. We show how one may obtain the result for $\tilde{S}(t) - S(t)$; the arguments for $\Lambda(t) - \Lambda(t)$ are straightforward. From the Duhamel equations [see, for instance, Gill (1994)], we have

$$\tilde{S}(t) - S(t) = -S(t)\int_0^t \frac{\tilde{S}(s-)}{S(s)} \frac{Y(s)}{\tilde{Y}(s)} \frac{1}{Y(s)}\, dM(s), \tag{3.2}$$

where $Y(s) = P(A_s)$, $\tilde{Y}(s) = Q(A_s)$ and

$$M(t) = \int M_x(t)\, d(Q - P)(x). \tag{3.3}$$

$M_x(t)$ is defined in Section 5. For further illustration, the arguments for the first order result are given. That is,

$$\begin{aligned} \tilde{S}(t) - S(t) &= \int T^{(1)}(x, K(P)(t))\, d(Q-P)(x) + o(d_{\mathscr{H}}(Q,P)) \\ &= -S(t)\int_0^t \frac{1}{Y(s)}\, dM(s) + o(d_{\mathscr{H}}(Q,P)), \end{aligned} \tag{3.4}$$

which indicates that it is necessary to show that

$$\left| \int_0^t \frac{dM(s)}{Y(s)} - \int_0^t \frac{\tilde{S}(s-)}{S(s)} \frac{Y(s)}{\tilde{Y}(s)} \frac{dM(s)}{Y(s)} \right| = o(d_{\mathscr{H}}(Q,P)). \tag{3.5}$$

From (3.2) one can expand $\tilde{S}(s-)/S(s)$ and then combine terms on the l.h.s. of (3.5) to get

$$\left| \int_0^t \left[\frac{(Y(s) - \tilde{Y}(s))}{\tilde{Y}(s)} - \frac{Y(s)}{\tilde{Y}(s)} \int_0^{s-} \frac{\tilde{S}(u-)Y(u)}{S(u)\tilde{Y}(u)} \frac{dM(u)}{Y(u)} \right] \frac{dM(s)}{Y(s)} \right|. \tag{3.6}$$

Notice that for constants K_1 and K_2,

$$\left| \frac{Y(s) - \tilde{Y}(s)}{\tilde{Y}(s)} \right| \le K_1 d_{\mathscr{H}}(Q,P) \tag{3.7}$$

and

$$\left| \frac{Y(s)}{\tilde{Y}(s)} \int_0^{s-} \frac{\tilde{S}(u-)Y(u)}{S(u)\tilde{Y}(u)} \frac{dM(u)}{Y(u)} \right| \le K_2 d_{\mathscr{H}}(Q,P), \tag{3.8}$$

which gives the desired result. The arguments for the third-order result are similar and upon applying the Duhamel equation recursively on $\tilde{S}(s-)/S(s)$ and gathering terms, one arrives at the following:

$$\left|\int_0^t V(s)\frac{dM(s)}{Y(s)}\right| \le KO\big(d_{\mathscr{H}_\Pi}(Q,P)\big)^3, \tag{3.9}$$

where

$$\begin{aligned} V(s) = {} & \frac{\big(\tilde{Y}(s)-Y(s)\big)^3}{Y^2(s)\tilde{Y}(s)} - \frac{\big(\tilde{Y}(s)-Y(s)\big)^2}{Y(s)\tilde{Y}(s)}\int_0^{s-}\frac{dM(u)}{Y(u)} \\ & + \int_0^{s-}\left[1-\frac{\tilde{S}(u-)Y(u)}{S(u)\tilde{Y}(u)} - \int_0^{u-}\frac{dM(v)}{Y(v)} - \frac{\big(\tilde{Y}(u)-Y(u)\big)}{Y(u)}\right]\frac{dM(u)}{Y(u)} \\ & - \frac{\big(\tilde{Y}(s)-Y(s)\big)}{\tilde{Y}(s)}\int_0^{s-}\left[1-\frac{\tilde{S}(u-)Y(u)}{S(u)\tilde{Y}(u)}\right]\frac{dM(u)}{Y(u)}. \end{aligned}$$

Furthermore, since $\mathscr{H}_\Lambda$ and $\mathscr{H}_\Pi$ are Donsker classes [see, for instance, Pollard (1990)] and by the result of Praestgaard and Wellner (1993) we have that for $Q = P_n$, or P_w; that is, for weights which satisfy conditions A1–A5 in Praestgaard and Wellner (1993), that $d_{\mathscr{H}}(Q,P) = O_p(n^{-1/2})$. This implies that these expansions are valid up to $o_p(n^{-1})$ for the classical bootstrap, the Bayesian bootstrap and the weights considered in Section 2, among many others. □

REMARK 3.2. Sufficient almost sure results for remainders of general von Mises functionals were obtained by Barbe and Bertail [(1995), Theorem 2.1] under a weaker (than Donsker) condition on the class of functions. That is, we have $d_{\mathscr{H}}(Q,P) = O_p(n^{-1/2}(\log n)^{1/2})$ a.s. P and $d_{\mathscr{H}}(P_w,P_n) = O_w(n^{-1/2}(\log n)^{1/2})$ a.s. P. This is attractive, since in general one may not be able to verify a Donsker class condition on $\mathscr{H}_T$ where, for instance, one may need to include $T^{(2)}$ and $T^{(3)}$. Their Theorem 2.1 gives us the almost sure results in Section 2. Further, if one is interested in LIL-type results, one may perhaps use the work of Arcones and Giné (1995).

4. Validity of expansions. We now proceed to show how one may verify the necessary technical conditions as outlined in Lai and Wang (1993) to obtain the validity of the third-order Edgeworth expansions of the sampling distributions. There they show that the Edgeworth expansion for the case of the cumulative hazard (assuming a known variance) is valid up to the third order. Their arguments involve an eigenvalue-type condition similar to but more general than that in Bickel, Götze and van Zwet (1986). We believe the results of Götze (1979, 1984) would be of use here as well. As in Lai and

Wang (1993), let $g_v(z) = \int h_v(s)\, dM_z(s)$ where $h_v(s) = \Lambda^v(s)$. Now define the operators,

$$W_v = \int T^{(2)}(x, z, K(P))\, g_v(z)\, dP(z),$$

$$S_v = \int f^{(2)}(x, z, \Lambda(t))\, g_v(z)\, dP(z),$$

$$U_v = \int f^{(2)}(x, z, K(P))\, g_v(z)\, dP(z),$$

where

$$f^{(2)}(x, z, \Lambda(t)) = \frac{T^{(2)}(x, z, \Lambda(t))}{C^{1/2}(t)} - \frac{T^{(1)}(z, \Lambda(t)) S^{(1)}(x, C(t))}{2C^{3/2}(t)} - \frac{T^{(1)}(x, \Lambda(t)) S^{(1)}(z, C(t))}{2C^{3/2}(t)},$$

$$f^{(2)}(x, z, K(P)) = \frac{T^{(2)}(x, z, K(P))}{S(t) C^{1/2}(t)} - \frac{T^{(1)}(z, K(P)) S^{(1)}(x, S^2 C(t))}{2S^3(t) C^{3/2}(t)} - \frac{T^{(1)}(x, K(P)) S^{(1)}(z, S^2(t) C(t))}{2S^3(t) C^{3/2}(t)}.$$

REMARK. The second canonical gradients associated with the functional $f(Q)$ evaluated at P are $f^{(2)}(x, z, \Lambda(t))$ and $f^{(2)}(x, z, K(P)(t))$ in the case of the cumulative hazard and survival function, respectively. The first canonical gradient is of $S^2(P)$ denoted by $S^{(1)}$. See Section 5 for more details.

In order to verify the third-order Edgeworth expansions in the case of the Kaplan–Meier estimator (standardized by the true variance), the "studentized" cumulative hazard, and the "studentized" Kaplan–Meier estimator, one needs now to show the linear independence of the vectors W_v, S_v, U_v, for $v = 1, \dots, k$ (which works for any $k \geq 1$), in the respective cases. We do not explicitly treat the case of $e^{-\hat{\Lambda}(t)}$ here, but note that the required result follows easily from arguments similar to those below. We first show that the condition is satisfied for the Kaplan–Meier estimator. Let us suppose that $\sum_{v=1}^{k} a_v W_v = 0$ a.s., then we have that

$$\begin{aligned}\int_0^t \frac{\hat{Y}_x(s) - Y(s)}{Y(s)} \sum_{v=1}^{k} a_v h_v(s)\, d\Lambda(s) &= \int_0^t \left[\int_0^{s-} \sum_{v=1}^{k} a_v h_v(u)\, d\Lambda(u) \right] \frac{dM_x(s)}{Y(s)} \\ &\quad - T^{(1)}(x, \Lambda(t)) \int_0^t \sum_{v=1}^{k} a_v h_v(s)\, d\Lambda(s) \quad \text{a.s.}\end{aligned}$$

Note the similarity between this and equation (2.11) in Lai and Wang (1993). It follows that we may use the remainder of their argument to obtain the desired result. That is, compute the conditional expectations under a filtration $\mathscr{F}_\gamma$, for any $\gamma < t$, and then proceed to take variances on both sides of the resulting expressions. Using standard martingale results for counting process models, one can easily calculate these. The argument is then finished off by using the absolute continuity of Λ. For the case of the studentized version of the hazard $\sum_{v=1}^k a_v S_v = 0$ a.s. implies

$$\begin{aligned}
&2C(t)\int_0^t \{Y(s) - \hat{Y}_x(s)\}\Big/ Y(s) \sum_{v=1}^k a_v h_v(s)\, d\Lambda(s) \\
&\quad -2C(t)\int_0^t \left[\int_0^{s-} \sum_{v=1}^k a_v h_v(u)\, d\Lambda(u)\right] \frac{dM_x(s)}{Y(s)} \\
&\qquad = S^{(1)}(x, C(t)) \int_0^t \sum_{v=1}^k a_v h_v(s)\, d\Lambda(s) \\
&\qquad\quad + T^{(1)}(x, \Lambda(t)) \int_0^t \frac{1}{Y(s)} \sum_{v=1}^k a_v h_v(s)\, d\Lambda(s) \\
&\qquad\quad + T^{(1)}(s, \Lambda(t)) \int_0^t \left[\int_0^{s-} \sum_{v=1}^k a_v h_v(u)\, d\Lambda(u)\right] dC(s) \quad \text{a.s.}
\end{aligned}$$

Now arguing along similar lines as above, again using the absolute continuity of Λ, one can obtain the required result although the resulting expressions will be a bit more complicated. The linear independence of the vectors U_v can be shown using similar arguments as well. These expansions are valid essentially because under (2.6) the kernels are bounded, and they contain an absolutely continuous part.

5. Canonical gradients. In this section we present the first three canonical gradients for the cumulative hazard and product integral representation of the survival function as well as the first two gradients of their asymptotic variances. The canonical gradients are simply the Gateaux derivatives of the above-mentioned functionals with the first canonical gradient being the influence function. They are perhaps the keystone to all the work in this paper and play a role at every level. We derive these results based on the functionals evaluated at $P = P_{FG}$ rather than, as is the case of Reid (1981), where the influence curve is computed by taking derivatives with respect to the two (sub)survival functions. Let $M_x(t) = \int_0^t \{dN_x(s) - \hat{Y}_x(s)\, d\Lambda(s)\}$ where $N_x(t) = I(y \le t,\ \delta = 1)$ and $\hat{Y}_x(t) = I(y \ge t)$. The canonical gradients $T^{(1)}$, $T^{(2)}$, $T^{(3)}$ for $\Lambda(t)$ are given first below:

$$T^{(1)}(x, \Lambda(t)) = \int_0^t 1/Y(s)\, dM_x(s),$$

$$T^{(2)}(x, y, \Lambda(t)) = \int_0^t \left\{Y(s) - \hat{Y}_y(s)\right\}/Y^2(s)\, dM_x(s)$$
$$+ \int_0^t \left\{Y(s) - \hat{Y}_x(s)\right\}/Y^2(s)\, dM_y(s),$$
$$T^{(3)}(x, y, z, \Lambda(t)) = 2\int_0^t \left\{Y(s) - \hat{Y}_y(s)\right\}\left\{Y(s) - \hat{Y}_z(s)\right\}/Y^3(s)\, dM_x(s)$$
$$+2\int_0^t \left\{Y(s) - \hat{Y}_x(s)\right\}\left\{Y(s) - \hat{Y}_x(s)\right\}/Y^3(s)\, dM_y(s)$$
$$+2\int_0^t \left\{Y(s) - \hat{Y}_y(s)\right\}\left\{Y(s) - \hat{Y}_x(s)\right\}/Y^3(s)\, dM_x(s)$$

and for $S(t) = \prod_{s \le t}(1 - \Delta\Lambda(s)) = K(P)$ (suppressing the dependence on t),

$$T^{(1)}(x, K(P)) = -S(t)T^{(1)}(x, \Lambda(t)),$$
$$T^{(2)}(x, y, K(P)) = -S(t)\Big\{T^{(2)}(x, y, \Lambda(t)) - T^{(1)}(x, \Lambda(t))T^{(1)}(y, \Lambda(t))$$
$$+ \sum_{s \le t} \Delta T^{(1)}(x, \Lambda(s))\Delta T^{(1)}(y, \Lambda(s))\Big\}$$

and

$$T^{(3)}(x, y, z, K(P))$$
$$= -S(t)\Big\{T^{(3)}(x, y, z, \Lambda(t)) + T^{(1)}(x, \Lambda(t))T^{(1)}(y, \Lambda(t))T^{(1)}(z, \Lambda(t))$$
$$-\big[T^{(1)}(x, \Lambda(t))T^{(2)}(y, z, \Lambda(t))$$
$$+T^{(1)}(y, \Lambda(t))T^{(2)}(x, z, \Lambda(t))$$
$$+T^{(1)}(z, \Lambda(t))T^{(2)}(x, y, \Lambda(t))\big]$$
$$+T^{(1)}(x, \Lambda(t)) \sum_{s \le t} \Delta T^{(1)}(y, \Lambda(s))\, \Delta T^{(1)}(z, \Lambda(s))$$
$$+T^{(1)}(y, \Lambda(t)) \sum_{s \le t} \Delta T^{(1)}(x, \Lambda(s))\, \Delta T^{(1)}(z, \Lambda(s))$$
$$+T^{(1)}(z, \Lambda(t)) \sum_{s \le t} \Delta T^{(1)}(x, \Lambda(s))\, \Delta T^{(1)}(y, \Lambda(s))$$
$$+ \sum_{s \le t} \Delta T^{(1)}(x, \Lambda(s))\, \Delta T^{(2)}(y, z, \Lambda(s)$$
$$+ \sum_{s \le t} \Delta T^{(1)}(y, \Lambda(s))\, \Delta T^{(2)}(x, z, \Lambda(s))$$
$$+ \sum_{s \le t} \Delta T^{(1)}(z, \Lambda(s))\, \Delta T^{(2)}(x, y, \Lambda(s))$$
$$-2 \sum_{s \le t} \Delta T^{(1)}(x, \Lambda(s))\, \Delta T^{(1)}(y, \Lambda(s))\, \Delta T^{(1)}(z, \Lambda(s))\Big\},$$

where $\Delta U(s) = U(s) - U(s-)$.

REMARK 5.1. The $T^{(1)}$, $T^{(2)}$, $T^{(3)}$ corresponding to the functional $e^{-\Lambda(t)}$ is the same as for $K(P)$ except for the omission of the Δ terms above. This gives a better idea of the difference between the two estimators $\hat{S}(t)$ and $e^{-\hat{\lambda}(t)}$, which becomes apparent in their respective Edgeworth expansions.

The first two canonical gradients for $S^2(P)$, denoted by $S^{(1)}$ and $S^{(2)}$, are now given in the particular cases where $S^2(P)$ represents $C(t)$ and $S^2(t)C(t)$ which are used in the "studentized" expansions:

$$S^{(1)}(x, C(t)) = \int_0^t \frac{1}{Y^2(s)}\, dM_x(s) + \int_0^t \frac{\{Y(s) - \hat{Y}(s)\}}{Y(s)}\, dC(s),$$

$$\begin{aligned} S^{(2)}(x, y, C(t)) = &- 2\int_0^t \frac{\{\hat{Y}_y(s) - Y(s)\}}{Y^3(s)} \{dN_x(s) - dN(s)\} \\ &-2\int_0^t \frac{\{\hat{Y}_x(s) - Y(s)\}}{Y^3(s)} \{dN_y(s) - dN(s)\} \\ &+6\int_0^t \frac{\{\hat{Y}_x(s) - Y(s)\}\{\hat{Y}_y(s) - Y(s)\}}{Y^2(s)}\, dC(s), \end{aligned}$$

and

$$\begin{aligned} S^{(1)}\big(x, S^2(t)C(t)\big) &= 2S(t)T^{(1)}(x, K(P))C(t) + S^2(t)S^{(1)}(x, C(t)) \\ S^{(2)}\big(x, y, S^2(t)C(t)\big) &= 2S(t)T^{(2)}(x, y, K(P))C(t) \\ &\quad +2T^{(1)}(x, K(P))T^{(1)}(y, K(P))C(t) \\ &\quad +2S(t)T^{(1)}(x, K(P))S^{(1)}(y, C(t)) \\ &\quad +2S(t)T^{(1)}(y, K(P))S^{(1)}(x, C(t)) \\ &\quad +S^2(t)S^{(2)}(x, y, C(t)). \end{aligned}$$

REMARK 5.2. As mentioned previously, for the studentized version of the statistics considered here, one needs to obtain the first three canonical gradients for $f(Q) = (T(Q) - T(P))/S(Q)$ evaluated at P, say $f^{(1)}$, $f^{(2)}$, $f^{(3)}$. These are constructed based on $T^{(1)}$, $T^{(2)}$, $T^{(3)}$ and $S^{(1)}$, $S^{(2)}$. We omit the details.

6. Edgeworth expansions for the sampling distributions. Here we give the explicit form of the Edgeworth expansions for the sampling distribution of the statistics considered. For some brevity, we give the general form of the expansion below and then explicitly the cumulant terms in the respective cases:

$$
\begin{aligned}
H_n(x, T(P_n), T(P)) = \Phi(x) &- n^{-1/2}\left(\frac{K_{1,1}(P)}{2} + \frac{K_{1,3}(P)}{6}(x^2 - 1)\right)\phi(x) \\
&- n^{-1}\left(\frac{K_{2,2}(P)}{2}x + \frac{K_{2,4}(P)}{24}(x^3 - 3x)\right. \\
&\qquad \left. + \frac{K_{2,6}(P)}{72}(x^5 - 10x^3 + 15x)\right)\phi(x) \\
&+ o(n^{-1}).
\end{aligned}
$$

The expansion for $K_n(x, T(P_n), T(P))$ has the same form with the $K_{i,j}(P)$ replaced by $K'_{i,j}(P)$. The cumulants $K_{i,j}(P)$ and $K'_{i,j}(P)$ are calculated based upon expectations involving the canonical gradients. Detailed expressions for these are given in Barbe and Bertail (1995) in the general setting. The cumulants associated with the Kaplan–Meier estimator in the nonstudentized case are

$$
\begin{aligned}
K_{1,1}(P) &= 0, \\
K_{1,3}(P) &= -\left[A_3(t) - (3/2)C^2(t)\right]C^{-3/2}(t), \\
K_{2,2}(P) &= \left[C(t) + (3/2)C^2(t) + A_3(t)\right]C^{-1}(t), \\
K_{2,4}(P) &= \frac{\left[A_4(t) + 22C^3(t) + 20\int C(s)\,dA_3(s)\right]}{C^2(t)} \\
&\quad + \frac{\left[24\int C^3(s)\,d\Lambda(s)\right]}{C^2(t)} - 3, \\
K_{2,6}(P) &= K_{1,3}^2(P),
\end{aligned}
$$

where $C(t) = \int_0^t (d\Lambda(s)/Y(s))$ and $A_j(t) = \int_0^t (d\Lambda(s)/Y^{j-1}(s))$ for $j = 3, 4$. The cumulants for $H_n(x, \hat{\Lambda}(t), \Lambda(t))$ are,

$$
\begin{aligned}
K_{1,1}(P) &= 0, \\
K_{1,3}(P) &= \left[A_3(t) + (3/2)C^2(t)\right]C^{-3/2}(t), \\
K_{2,2}(P) &= \left[C(t) + A_3(t)\right]C^{-1}(t), \\
K_{2,4}(P) &= \frac{\left[A_4(t) + 2C^3(t) - 4\int C(s)\,dA_3(s)\right]}{C^2(t)} \\
&\quad + \frac{\left[24\int C^3(s)\,d\Lambda(s) + 12C(t)A_3(t)\right]}{C^2(t)} - 3, \\
K_{2,6}(P) &= K_{1,3}^2(P)
\end{aligned}
$$

and for $H_n(x, e^{-\hat{\Lambda}(t)}, S(t))$

$$K_{1,1}(P) = C^{1/2}(t),$$
$$K_{1,3}(P) = -\left[A_3(t) - (3/2)C^2(t)\right]C^{-3/2}(t),$$
$$K_{2,2}(P) = \left[C(t) + (1/4)C^2(t)\right]C^{-1}(t),$$
$$K_{2,4}(P) = \frac{\left[A_4(t) + 9C^3(t) - 4\int C(s)\,dA_3(s)\right]}{C^2(t)} + \frac{\left[24\int C^3(s)\,d\Lambda(s) - 2C(t)A_3(t)\right]}{C^2(t)} - 3,$$
$$K_{2,6}(P) = K_{1,3}^2(P).$$

In the studentized case, the calculations for the cumulants are a bit more tedious since they now involve the canonical gradients for $S^2(P)$ as well. Here we just give the explicit form for the first two cumulants. $K_n(x, \hat{S}(t), S(t))$ has cumulants

$$K'_{1,1}(P) = \left[A_3(t) - (3/2)C^2(t)\right]C^{-3/2}(t),$$
$$K'_{1,3}(P) = 2\left[A_3(t) - (3/2)C^2(t)\right]C^{-3/2}(t),$$

and $K_n(x, \tilde{\Lambda}(t), \Lambda(t))$ has cumulants

$$K'_{1,1}(P) = -\left[A_3(t) + (1/2)C^2(t)\right]C^{-3/2}(t),$$
$$K'_{1,3}(P) = -2A_3(t)C^{-3/2}(t).$$

7. Concluding remarks. This work is a detailed adaptation of the methods of Barbe and Bertail (1995) to a complex situation. It blends and hopefully highlights some ideas from empirical process theory, martingale theory and the theory of functional derivatives which may be of independent interest. It is hoped the reader has been given an idea of how to apply these techniques to other situations and also that indeed the classical bootstrap may not always be the best choice. It is believed that the utilization of these weighted bootstrap methods from a practical viewpoint will not simply come about by the technical measures of accuracy illustrated here alone. Section 2.2 was written mostly with the potential user in mind. One will find that the choice of gamma weights in this setting is easy to implement in standard packages such as Splus. The fact that the weights are continuous does away with ties which occur with the classical bootstrap and results in slightly faster running times. Some preliminary simulations have been conducted, but it is felt that such a study should be done on a larger scale. In this setting there is the potentially interesting question of approximating a binomial type distribution by a continuous one analogous to the case of continuity corrections for normal approximations to binomial distributions. It is not claimed that the choice of gamma weights is optimal in a global sense; it will be interesting to compare its performance to that of other weights, and this work gives a more precise guideline to the choice of possible schemes. Nevertheless, the ease of use of gamma type weights, and its performance under the two criteria considered here, point to a very formidable competitor to the classical scheme rather than just a mathematically interesting doppelgänger. It is hoped that this work has also reminded the reader that bootstrap methods are viable alternatives to other techniques used to approximate posterior quantities.

Acknowledgments. Very special thanks to Phillipe Barbe and Patrice Bertail for graciously sharing their work, and office space, with this author and clarifying some key points. Special thanks also to M. Bruegel and R. Ayarza. Thanks to Messieurs Bertail, C. Robert, J. M. Robin, Grignon, and members of the CREST and INRA for the opportunity to spend some months in their institutes in Paris. Thanks to R. Gill, Mark van der Laan, Deb Burr, Odile Pons and M. Arcones for their preprints. Thanks also to Stephen Fienberg and David Mason. Thanks to Lenore Cowen, Carey Priebe, Cheng Cheng, Dan Naiman and Alan Goldman for their strong encouragement and inputs. I would like to express deep appreciation to Albert Y. Lo for his continued support and thoughtful advice.

REFERENCES

Akritas, M. G. (1986). Bootstrapping the Kaplan–Meier estimator. *J. Amer. Statist. Assoc.* **81** 1032–1038.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Arcones, M. A. and Giné, E. (1992). On the bootstrap of *M*-estimators and other statistical functionals. In *Exploring the Limits of the Bootstrap* (R. Lepage and L. Billard, eds.) 13–47. Wiley, New York.

Arcones, M. A. and Giné, E. (1995). On the law of the iterated logarithm for canonical *U*-statistics and processes. *Stochastic Process. Appl.* **58** 217–245.

Babu, J. G. (1991). Asymptotic theory for estimators under random censorship. *Probab. Theory Related Fields* **90** 275–290.

Barbe, P. and Bertail, P. (1995). *The Weighted Bootstrap. Lecture Notes in Statist.* **98**. Springer, New York.

Bertail, P. (1992). *La Méthode du Bootstrap, Quelques Applications et Résultats Théoretiques*. Ph.D. dissertation, Univ. Paris IX.

Bickel, P. J., Götze, F. and van Zwet, W. R. (1986). The Edgeworth expansions for *U*-statistics of degree 2. *Ann. Statist.* **14** 1463–1484.

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

Brunner, L. J. and Lo, A. Y. (1996). Limiting posterior distributions under mixture of conjugate priors. *Statist. Sinica* **6** 187–197.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *J. Amer. Statist. Assoc.* **89** 1290–1302.

Chang, M. N. (1991). Edgeworth expansions for the Kaplan–Meier estimator. *Comm. Statist. Theory Methods* **20** 2479–2494.

Chen, K. and Lo, S. H. (1996). On bootstrap accuracy with censored data. *Ann. Statist.* **24** 569–595.

Damien, P., Laud, P. W. and Smith, A. F. M. (1996). Implementation of Bayesian non-parametric inference based on beta processes. *Scand. J. Statist.* **23** 27–36.

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

DUDLEY, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In *Probability in Banach Spaces* (E. Eberlein, J. Kuelbs and M. B. Marsuc, eds.) **7** 63–82. Birkhäuser, Boston.

DUDLEY, R. M. (1992). Frechet differentiability, p-variation and uniform Donsker classes. *Ann. Probab.* **20** 1968–1982.

DUDLEY, R. M. (1994). The order of the remainder term in derivatives of composition and inverse operators for p-variation norms. *Ann. Statist.* **22** 1–20.

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

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

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

GILL, R. D. (1994). Lectures on survival analysis. *Lectures on Probability Theory*, *Ecole d'Eté de Probabilitiés de Saint-Flour XXII–1992*. *Lecture Notes in Math.* **1581** 115–241. Springer, Berlin.

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical functions. *Ann. Probab.* **18** 851–869.

GÖTZE, F. (1979). Asymptotic expansions for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50** 333–355.

GÖTZE, F. (1984). Expansions for von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **65** 599–625.

GROSS, S. T. and LAI, T. L. (1996). Bootstrap methods for truncated and censored data. *Statist. Sinica* **6** 509–530.

GUILLOU, A. (1995). Weighted bootstraps for studentized statistics. *C.R. Acad. Sci. Paris Sér. I Math* **320** 1379–1384.

HALL, P. and MAMMEN, P. (1994). On general resampling algorithms and their performance in distribution estimation. *Ann. Statist.* **22** 2011–2030.

HAEUSLER, E., MASON, D. and NEWTON, M. A. (1992). Weighted bootstrapping of means. *CWI Quarterly* **5** 213–228.

HELMERS, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a studentized U-statistic. *Ann. Statist.* **19** 470–484.

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

JAMES, L. F. (1993). The bootstrap, Bayesian bootstrap and random weighted methods for censored data models. Ph.D. dissertation, SUNY, Buffalo.

LAI, T. L. and WANG, J. Q. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap methods. *Statist. Sinica* **3** 517–542.

LO, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā Ser. A* **53** 320–333.

LO, A. Y. (1993a). A Bayesian bootstrap for censored data. *Ann. Statist.* **21** 100–123.

LO, A. Y. (1993b). A Bayesian method for weighted sampling. *Ann. Statist.* **21** 2138–2148.

MASON, D. M. and NEWTON, M. (1992). A rank statistic approach to the consistency of a general bootstrap. *Ann. Statist.* **20** 1611–1624.

PONS, O. and DE TURCKHEIM, E. (1991). Von Mises method, bootstrap and Hadamard differentiability for nonparametric general models. *Statistics* **22** 205–214.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. IMS, Hayward, CA.

PRAESTGAARD, J. T. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086.

REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9** 78–92.

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

SHAO, J. and TU, D. (1995). *The Jacknife and Bootstrap*. Springer, New York.

VAN DER LAAN, M. J. (1993). Efficient and inefficient estimation in semiparametric models. Ph.D. dissertation, Univ. Utrecht.

WELLS, M. T. and TIWARI, R. C. (1994). Bootstrapping a Bayes estimator of a survival function with censored data. *Ann. Inst. Statist. Math.* **46** 487–495.

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

DEPARTMENT OF MATHEMATICAL SCIENCES
GWC WHITING SCHOOL OF ENGINEERING
JOHNS HOPKINS UNIVERSITY
34TH & CHARLES STREETS
BALTIMORE, MARYLAND 21218
E-MAIL: james@brutus.mts.jhu.edu