The Annals of Statistics
1996, Vol. 24, No. 2, 569–595

ON BOOTSTRAP ACCURACY WITH CENSORED DATA[1]

BY KANI CHEN AND SHAW-HWA LO

HKUST and Columbia University

In survival analysis with censored data, we consider three closely related survival function estimators: the Kaplan–Meier, Nelson and moment estimators. We derive the Edgeworth expansions for these three estimators with Studentization. Edgeworth expansions for the corresponding bootstrap statistics are also given. It is found that the bootstrap approximation is better than the normal approximation for the Studentized Kaplan–Meier and Nelson estimators, but not so for the Studentized moment estimator. With these results, we construct bootstrap-based confidence intervals with better coverage probabilities. We also include some simulations which show strong agreement with our theoretical findings.

1. Introduction. Lifetime data with incomplete observations often arise in biometry and reliability theory. In some medical studies, each subject is followed from an entrance time to an exit time, and whether the exit is due to death or other reasons is recorded. With this type of right-censored data, the product-limit estimator of Kaplan and Meier (1958) (K-M estimator) has been generally accepted as the estimator of the underlying survival function. Inference with the K-M estimator that appeared in the literature mainly relies on its asymptotic normality, which was first established in Breslow and Crowley (1974) and later extended in Gill (1983). For example, the asymptotic normality of the K-M estimator can be used to build a confidence interval for the survival function at some fixed time t, and the confidence band for the survival curve can be constructed based on the fact that the K-M process, with appropriate Studentization, converges to a Brownian bridge [see, e.g., Hall and Wellner (1980) and Gill (1983)].

As an alternative to the normal approximation, the bootstrap method was proposed in Efron (1981) to approximate the distribution of the K-M estimator. It was shown later [e.g., in Lo and Singh (1986) and Akritas (1986), among others] that the bootstrap gives a correct first order asymptotic approximation of the K-M process and its quantile process. However, to our knowledge there has been no report in the literature that the bootstrap provides a better approximation and hence derives better confidence intervals for survival function than those constructed based on the normal approximations.

Received May 1993; revised May 1995.
[1]Research supported in part by AFOSR Grant F49620-94-1-0035.
AMS 1991 *subject classifications*. 62G05, 62E20.
Key words and phrases. Kaplan–Meier estimator, Nelson estimator, moment estimator, Greenwood formula, Edgeworth expansion.

It is well known that under certain regularity conditions, the bootstrap approximation is better than the normal approximation for a broad class of Studentized statistics. The case of sample mean was first observed in Abramovitch and Singh (1985) and more general cases dealing with smooth functions of sample mean can be found in Hall (1988). Similar results were also obtained in Helmers (1991) for Studentized U-statistics. In summary, it is generally believed that the bootstrap of Studentized statistics (or, more generally, pivotal statistics) leads to confidence intervals with better coverage probabilities. In contrast to Studentized statistics, the bootstrap of statistics without Studentization has an error, usually of order $n^{-1/2}$, which is no better than the normal approximation in terms of order. For more discussions of this issue, the readers are referred to Helmers (1991) and Hall (1992).

The question that naturally arises in survival analysis is whether the bootstrap approximation for the Studentized K-M estimator is better than the normal approximation. The purpose of this article is to address this problem. Our findings indicate that the answer is indeed positive. We show that the difference between the distribution function of the Studentized K-M estimator and that of its bootstrap analogue is $o(n^{-1/2})$ almost surely. From this result we prove that the bootstrap offers a confidence interval with better coverage probability than the normal approximation does. To support our results we present some simulations with small sample size and heavy censorship. The simulations strongly support our theoretical findings. In addition to the Studentized K-M estimator, we also study the closely related Nelson estimator [also called Nelson–Aalen estimator; see, e.g., Altshuler (1970)] and the moment estimator [cf. Prentice (1978) and Cuzick (1985)], as well as the estimators of the corresponding cumulative hazard functions. Furthermore, the explicit formulas of the Edgeworth expansions of the Studentized statistics are obtained in all three cases.

It is noted that the whole analysis involved in deriving these results cannot be perceived as a special case of what seems to be the belief that the bootstrap approximation is better for Studentized statistics. In fact, the results obtained in this paper are, to a certain extent, surprising. First, the bootstrap of the Studentized moment estimator of the survival function is no better than the normal approximation, while the bootstrap of the Studentized Nelson and K-M estimators do perform better. These three asymptotically equivalent estimators have quite different behavior in their effects of the bootstrap. Second, the K-M estimator is a natural generalization of the empirical survival function when data are complete: the bootstrap accuracy of the Studentized K-M estimator is expected to conform with that of the Studentized empirical survival function. It turns out that the former is second order accurate and the latter is only first order accurate. The reason for this is that the empirical distribution is the average of iid Bernoulli variables whose distribution is obviously latticed. The bootstrap with Studentization reported in the literature is restricted to the cases of *nice* statistics, such as sample means or U-statistics and their smooth functions [cf. Helmers (1992); Hall (1988, 1992)]. However, the K-M estimator and the moment estimator, taking product forms and complicated by Studentization, have irregular behavior in bootstrap accuracy and hence require a rather delicate treatment. As a result, the Edgeworth expansions for the Studentized K-M estimator and its bootstrap cannot be directly derived. In fact, representing the Studentized K-M estimator as a sum of iid variables plus an error term does not provide us enough tools to tackle the problem [cf. Lo and Singh (1986)]. The approaches we adopt to encounter the difficulties are to first write the target statistic as a U-statistic plus smaller remainder terms. We then derive the Edgeworth expansions of this U-statistic and employ various exponential inequalities to control the orders of the remainder terms and show their negligibility. In this way, we derive the Edgeworth expansions for the Studentized Nelson estimator. The expansions of the Studentized K-M and moment estimators are obtained by observing the fact that the differences multiplied by n among the three estimators are asymptotic constants. The expansion for the bootstrap of the Studentized Nelson estimator, though more complicated, bears the same idea and is performed in a similar fashion. To deal with the bootstrap of the Studentized K-M estimator, we consider a smoothed version of the bootstrap. This allows us to obtain the expansions for the bootstrap of the three Studentized estimators and thus conclude their bootstrap approximations.

In Section 2 we present Edgeworth expansions for several Studentized statistics, including the estimator of the cumulative hazard function, the K-M, the Nelson and the moment estimator of the survival function. In Section 3, we derive the Edgeworth expansions for the corresponding bootstrap statistics. By comparison of the coefficients of the expansions, we conclude that the bootstraps of the Studentized K-M and Nelson estimators are better than the normal approximations, while the bootstrap for the Studentized moment estimator is not. The result is then extended to address the issue of coverage probabilities of confidence intervals in Section 4. Finally we give an example of simulations with small sample but rather heavy censorship. The example shows strong agreement with the theoretical findings presented in this paper. Some of the proofs are deferred to the Appendix.

Before we move on to the next section, we introduce some notations and relevant estimates as follows.

Let (X_i, Y_i), $i = 1, \dots, n$, be iid pairs of nonnegative random variables, with X_i independent of Y_i for each i. Let F and G denote the distribution functions of the X and Y populations, respectively. The distribution function F is known as the survival distribution and G is known as the censoring distribution. We denote by $\overline{F}$ and $\overline{G}$ the survival functions of X and Y, respectively (i.e., $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$). Let δ_i denote the usual indicator function of the censoring, that is, $\delta_i = 1$ if $X_i \le Y_i$ and $\delta_i = 0$ if $X_i > Y_i$. The observed data consist of (Z_i, δ_i), $1 \le i \le n$, where $Z_i = X_i \wedge Y_i$ for $1 \le i \le n$.

Let $\overline{H}_1(t) = P(Z_1 \ge t,\ \delta_1 = 1)$ and $\overline{H}(t) = P(Z_1 \ge t)$ be the subsurvival and survival functions of the observations, respectively, and let $\Lambda(t) = -\int_0^t d\overline{H}_1/\overline{H}$ be the cumulative hazard function of the X population. Let

$\hat{\bar{H}}_1(\cdot)$, $\hat{\bar{H}}(\cdot)$ and $\hat{\Lambda}(\cdot)$ be the corresponding sample analogue of the above three functions, that is,

$$\hat{\bar{H}}_1(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i \ge t,\, \delta_i = 1), \qquad \hat{\bar{H}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i \ge t)$$

and

$$\hat{\Lambda}(t) = -\int_0^t \frac{d\hat{\bar{H}}_1}{\hat{\bar{H}}} = \sum^{*} \frac{r_i\delta_i}{n_i},$$

where $r_i = \#\{j;\, Z_j = Z_i\}$, $n_i = \#\{j;\, Z_j \ge Z_i\}$ and $*$ runs over all i such that $Z_i \le t$.

With these notations, the K-M estimator, the moment estimator and the Nelson estimator, denoted by $\hat{\bar{F}}$, $\hat{\bar{F}}_M$ and $\hat{\bar{F}}_N$, can be expressed as

$$\hat{\bar{F}}(t) = \prod^{*}\left(1 - \frac{r_i}{n_i}\right)^{\delta_i}, \tag{1.1}$$

$$\hat{\bar{F}}_M(t) = \prod^{*}\left(1 - \frac{r_i}{n_i + 1}\right)^{\delta_i} \tag{1.2}$$

and

$$\hat{\bar{F}}_N(t) = \exp\bigl(-\hat{\Lambda}(t)\bigr), \tag{1.3}$$

respectively. The variance of $\hat{\bar{F}}(t)$ estimated via the Greenwood (1926) formula is $\hat{\bar{F}}(t)^2\hat{\sigma}_G^2(t)/n$, where

$$\hat{\sigma}_G^2(t) = \sum^{*} \frac{n\delta_i r_i}{n_i(n_i - r_i)}. \tag{1.4}$$

Likewise, one can use $\hat{\bar{F}}_N^2\hat{\sigma}_G^2/n$ and $\hat{\bar{F}}_M^2\hat{\sigma}_G^2/n$ to estimate the variances of $\hat{\bar{F}}_N$ and $\hat{\bar{F}}_M$, respectively.

Notice that $\hat{\sigma}_G^2(t)$ is an estimator of $\sigma^2(t) = -\int_0^t \bar{H}^{-2}\, d\bar{H}_1$. It can also be written as $-\int_0^t (\hat{\bar{H}}(s)\hat{\bar{H}}(s+))^{-1}\, d\hat{\bar{H}}_1(s)$. In our proofs we shall often use the sample analogue of $\sigma^2(t)$, that is,

$$\hat{\sigma}^2(t) = -\int_0^t \hat{\bar{H}}^{-2}\, d\hat{\bar{H}}_1.$$

2. The Edgeworth expansions for the Studentized estimators. We assume throughout the paper that:

(2.1) F and G are continuous; t is a fixed positive number such that $\bar{F}(t)\bar{G}(t) > 0$ and $G(t) > 0$.

The above assumption is reasonable. If $\bar{G}(t) = 0$ and $\bar{F}(t) > 0$, the asymptotic normality of the Studentized K-M estimator does not hold [cf. Chen and

Ying (1996)]. On the other hand, the case with $\bar{F}(t) = 0$ is obviously not of interest to us. In order to obtain the Edgeworth expansion for the Studentized K-M estimator, it is therefore natural to restrict our attention to the estimation of the survival function at a *fixed* time t with $\bar{F}(t)\bar{G}(t) > 0$. The assumption $G(t) > 0$ is necessary for the second order accuracy of the bootstrap. Without this assumption the K-M estimator at time t reduces to the empirical survival function, which has quite different expansion and bootstrap accuracy.

REMARK. The validity of the expansion of the sum of iid variables or U-statistics typically requires Cramér's condition plus other moment conditions. In the current case, the assumption (2.1) is strong enough to warrant all the expansions in this paper. Thanks to this assumption, the U-statistics we consider in this paper are bounded and the distributions of their kernels are continuous. Therefore the Edgeworth expansions are valid.

The presentation of this section is arranged as follows. We first derive the Edgeworth expansion for the Studentized cumulative hazard estimator by expressing it as sum of two statistics θ and γ_3 (Proposition 1). The expansion of θ and the negligibility of γ_3 are shown in Lemmas 1 and 2 in the Appendix. (The statistic θ has a special form, and the technique to handle it is also used in the proof of Theorem 1.) Since the Nelson estimator is just the exponential function of the cumulative hazard estimator, the result in Proposition 1 is inherited to show the Edgeworth expansion for the Studentized Nelson estimator in Theorem 1. We then prove that the differences among the Nelson estimator, the K-M estimator and the moment estimator are asymptotically constant on the order of n^{-1}. The expansions for the Studentized K-M and moment estimators are easily derived in light of the parallel result obtained in Theorem 1.

Because t is fixed throughout the paper, for simplicity we shall write the function value at time t as the function itself without confusion [e.g., $\hat{\Lambda} \equiv \hat{\Lambda}(t)$]. We shall also denote a function of (z_i, δ_i) as a function of z_i only. For example, $f((z_i, \delta_i))$ will be expressed as $f(z_i)$.

PROPOSITION 1. *Under the assumption in* (2.1), *one can write*

$$P\left(\frac{n^{1/2}(\hat{\Lambda} - \Lambda)}{\hat{\sigma}_G} < x\right) = \Phi(x) + n^{-1/2}\phi(x)\left(\kappa_1 x^2 + \kappa_2\right) + o(n^{-1/2}) \tag{2.2}$$

uniformly in x, *where* Φ *and* ϕ *are the distribution function and density function of the standard normal distribution*, *respectively*, *and* κ_1 *and* κ_2 *are constants depending on* F, G *and* t *only*.

PROOF. Recall the definitions of $\hat{\Lambda}(t)$ and $\Lambda(t)$ given in the previous section. We can write

$$
\begin{aligned}
\hat{\Lambda}(t) - \Lambda(t)
&= -\int_0^t \frac{d\hat{\overline{H}}_1}{\hat{\overline{H}}} + \int_0^t \frac{d\overline{H}_1}{\overline{H}} \\
&= -\int_0^t \frac{d\left(\hat{\overline{H}}_1 - \overline{H}_1\right)}{\overline{H}} + \int_0^t \frac{\hat{\overline{H}} - \overline{H}}{\overline{H}^2}\, d\overline{H}_1 + \int_0^t \frac{\hat{\overline{H}} - \overline{H}}{\overline{H}^2}\, d\left(\hat{\overline{H}}_1 - \overline{H}_1\right) \\
&\quad - \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^2}{\overline{H}^3}\, d\overline{H}_1 - \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^2}{\overline{H}^3}\, d\left(\hat{\overline{H}}_1 - \overline{H}_1\right) \\
&\quad + \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^3}{\overline{H}^3 \hat{\overline{H}}}\, d\hat{\overline{H}}_1 \\
&= -2\int_0^t \frac{d\hat{\overline{H}}_1}{\overline{H}} + 2\int_0^t \frac{\hat{\overline{H}}}{\overline{H}^2}\, d\overline{H}_1 + \int_0^t \frac{\hat{\overline{H}}}{\overline{H}^2}\, d\hat{\overline{H}}_1 - \int_0^t \frac{\hat{\overline{H}}^2}{\overline{H}^3}\, d\overline{H}_1 \\
&\quad - \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^2}{\overline{H}^3}\, d\left(\hat{\overline{H}}_1 - \overline{H}_1\right) + \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^3}{\overline{H}^3 \hat{\overline{H}}}\, d\hat{\overline{H}}_1 \\
&= \frac{1}{n^2} \sum_{1 \le i < j \le n} h(z_i, z_j) + \gamma,
\end{aligned}
\tag{2.3}
$$

where

$$
\begin{aligned}
h(z_i, z_j) &= \frac{2\delta_i 1_{[z_i \le t]}}{\overline{H}(z_i)} + \frac{2\delta_j 1_{[z_j \le t]}}{\overline{H}(z_j)} + 2\int_0^t \frac{1_{[z_i \ge s]}}{\overline{H}^2(s)}\, d\overline{H}_1(s) \\
&\quad + 2\int_0^t \frac{1_{[z_j \ge s]}}{\overline{H}^2(s)}\, d\overline{H}_1(s) + \int_0^t \frac{1_{[z_i \ge s]}}{\overline{H}^2(s)}\, d1_{[z_j \ge s,\, \delta_j = 1]} \\
&\quad + \int_0^t \frac{1_{[z_j \ge s]}}{\overline{H}^2(s)}\, d1_{[z_i \ge s,\, \delta_i = 1]} - 2\int_0^t \frac{1_{[z_i \ge s,\, z_j \ge s]}}{\overline{H}^3(s)}\, d\overline{H}_1(s)
\end{aligned}
\tag{2.4}
$$

and

$$
\begin{aligned}
\gamma = &-\frac{2}{n}\int_0^t \frac{d\hat{\overline{H}}_1}{\overline{H}} + \frac{2}{n}\int_0^t \frac{\hat{\overline{H}}}{\overline{H}^2}\, d\overline{H}_1 + \frac{1}{n}\int_0^t \frac{d\hat{\overline{H}}_1}{\overline{H}^2} - \frac{1}{n}\int_0^t \frac{\hat{\overline{H}}}{\overline{H}^3}\, d\overline{H}_1 \\
&- \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^2}{\overline{H}^3}\, d\left(\hat{\overline{H}}_1 - \overline{H}_1\right) + \int_0^t \frac{(\hat{\overline{H}} - \overline{H})^3}{\overline{H}^3 \hat{\overline{H}}}\, d\hat{\overline{H}}_1.
\end{aligned}
\tag{2.5}
$$

Define $g(z_1) = E(h(z_1, z_2) | (z_1, \delta_1))$. Then

$$
\begin{aligned}
g(z_1) = {} & \frac{2\delta_1 1_{[z_1 \le t]}}{\overline{H}(z_1)} + 2\int_0^t \frac{1_{[z_1 \ge s]}}{\overline{H}^2(s)}\, d\overline{H}_1(s) + \int_0^t \frac{1_{[z_1 \ge s]}}{\overline{H}^2(s)}\, d\overline{H}_1(s) \\
& - \frac{\delta_1 1_{[z_1 \le t]}}{\overline{H}(z_1)} - 2\int_0^t \frac{1_{[z_1 \ge s]}}{\overline{H}^2(s)}\, d\overline{H}_1(s) \qquad (2.6) \\
= {} & \frac{\delta_1 1_{[z_1 \le t]}}{\overline{H}(z_1)} + \int_0^t \frac{1_{[z_1 \ge s]}}{\overline{H}^2(s)}\, d\overline{H}_1(s).
\end{aligned}
$$

Now write

$$
\begin{aligned}
\hat{\sigma}^2 - \sigma^2 &= -\int_0^t \frac{d\hat{\overline{H}}_1}{\hat{\overline{H}}^2} + \int_0^t \frac{d\overline{H}_1}{\overline{H}^2} \qquad (2.7) \\
&= \frac{1}{n}\sum_{i=1}^n f(z_i) + \gamma_1 \quad \text{(say)},
\end{aligned}
$$

where

$$
f(z_1) = \frac{\delta_1 1_{[z_1 \le t]}}{\overline{H}^2(z_1)} + 2\int_0^t \frac{1_{[z_1 \ge s]}}{\overline{H}^3(s)}\, d\overline{H}_1(s) - \int_0^t \frac{d\overline{H}_1}{\overline{H}^2},
$$

$$
\gamma_1 = 2\int_0^t \frac{\hat{\overline{H}} - \overline{H}}{\overline{H}^3}\, d\left(\hat{\overline{H}}_1 - \overline{H}_1\right) - \int_0^t \frac{\left(\hat{\overline{H}} - \overline{H}\right)^2 \left(\overline{H} + 2\hat{\overline{H}}\right)}{\hat{\overline{H}}^2 \overline{H}^3}\, d\hat{\overline{H}}_1.
$$

It follows that

$$
\begin{aligned}
\frac{\sigma}{\hat{\sigma}} &= 1 - \frac{\hat{\sigma}^2 - \sigma^2}{2\sigma^2} - \frac{(\hat{\sigma}^2 - \sigma^2)(2\sigma^2 - \hat{\sigma}^2 - \hat{\sigma}\sigma)}{2\sigma^2\hat{\sigma}(\hat{\sigma} + \sigma)} \\
&= 1 - \frac{1}{2n\sigma^2}\sum_{i=1}^n f(z_i) - \frac{\gamma_1}{2\sigma^2} + \frac{(\hat{\sigma} - \sigma)^2(\hat{\sigma} + 2\sigma)}{2\hat{\sigma}\sigma^2} \qquad (2.8) \\
&= 1 - \frac{1}{2\sigma^2 n}\sum_{i=1}^n f(z_i) + \gamma_2,
\end{aligned}
$$

where

$$
\gamma_2 = -\frac{\gamma_1}{2\sigma^2} + \frac{(\hat{\sigma} - \sigma)^2(\hat{\sigma} + 2\sigma)}{2\hat{\sigma}\sigma^2}. \qquad (2.9)
$$

From (2.3), (2.7) and (2.8), we have

$$
\begin{aligned}
\frac{n^{1/2}(\hat{\Lambda}-\Lambda)}{\hat{\sigma}} &= \frac{n^{1/2}}{2\sigma}\left(\frac{2}{n^2}\sum_{i<j} h(z_i, z_j) + 2\gamma\right)\frac{\sigma}{\hat{\sigma}} \\
&= \frac{n^{1/2}}{2\sigma}\left(\frac{2}{n^2}\sum_{i<j} h(z_i, z_j)\right)\left(1 - \frac{1}{2n\sigma^2}\sum_{i=1}^{n} f(z_i)\right) \\
&\quad + \frac{n^{1/2}\gamma_2}{2\sigma}\left(\frac{2}{n^2}\sum_{i<j} h(z_i, z_j)\right) + \frac{n^{1/2}\gamma}{\hat{\sigma}} \\
&= \theta + \gamma_3,
\end{aligned}
\tag{2.10}
$$

where

$$
\begin{aligned}
\theta &= \frac{n^{1/2}}{2\sigma}\left(\frac{2}{n^2}\sum_{i<j} h(z_i, z_j)\right)\left(1 - \frac{1}{2n\sigma^2}\sum_{i=1}^{n} f(z_i)\right), \\
\gamma_3 &= \frac{\gamma_2 n^{1/2}}{2\sigma}\left(\frac{2}{n^2}\sum_{i<j} h(z_i, z_j)\right) + \frac{n^{1/2}\gamma}{\hat{\sigma}}.
\end{aligned}
\tag{2.11}
$$

Lemmas 1 and 2 in the Appendix give the Edgeworth expansion of θ and the negligibility of γ_3. Therefore using the delta method, we conclude that

$$
\begin{aligned}
P&\left(\frac{n^{1/2}(\hat{\Lambda}-\Lambda)}{\hat{\sigma}} < x\right) \\
&= P(\theta + \gamma_3 \le x) \\
&= \Phi(x) + n^{-1/2}\phi(x)\left(\kappa_1 x^2 + \kappa_2\right) + o(n^{-1/2})
\end{aligned}
\tag{2.12}
$$

uniformly in x. Now we have proved (2.2) with $\hat{\sigma}_G$ replaced by $\hat{\sigma}$ there. To show (2.2), write

$$
\frac{n^{1/2}(\hat{\Lambda}-\Lambda)}{\hat{\sigma}_G} = \frac{n^{1/2}(\hat{\Lambda}-\Lambda)}{\hat{\sigma}} + \frac{n^{1/2}(\hat{\Lambda}-\Lambda)}{\hat{\sigma}}\left(\frac{\hat{\sigma}}{\hat{\sigma}_G} - 1\right)
$$

and it suffices to show

$$
P\left(\left|\frac{\hat{\sigma}}{\hat{\sigma}_G} - 1\right| > n^{-2/3}\right) = o(n^{-1/2}). \tag{2.13}
$$

Notice that

$$
\left|\frac{\hat{\sigma}}{\hat{\sigma}_G} - 1\right| \le \left|\frac{\hat{\sigma}^2 - \hat{\sigma}_G^2}{\hat{\sigma}_G^2}\right| = \frac{1}{n\hat{\sigma}_G^2}\int_0^t \frac{d\hat{\overline{H}}_1(s)}{\hat{\overline{H}}^2(s)\hat{\overline{H}}(s+)} \le \frac{1}{n\hat{\overline{H}}(t+)}
$$

and we have

$$P\left(\left|\frac{\hat{\sigma}}{\hat{\sigma}_G} - 1\right| > n^{-2/3}\right) \le P\left(\hat{\overline{H}}(t+) \le n^{-1/3}\right) = o(n^{-\lambda}) \tag{2.14}$$

for arbitrary $\lambda > 0$. This establishes (2.13) and the proof is complete. □

REMARK. The explicit formulas of κ_1 and κ_2 are given in (A.4) in the Appendix. They are functions of the population distribution F, G and time t only.

Now we are ready to derive the Edgeworth expansion for the Studentized Nelson estimator.

THEOREM 1. *Under the assumption in* (2.1), *one can write*

$$P\left(\frac{n^{1/2}\left(\hat{\overline{F}}_N - \overline{F}\right)}{\hat{\overline{F}}_N \hat{\sigma}_G} \le x\right) = \Phi(x) + n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2\right) + o(n^{-1/2}) \tag{2.15}$$

uniformly in x, *where* $\tilde{\kappa}_1 = -\kappa_1 + \sigma/2$ *and* $\tilde{\kappa}_2 = -\kappa_2$.

PROOF. In view of (2.13) and the delta method, it suffices to show (2.15) with $\hat{\sigma}_G$ replaced by $\hat{\sigma}$ on the left-hand side. By Taylor expansion, we can write

$$\frac{n^{1/2}\left(\hat{\overline{F}}_N - \overline{F}\right)}{\hat{\overline{F}}_N \hat{\sigma}} = -\frac{n^{1/2}(\hat{\Lambda} - \Lambda)}{\hat{\sigma}}\left(1 + \frac{1}{2}(\hat{\Lambda} - \Lambda) + \frac{\xi}{6}(\hat{\Lambda} - \Lambda)^2\right),$$

where ξ satisfies $|\xi - 1| \le |\exp(\hat{\Lambda} - \Lambda) - 1|$. From the expression of $\hat{\Lambda} - \Lambda$ in (2.3), we can further write

$$\frac{n^{1/2}\left(\hat{\overline{F}}_N - \overline{F}\right)}{\hat{\overline{F}}_N \hat{\sigma}} = -\frac{n^{1/2}(\hat{\Lambda} - \Lambda)}{\hat{\sigma}}\left(1 + \frac{1}{2n}\sum_{i=1}^{n} g(z_i)\right) - \frac{n^{1/2}(\hat{\Lambda} - \Lambda)}{\hat{\sigma}} \times\left(\frac{1}{2n^2}\sum_{i<j}\varphi(z_i, z_j) - \frac{1}{2n^2}\sum_{i=1}^{n} g(z_i) + \frac{\gamma}{2} + \frac{\xi}{6}(\hat{\Lambda} - \Lambda)^2\right),$$

where $\varphi(z_i, z_j) = h(z_i, z_j) - g(z_i) - g(z_j)$. Since $g(z_i) = E(h(z_i, z_j)|(z_i, \delta_i))$ and $Eg(z_i) = 0$, we conclude $E((1/2n^2)\sum_{i<j}\varphi(z_i, z_j))^2 = O(n^{-2})$ and $E((1/2n^2)\sum_{i=1}^{n} g(z_i))^2 = O(n^{-2})$. In view of Lemma 2 in the Appendix and the Edgeworth expansion of $n^{1/2}(\hat{\Lambda} - \Lambda)/\hat{\sigma}$, one can show the second term in the right-hand side above is clearly negligible. For the first term, we notice

that $n^{1/2}(\hat{\Lambda} - \Lambda)/\hat{\sigma} = \theta + \gamma_3$, γ_3 is negligible (Lemma 2), and θ can be written as

$$\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\left(\tilde{h}(z_i, z_j) - \frac{1}{n\sigma^2}E(g(z_1)f(z_1))\right)$$

plus some negligible terms [Lemma 1; $\tilde{h}$ is a function defined in (A.2)]. Hence it suffices to show

$$L \equiv -\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\left(\tilde{h}(z_i, z_j) - \frac{1}{n\sigma^2}E(g(z_1)f(z_1))\right)\left(1 + \frac{1}{2n}\sum_{i=1}^{n} g(z_i)\right)$$

has the same Edgeworth expansion as the right-hand side of (2.15). Now L can be further written as a U-statistic L_1 plus four negligible terms in a similar fashion as we express θ in Lemma 1. The U-statistic L_1 can be expressed as

$$L_1 = -\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\Bigg(\tilde{h}(z_i, z_j) + g(z_i)g(z_j) - \frac{1}{n\sigma^2}E(g(z_1)f(z_1)) + \frac{\sigma^2}{n}\Bigg).$$

From Theorem 1.2 of Bickel, Götze and van Zwet (1986), we have

$$P(L_1 \le x) = \Phi(x) + n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2\right) + o(n^{-1/2})$$

uniformly in x, where $\tilde{\kappa}_1 = -\kappa_1 + \sigma/2$ and $\tilde{\kappa}_2 = -\kappa_2$. The proof is thus complete. □

The following theorem gives the Edgeworth expansion of the Studentized K-M estimator.

THEOREM 2. *Under the assumption in* (2.1), *one can write*

$$P\left(\frac{n^{1/2}\left(\hat{\bar{F}} - \bar{F}\right)}{\hat{\bar{F}}\hat{\sigma}_G} \le x\right) = \Phi(x) + n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2 + \sigma/2\right) + o(n^{-1/2}) \tag{2.16}$$

uniformly in x.

PROOF. We shall first show

$$P\left(\left|\hat{\bar{F}} - \hat{\bar{F}}_N + \frac{\bar{F}\sigma^2}{2n}\right| > (n\log n)^{-1/2}\right) = o(n^{-1/2}). \tag{2.17}$$

To this end, following the notation in Section 1, let $r_i = \#\{j;\ z_j = z_i\}$ and $n_i = \#\{j;\ z_j \ge z_i\}$; Σ^* means summation over all z_i such that $z_i \le t$. The continuity of F and G entails $P(r_i = 1) = 1$. Let $\eta = 8$ if $n\hat{\bar{H}}(t+) \geq 2$; $= \infty$ otherwise. Now write

$$\left|\log \hat{\bar{F}} - \log \hat{\bar{F}}_N + \frac{\hat{\sigma}^2}{2n}\right| = \left|\sum^* \delta_i \log\left(1 - \frac{1}{n_i}\right) + \sum^* \frac{\delta_i}{n_i} + \sum^* \frac{\delta_i}{2n_i^2}\right| \tag{2.18}$$
$$\leq \eta \sum^* \frac{\delta_i}{3n_i^3} \leq \frac{\eta}{3n^2 \hat{\bar{H}}^3(t)}.$$

The first inequality appearing above is due to Taylor expansion. It is easy to see that

$$P\left(\left|\log \hat{\bar{F}} - \log \hat{\bar{F}}_N + \frac{\hat{\sigma}^2}{2n}\right| > (n \log n)^{-1}\right) = o(n^{-1/2}). \tag{2.19}$$

Applying Taylor expansion again, we have

$$P\left(\left|\frac{\hat{\bar{F}} - \hat{\bar{F}}_N}{\hat{\bar{F}}_N} + \frac{\hat{\sigma}^2}{2n}\right| > (n \log n)^{-1}\right) = o(n^{-1/2}). \tag{2.20}$$

Since $P(|\hat{\sigma}^2 - \sigma^2| > n^{-1/2}(\log n)^{-1}) = o(n^{-1/2})$ is proved in (A.7), the result (2.17) follows from (2.15) and (2.20). The theorem is an easy consequence of (2.17), (2.15) and the delta method. We omit the details. □

The Edgeworth expansion for the Studentized moment estimator can be derived similarly. Following the previous notations, write

$$\begin{aligned}
&\left|\log \hat{\bar{F}}_M - \log \hat{\bar{F}}_N - \frac{\hat{\sigma}^2}{2n}\right| \\
&\quad = \left|\sum^* \delta_i \log\left(1 - \frac{1}{n_i + 1}\right) + \sum^* \frac{\delta_i}{n_i} - \sum^* \frac{\delta_i}{2n_i^2}\right| \\
&\quad \leq \left|-\sum^* \frac{\delta_i}{n_i + 1} + \sum^* \frac{\delta_i}{n_i} - \sum^* \frac{\delta_i}{2n_i^2} - \sum^* \frac{\delta_i}{2(n_i + 1)^2}\right| + \eta \sum^* \frac{\delta_i}{3n_i^3} \\
&\quad \leq \sum^* \frac{\delta_i}{2n_i^2(n_i + 1)^2} + \eta \sum^* \frac{\delta_i}{3n_i^3} \leq \frac{2 + \eta}{3n^2 \hat{\bar{H}}^3(t)}.
\end{aligned}$$

Similarly, we have

$$P\left(\left|\hat{\bar{F}}_M - \hat{\bar{F}}_N - \frac{\bar{F}\sigma^2}{2n}\right| > (n \log n)^{-1}\right) = o(n^{-1/2}). \tag{2.21}$$

We thus proved the following theorem.

THEOREM 3. *Under the assumption* (2.1), *one can write*

$$P\left(\frac{n^{1/2}\left(\hat{\bar{F}}_M - \bar{F}\right)}{\hat{\bar{F}}_M \hat{\sigma}_G} \le x\right) = \Phi(x) + n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2 - \frac{\sigma}{2}\right) + o(n^{-1/2})$$

uniformly in x.

REMARK. From (2.17) and (2.21), the K-M estimator and the moment estimator are asymptotically different from the Nelson estimator by $\bar{F}\sigma^2/2n$ and $-\bar{F}\sigma^2/2n$. As a result, the differences of the Edgeworth expansions for these three estimators with Studentization are $\phi(x)\sigma/2$ or $\phi(x)\sigma$ on the order of $n^{-1/2}$.

3. The Edgeworth expansions for the bootstrap statistics. In this section we study the issue of bootstrap approximations. Recall that the bootstrap sample is obtained by simple random sample with replacement from the set of observations $\{(Z_i, \delta_i), i = 1, \dots, n\}$. Let $\{(Z_i^*, \delta_i^*), i = 1, \dots, n\}$ denote the bootstrap sample. We shall put a sign * on each statistic or function associated with the bootstrap sample. For example, P^* is the probability measure on the bootstrap sample space,

$$\hat{\bar{H}}_1^*(s) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i^* \ge s, \delta_i^* = 1), \qquad \hat{\bar{H}}^*(s) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i^* \ge s)$$

and

$$g^*(z_1^*) = \frac{\delta_1^* 1_{[z_1^* \le t]}}{\hat{\bar{H}}(z_1^*)} + \int_0^t \frac{1_{[z_1^* \ge s]}}{\hat{\bar{H}}^2(s)}\, d\hat{\bar{H}}_1(s),$$

and so forth. Now $\{\hat{\bar{H}}_1^*(s),\ s \in [0,\infty)\}$ and $\{\hat{\bar{H}}^*(s),\ s \in [0,\infty)\}$ are empirical processes with parent distributions $\{\hat{\bar{H}}_1(s),\ s \in [0,\infty)\}$ and $\{\hat{\bar{H}}(s),\ s \in [0,\infty)\}$, respectively.

The following theorem shows the bootstrap accuracy for the Studentized cumulative hazard estimator and for the Studentized Nelson estimator.

THEOREM 4. *Under the assumption in* (2.1), *we have*, *for any* $\delta > 0$ *and* $\lambda > 0$,

$$P\left(\sup_x \left| P^*\left(\frac{n^{1/2}(\hat{\Lambda}^* - \hat{\Lambda})}{\hat{\sigma}_G^*} < x\right) - P\left(\frac{n^{1/2}(\hat{\Lambda} - \Lambda)}{\hat{\sigma}_G} < x\right)\right| > \delta n^{-1/2}\right) = o(n^{-\lambda}) \tag{3.1}$$

and

$$P\left(\sup_x\left|P^*\left(\frac{n^{1/2}\left(\hat{\bar{F}}_N^* - \hat{\bar{F}}_N\right)}{\hat{\bar{F}}_N^*\hat{\sigma}_G^*} < x\right) - P\left(\frac{n^{1/2}\left(\hat{\bar{F}}_N - \bar{F}\right)}{\hat{\bar{F}}_N\hat{\sigma}_G} < x\right)\right| > \delta n^{-1/2}\right) = o(n^{-\lambda}). \tag{3.2}$$

It is noted that the second order accuracy of bootstrap presented in the above formulation implies second order accuracy in the sense of almost sure convergence or, more loosely, in probability convergence. In fact, by choosing δ small and $\lambda > 1$, we have, from the Borel–Cantelli lemma,

$$\sup_x\left|P^*\left(\frac{n^{1/2}(\hat{\Lambda}^* - \hat{\Lambda})}{\hat{\sigma}_G^*} < x\right) - P\left(\frac{n^{1/2}(\hat{\Lambda} - \Lambda)}{\hat{\sigma}_G} < x\right)\right| = o(n^{-1/2})$$

and

$$\sup_x\left|P^*\left(\frac{n^{1/2}\left(\hat{\bar{F}}_N^* - \hat{\bar{F}}_N\right)}{\hat{\bar{F}}_N^*\hat{\sigma}_G^*} < x\right) - P\left(\frac{n^{1/2}\left(\hat{\bar{F}}_N - \bar{F}\right)}{\hat{\bar{F}}_N\hat{\sigma}_G} < x\right)\right| = o(n^{-1/2})$$

almost surely. Similar results also hold for the Studentized K-M estimator following Theorem 5. The formulation of bootstrap accuracy that we choose to present in the theorems is necessary for a rigorous proof of second order accuracy of the bootstrap-based confidence intervals.

Proof of Theorem 4. Given $\lambda > 0$, it follows from the convergence of empirical process that there exists a set $\mathscr{E}_n$ with $P(\mathscr{E}_n^c) = o(n^{-\lambda})$ such that uniformly on $\mathscr{E}_n$,

$$\sup_{s\geq 0}\left|\hat{\bar{H}}_1(s) - \bar{H}_1(s)\right| = o(1) \quad \text{and} \quad \sup_{s\geq 0}\left|\hat{\bar{H}}(s) - \bar{H}(s)\right| = o(1). \tag{3.3}$$

Let $\mathscr{S} = [0, \infty) \times \{0, 1\}$. Recall the definition of the functions f, g, h and φ in Section 2. It is easy to see, uniformly on $\mathscr{E}_n$, we have

$$\begin{aligned}
&\sup_{(z,\delta)\in\mathscr{S}} |f^*(z) - f(z)| = o(1),\\
&\sup_{(z,\delta)\in\mathscr{S}} |g^*(z) - g(z)| = o(1),\\
&\sup_{\substack{(z,\delta)\in\mathscr{S}\\ (\tilde{z},\tilde{\delta})\in\mathscr{S}}} |h^*(z,\tilde{z}) - h(z,\tilde{z})| = o(1),\\
&\sup_{\substack{(z,\delta)\in\mathscr{S}\\ (\tilde{z},\tilde{\delta})\in\mathscr{S}}} |\varphi^*(z,\tilde{z}) - \varphi(z,\tilde{z})| = o(1)
\end{aligned} \tag{3.4}$$

and

$$\left|E^*\left(f^k(z_1^*)g^l(z_1^*)\varphi^m(z_1^*,z_2^*)\right) - E\left(f^k(z_1)g^l(z_1)\varphi^m(z_1,z_2)\right)\right| = o(1) \tag{3.5}$$

for any fixed nonnegative integers k, l and m. Now (3.4) and (3.5) immediately give

$$E^*\left(f^{*k}(z_1^*)g^{*l}(z_1^*)\varphi^{*m}(z_1^*,z_2^*)\right) = E\left(f^k(z_1)g^l(z_1)\varphi^m(z_1,z_2)\right) + o(1) \tag{3.6}$$

uniformly on $\mathscr{E}_n$.

To show (3.1), we mimic the proof of Proposition 1 in Section 2. We first prove

$$\sup_x \left| P^*\left(\frac{n^{1/2}(\hat{\Lambda}^* - \hat{\Lambda})}{\hat{\sigma}^*} < x\right) - \Phi(x) - n^{-1/2}\phi(x)\left(\kappa_1^* x^2 + \kappa_2^*\right)\right| = o(n^{-1/2}) \tag{3.7}$$

uniformly on $\mathscr{E}_n$, where κ_j^* is the bootstrap analogue of κ_j, $j = 1, 2$. [Note that κ_1^* and κ_2^* depend on the samples (z_i, δ_i), $1 \le i \le n$, but not bootstrap samples.] Recall that $\hat{\Lambda}^* = \int_0^t d\hat{\bar{H}}_1^*/\hat{\bar{H}}^*$ and $\hat{\Lambda} = \int_0^t d\hat{\bar{H}}_1/\hat{\bar{H}}$. From (2.10), we may rewrite $n^{1/2}(\hat{\Lambda}^* - \hat{\Lambda})/\hat{\sigma}^*$ as

$$\frac{n^{1/2}(\hat{\Lambda}^* - \hat{\Lambda})}{\hat{\sigma}^*} = \theta^* + \gamma_3^*, \tag{3.8}$$

where θ^* and γ_3^* are the bootstrap analogues of θ and γ_3, respectively. The result (3.7) is proved in view of (3.8) and Lemma 3 in the Appendix.

To complete the proof of (3.1), we need to show (3.7) still holds if κ_1^*, κ_2^* and $\hat{\sigma}^*$ are replaced by κ_1, κ_2 and $\hat{\sigma}_G^*$, respectively. To this end, we first notice that κ_1 and κ_2 defined in (A.4) in the Appendix have similar form to those that appeared in (3.6). Therefore uniformly on $\mathscr{E}_n$, $\kappa_1^* = \kappa_1 + o(1)$ and $\kappa_2^* = \kappa_2 + o(1)$. Second, we write

$$\left|\frac{\hat{\sigma}^*}{\hat{\sigma}_G^*} - 1\right| \le \left|\frac{\hat{\sigma}^{*2}}{\hat{\sigma}_G^{*2}} - 1\right| \le \frac{1}{n\hat{\bar{H}}^*(t+)} \sup_{s \ge 0}\left|\hat{\bar{H}}^*(s) - \hat{\bar{H}}^*(s+)\right| = \frac{1}{n\hat{\bar{H}}^*(t+)} \sup_{1 \le i \le n} r_i^*,$$

where $r_i^* = \#\{j\colon z_j^* = z_i\}$, $1 \le i \le n$. Notice that $\{r_1^*, \dots, r_n^*\}$ has a multinomial distribution $\mathscr{M}(n; 1/n, \dots, 1/n)$, and it follows that

$$P^*\left(\sup_{1 \le i \le n} r_i^* \ge n^{1/6}\right) \le nP^*\left(r_1^* \ge n^{1/6}\right) \le n \sum_{j \ge n^{1/6}} \binom{n}{j}\left(\frac{1}{n}\right)^j = o(n^{-1/2}).$$

Since $\hat{\bar{H}}(t+) = \bar{H}(t+) + o(1)$ uniformly on $\mathscr{E}_n$ and $\bar{H}(t+) > 0$, we have $P^*(\hat{\bar{H}}^*(t+) \le n^{-1/6}) = o(n^{-1/2})$ uniformly on $\mathscr{E}_n$. Therefore

$$\begin{aligned} P^*\left(\left|\frac{\hat{\sigma}^*}{\hat{\sigma}_G^*} - 1\right| > n^{-1/2}(\log n)^{-1}\right) &\le P^*\left(\sup_{1\le i\le n} r_i^* \ge n^{1/6}\right) + P^*\left(\hat{\bar{H}}^*(t+) \le n^{-1/6}\right) \\ &= o(n^{-1/2}) \end{aligned}$$

uniformly on $\mathscr{E}_n$. Using the delta method, (3.7) still holds with $\hat{\sigma}^*$ replaced by $\hat{\sigma}_G^*$. The proof of (3.1) is thus complete. The proof of (3.2) is similar and hence is omitted. □

REMARK. We point out here, that in proving Theorem 4, only the convergence of the empirical distributions and a rather loose bound on the rate of the convergence are used.

Our next objective is to derive the Edgeworth expansion for the bootstrap of the Studentized K-M estimator and moment estimator. Because of the presence of ties in the bootstrap sample, we cannot directly copy the proofs of Theorems 2 and 3. [The validity of (2.18) relies on the fact that $P(r_i = 1) = 1$, due to the continuity of the distribution functions F and G, but evidently $P^*(r_1^*) = 1$ is wrong.] We shall turn to another approach which might be of independent interest. The main idea may be briefly explained as follows. Instead of sampling from the empirical observations, we consider a slightly smoothed bootstrap procedure that samples from a continuous distribution which is close to the empirical distribution. As noted in the above remark, the bootstrap accuracy obtained in (3.1) and (3.2) should also hold for this smoothed bootstrap. Now we can avoid the ties in this smoothed bootstrap sample and follow the proofs of Theorem 2 to identify the asymptotic differences among the bootstraps of the three estimators of survival function. Interesting enough, we find that the K-M estimator under the empirical measure equals the Nelson estimator under the smoothed bootstrap measure. Thus we can use the bootstrap accuracy for the Studentized Nelson estimator to show the bootstrap accuracy for the Studentized K-M and moment estimators.

The construction of a smoothed version of the bootstrap is conceptually simple. Instead of assigning probability $1/n$ to each observation, we assign probability $1/n$ to a small interval containing that observation. To be more specific, let ε_n be a small positive number such that $\varepsilon_n < \frac{1}{2}\min_{1\le i<j\le n}(|z_i - z_j|, |z_i - t|, |z_j - t|, n^{-n})$. This guarantees that the small intervals centered on each z_i with radius ε_n are all disjoint. (Note that t is fixed and F and G are both continuous. This implies $\min_i\{|z_i - t|\} > 0$ with probability 1.) Let $\{(z_i', \delta_i'),\ i = 1,\dots,n\}$ be the smoothed bootstrap sample sampled from a population with density function $(1/2n\varepsilon_n)\sum_{i=1}^n 1_{[|z'-z_i|\le\varepsilon_n,\ \delta'=\delta_i]}$ [as a function of (z', δ')]. It can be seen that this population, depending on the observations $\{(Z_i, \delta_i), i = 1, \dots, n\}$, is very close to the empirical distribution. We shall put a prime (′) on each statistic or function associated with $\{(z'_i, \delta'_i), i = 1, \dots, n\}$. Let

$$\hat{\hat{H}}'_1(s) = \frac{1}{n}\#\{j;\, z'_j \ge s,\, \delta'_j = 1\}, \qquad \hat{\hat{H}}'(s) = \frac{1}{n}\#\{j;\, z'_j \ge s\},$$

$$\overline{H}'_1(s) = P'(z'_1 \ge s,\, \delta'_1 = 1) \quad \text{and} \quad \overline{H}'(s) = P'(z'_1 \ge s).$$

Clearly

$$\sup_{s\ge 0}|\hat{\hat{H}}_1(s) - \overline{H}'_1(s)| \le n^{-1} \quad \text{and} \quad \sup_{s\ge 0}|\hat{\hat{H}}(s) - \overline{H}'(s)| \le n^{-1}.$$

Recall $\mathscr{E}_n$ in (3.3). We have uniformly on $\mathscr{E}_n$,

$$\sup_{s\ge 0}\left|\overline{H}'_1(s) - \overline{H}_1(s)\right| = o(1) \quad \text{and} \quad \sup_{s\ge 0}\left|\overline{H}'(s) - \overline{H}(s)\right| = o(1).$$

One then follows the proof of Theorem 4 line by line and concludes that

$$\sup_x\left|P'\left(\frac{n^{1/2}(\hat{\Lambda}' - \Lambda')}{\hat{\sigma}'_G} < x\right) - \Phi(x) - n^{-1/2}\phi(x)\left(\kappa_1 x^2 + \kappa_2\right)\right| = o(n^{-1/2})$$

and

$$\begin{aligned}\sup_x \Bigg|P'\left(\frac{n^{1/2}\left(\exp(-\hat{\Lambda}') - \exp(-\Lambda')\right)}{\exp(-\hat{\Lambda}')\,\hat{\sigma}'_G} < x\right) \\ - \Phi(x) - n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2\right)\Bigg| = o(n^{-1/2})\end{aligned} \tag{3.9}$$

uniformly on $\mathscr{E}_n$. Note that

$$\hat{\hat{F}}'_N = \exp(-\hat{\Lambda}'), \qquad \Lambda' = -\int_0^t \frac{d\overline{H}'_1}{\overline{H}'},$$

$$\hat{\hat{F}}' = \prod^*\left(1 - \frac{1}{n'_i}\right)^{\delta'_i} \quad \text{and} \quad \hat{\hat{F}}'_M = \prod^*\left(1 - \frac{1}{n'_i + 1}\right)^{\delta'_i},$$

where $n'_i = \#\{j;\, z'_j \ge z'_i\}$ and the product runs over all z'_i such that $z'_i \le t$. From (3.9), with the absence of ties in this smoothed bootstrap sampling, we can use Taylor expansion and mimic the proof of Theorems 2 and 3 to show

$$\begin{aligned}\sup_x \Bigg|P'\left(\frac{n^{1/2}\left(\hat{\hat{F}}' - \exp(-\Lambda')\right)}{\hat{\hat{F}}'\hat{\sigma}'_G} < x\right) \\ - \Phi(x) - n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2 + \frac{\sigma}{2}\right)\Bigg| = o(n^{-1/2})\end{aligned} \tag{3.10}$$

and

$$\sup_x \left| P'\left(\frac{n^{1/2}\left(\hat{\hat{F}}'_M - \exp(-\Lambda')\right)}{\hat{\hat{F}}'_M \hat{\sigma}'_G} < x \right) - \Phi(x) - n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2 - \frac{\sigma}{2}\right) \right| = o(n^{-1/2}) \tag{3.11}$$

uniformly on $\mathscr{E}_n$.

Now define $Z_i^* = Z_j$. If $|Z'_i - Z_j| \le \varepsilon_n$ for some $1 \le j \le n$, then $\{(Z_i^*, \delta_i^*), i = 1, \dots, n\}$ is clearly the ordinary bootstrap sample sampled uniformly over the observations $\{(Z_i, \delta_i), 1 \le i \le n\}$. With some easy calculations we have

$$\exp(-\Lambda') = \hat{\hat{F}}, \qquad \hat{\hat{F}}' = \hat{\hat{F}}^*, \qquad \hat{\hat{F}}'_M = \hat{\hat{F}}^*_M \quad \text{and} \quad \hat{\sigma}'_G = \hat{\sigma}^*_G.$$

In view of this key relationship, (3.10) and (3.11), we have

$$\sup_x \left| P^*\left(\frac{n^{1/2}\left(\hat{\hat{F}}^* - \hat{\hat{F}}\right)}{\hat{\hat{F}}^* \hat{\sigma}^*_G} < x \right) - \Phi(x) - n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2 + \frac{\sigma}{2}\right) \right| = o(n^{-1/2}) \tag{3.12}$$

and

$$\sup_x \left| P'\left(\frac{n^{1/2}\left(\hat{\hat{F}}^*_M - \hat{\hat{F}}\right)}{\hat{\hat{F}}^*_M \hat{\sigma}^*_G} < x \right) - \Phi(x) - n^{-1/2}\phi(x)\left(\tilde{\kappa}_1 x^2 + \tilde{\kappa}_2 - \frac{\sigma}{2}\right) \right| = o(n^{-1/2}) \tag{3.13}$$

uniformly on $\mathscr{E}_n$.

Now a comparison of the Edgeworth expansions in (3.12) with that in Theorem 2 gives the following result.

THEOREM 5. *Under the assumption in* (2.1), *we have*, *for arbitrary numbers* $\delta > 0$ *and* $\lambda > 0$,

$$P\left(\sup_x \left| P^*\left(\frac{n^{1/2}\left(\hat{\hat{F}}^* - \hat{\hat{F}}\right)}{\hat{\hat{F}}^* \hat{\sigma}^*_G} < x \right) - P\left(\frac{n^{1/2}\left(\hat{\hat{F}} - \overline{F}\right)}{\hat{\hat{F}} \hat{\sigma}_G} < x \right) \right| > \delta n^{-1/2} \right)$$
$$= o(n^{-\lambda}).$$

To obtain the Edgeworth expansion for the bootstrap of the Studentized moment estimator, note that the Studentized statistics in the left-hand side of (3.13) is only slightly different from our target statistic $n^{1/2}(\hat{\bar{F}}^*_M - \hat{\bar{F}}_M)/(\hat{\bar{F}}^*_M \hat{\sigma}^*_G)$. In fact, it follows from (2.17) and (2.21) that

$$P\left(\left|\hat{\bar{F}}_M - \hat{\bar{F}} - \frac{\bar{F}\sigma^2}{n}\right| > 2(n\log n)^{-1}\right) = o(n^{-1/2}). \tag{3.14}$$

Again applying the delta method together with (3.13) and (3.14), we deduce the following theorem.

THEOREM 6. *Under the assumption in* (2.1), *with probability* 1,

$$P^*\left(\frac{n^{1/2}\left(\hat{\bar{F}}^*_M - \hat{\bar{F}}_M\right)}{\hat{\bar{F}}^*_M \hat{\sigma}^*_G} < x\right) - P\left(\frac{n^{1/2}\left(\hat{\bar{F}}_M - \bar{F}\right)}{\hat{\bar{F}}_M \hat{\sigma}_G} < x\right)$$
$$= \phi(x)\sigma n^{-1/2} + o(n^{-1/2})$$

uniformly in x.

Theorems 4, 5 and 6 give the results on the asymptotic bootstrap accuracy for the Studentized statistics as we promised in Section 1. These results indicate that bootstrapping the Studentized K-M and Nelson estimators offer better alternatives than the commonly used normal approximations. The message is especially important and valuable when data are of moderate size or censoring is heavy so that the normal approximation could be inadequate. This issue together with the issue of coverage probabilities will be addressed in the next section.

4. The bootstrap accuracy for confidence intervals. In this section we demonstrate how to construct confidence intervals with better coverage probabilities. We also give an example with small sample size and heavy censorship. The simulations reported in the example show strong agreement with out theoretical findings and suggestions. The confidence intervals discussed in this section are all based on Studentized statistics.

To highlight the point, we shall focus our attention on the confidence intervals constructed with the K-M estimator. Similar conclusions hold for the Nelson estimator as well. Consider, for instance, the one-sided level $1-\alpha$ confidence intervals of the survival function. The normal-based confidence interval can be written as

$$I_{1-\alpha} \equiv \left(-\infty, \hat{\bar{F}} - z_\alpha \hat{\bar{F}}\hat{\sigma}_G n^{-1/2}\right),$$

where $z_\alpha = \Phi^{-1}(\alpha)$, and the bootstrap-based confidence interval is

$$J_{1-\alpha} \equiv \left(-\infty, \hat{\bar{F}} - \hat{q}_\alpha \hat{\bar{F}}\hat{\sigma}_G n^{-1/2}\right),$$

where $\hat{q}_\alpha$ is the α-quantile of the distribution of the bootstrap statistic $n^{1/2}(\hat{\bar{F}}^* - \hat{\bar{F}})/(\hat{\bar{F}}^*\hat{\sigma}_G^*)$ under P^*. More specifically, $\hat{q}_\alpha$ is the largest q which satisfies

$$P^*\left(\frac{n^{1/2}\left(\hat{\bar{F}}^* - \hat{\bar{F}}\right)}{\hat{\bar{F}}^*\hat{\sigma}_G^*} \le q\right) \le \alpha.$$

Clearly $\hat{q}_\alpha$ depends on the sample $\{(Z_i, \delta_i);\ 1 \le i \le n\}$. It is easy to see that

$$P\left(\bar{F}(t) \in I_{1-\alpha}\right) = 1 - \alpha + n^{-1/2}\phi(z_\alpha)\left(\tilde{\kappa}_1 z_\alpha^2 + \tilde{\kappa}_2\right) + o(n^{-1/2}).$$

Therefore, the normal-based confidence interval has an error of coverage probability on the order of $n^{-1/2}$. On the other hand, the bootstrap-based confidence interval has an error of coverage probability of $o(n^{-1/2})$, as proved in the following theorem. We shall also show that the bootstrap approximation of the quantiles is also better than the normal approximation. More specifically, let $q_{1-\alpha}$ be the $(1-\alpha)$-quantile of the distribution of the target statistic $n^{-1/2}(\hat{\bar{F}} - \bar{F})/(\hat{\bar{F}}\hat{\sigma}_G)$. Then it is clear that $z_{1-\alpha} = q_{1-\alpha} + O(n^{-1/2})$. However, as we shall show $\hat{q}_{1-\alpha} = q_{1-\alpha} + o(n^{-1/2})$ almost surely.

REMARK. When the distribution of target statistic is symmetric on the order of $n^{-1/2}$, the bootstrap-based two-sided equal-length confidence interval is no better than the normal-based two-sided confidence interval in terms of coverage probabilities [see, e.g., Hall (1992) for a detailed discussion]. However, the errors of lower and upper limits of the bootstrap-based confidence intervals are $o(n^{-1/2})$ and those of normal-based confidence intervals are $O(n^{-1/2})$. If we consider the coverage probabilities of nonsymmetric confidence intervals, the bootstrap-based intervals are usually better.

The following theorem summarizes the bootstrap accuracy in estimation of quantiles and the coverage probabilities of one-sided confidence intervals

THEOREM 7. *Under the assumption in* (2.1), *we have*

$$\hat{q}_\alpha = q_\alpha + o(n^{-1/2}) \tag{4.1}$$

almost surely and

$$P\left(\bar{F}(t) \in J_{1-\alpha}\right) = 1 - \alpha + o(n^{-1/2}) \tag{4.2}$$

for fixed $0 < \alpha < 1$.

PROOF. The proof is based on standard delta method. Let ε be an arbitrary positive number. From Theorem 5, we consider a set $\mathscr{E}_n$ with $P(\mathscr{E}_n^c) = o(n^{-\lambda})$ and $\lambda > 1$ such that

$$\sup_x \left| P^*\left(\frac{n^{1/2}\left(\hat{\bar{F}}^* - \hat{\bar{F}}\right)}{\hat{\bar{F}}^*\hat{\sigma}_G^*} < x\right) - P\left(\frac{n^{1/2}\left(\hat{\bar{F}} - \bar{F}\right)}{\hat{\bar{F}}\hat{\sigma}_G} < x\right)\right| \le \delta n^{-1/2} \tag{4.3}$$

on $\mathscr{E}_n$ for some δ, $0 < \delta < \varepsilon$. Let $x = q_\alpha + \varepsilon n^{-1/2}$. Since

$$P\left(\frac{n^{1/2}(\hat{\bar{F}} - \bar{F})}{\hat{\bar{F}}\hat{\sigma}_G} < q_\alpha + \varepsilon n^{-1/2}\right) = \alpha + \phi(z_\alpha)\varepsilon n^{-1/2} + o(n^{-1/2}),$$

it follows from (4.3) that

$$P^*\left(\frac{n^{1/2}(\hat{\bar{F}}^* - \hat{\bar{F}})}{\hat{\bar{F}}^*\hat{\sigma}_G^*} < q_\alpha + \varepsilon n^{-1/2}\right) > \alpha + \phi(z_\alpha)(\varepsilon - \delta)n^{-1/2} + o(n^{-1/2})$$

uniformly on $\mathscr{E}_n$. Because $\delta < \varepsilon$, the right-hand side of the above inequality is larger than α for large n. Therefore we conclude that on $\mathscr{E}_n$, $\hat{q}_\alpha \le q_\alpha + \varepsilon n^{-1/2}$ for large n. Likewise we can also show $\hat{q}_\alpha \ge q_\alpha - \varepsilon n^{-1/2}$ on $\mathscr{E}_n$ for large n. Since ε can be arbitrarily small, (4.1) follows from the Borel–Cantelli lemma. To show (4.2), write

$$\begin{aligned} P(\bar{F}(t) \in J_{1-\alpha}) &= P\left(\frac{n^{1/2}(\hat{\bar{F}} - \bar{F})}{\hat{\bar{F}}\hat{\sigma}_G} > \hat{q}_\alpha\right) \\ &\le P\left(\frac{n^{1/2}(\hat{\bar{F}} - \bar{F})}{\hat{\bar{F}}\hat{\sigma}_G} > q_\alpha - \varepsilon n^{-1/2}\right) + P(\hat{q}_\alpha \le q_\alpha - \varepsilon n^{-1/2}) \\ &= 1 - \alpha + \varepsilon\phi((q_\alpha))n^{-1/2} + o(n^{-1/2}). \end{aligned}$$

Since ε can be arbitrarily small, we have proved that the right-hand side of (4.2) is smaller than $1 - \alpha + o(n^{-1/2})$. The "bigger than" part can be proved analogously and (4.2) follows. □

We now consider one example in connection with the theoretical results we obtained so far in the bootstrap accuracy of confidence intervals. Let the distribution function of lifetime F be an exponential distribution with parameter 1, that is, $F(x) = 1 - \exp(-x)$ for $x > 0$, and let the distribution function of the censoring variable G be a uniform distribution over $(0, 1)$. Table 1 shows the coverage probabilities of one-sided confidence intervals constructed based on both normal approximations and bootstrap approximations.

We consider only two fixed time points, $t = 0.8$ and $t = 0.4$. It is clear that the former corresponds to a case with heavier censorship. The percentages of censored observations are roughly 36% when $t = 0.8$ and 27% when $t = 0.4$, according to the formula $\int_0^t (1 - x)e^{-x}\,dx$. (Note that the observations have no effect on the estimators if they are larger than t, regardless of whether

TABLE 1
*Coverage probabilities of confidence intervals**

	Time t = 0.8						**Time t = 0.4**					
	K-M		**Nelson**		**Moment**		**K-M**		**Nelson**		**Moment**	
Nominal	**N**	**B**	**N**	**B**	**N**	**B**	**N**	**B**	**N**	**B**	**N**	**B**
0.975	0.928	0.967	0.923	0.958	0.918	0.945	0.932	0.975	0.930	0.971	0.925	0.967
0.95	0.897	0.940	0.888	0.924	0.880	0.902	0.900	0.954	0.896	0.948	0.890	0.940
0.90	0.844	0.889	0.829	0.863	0.815	0.825	0.852	0.907	0.843	0.896	0.836	0.880
0.85	0.795	0.838	0.775	0.803	0.753	0.751	0.805	0.856	0.793	0.842	0.783	0.818
0.80	0.746	0.788	0.721	0.748	0.695	0.684	0.763	0.804	0.751	0.788	0.734	0.759

*"N" stands for normal approximation; "B" stands for bootstrap approximation. Sample size $n = 20$. The bootstrap approximations are based on 1000 repetitions.

they are censored or not.) Since $t = 0.8$ is close to 1, which is the upper bound of the support of the censoring distribution, one expects to have rougher normal approximations than the case when $t = 0.4$. Indeed, one can read from Table 1 that all three "N" columns at time $t = 0.4$ are uniformly better than the corresponding "N" columns at $t = 0.8$. On the other hand, in both cases, the bootstrap-based K-M and Nelson estimators provide uniformly better confidence intervals than the corresponding normal approximations.

It is fair to say, in the case of $t = 0.8$, the improvements of "B" over "N" are 80% for K-M at all nominal levels and over 60% for Nelson at all levels except levels 0.85 and 0.80. In the case of $t = 0.4$, the improvements of "B" over "N" are even greater: over 90% for both K-M and Nelson at all nominal levels. [Note that the percentage of improvement is calculated based on the normal-based error. For example, when $t = 0.4$ and K-M is used, the base error is $0.975 - 0.932 = 0.043$ at nominal level 0.975. The percentage of improvement of "B" over "N" is thus $1 - (0.975 - 0.975)/0.043$, which is 100% improvement.] On the contrary, one can also see from the table that the bootstrap approximation of the Studentized moment estimator does not always do a better job than the normal approximation. For example, when $t = 0.8$ and nominal levels are 0.85 and 0.80, the normal-based confidence intervals offer better coverage probabilities. This shows a good agreement with our theory (Theorem 6) in this paper.

APPENDIX

This Appendix consists of three lemmas.

LEMMA 1. *Under the assumption in* (2.1), *one can write*

$$P(\theta \le x) = \Phi(x) + n^{-1/2}\phi(x)\big(\kappa_1 x^2 + \kappa_2\big) + o(n^{-1/2}) \tag{A.1}$$

uniformly in x, where θ is defined in (2.11).

PROOF. Recall that $\varphi(z_i, z_j) = h(z_i, z_j) - g(z_i) - g(z_j)$. Then one can write

$$
\begin{aligned}
\theta &= \frac{n^{1/2}}{2\sigma}\left(\frac{2}{n^2}\sum_{i<j} h(z_i, z_j)\right)\left(1 - \frac{1}{2n\sigma^2}\sum_{i=1}^{n} f(z_i)\right) \\
&= \frac{(n-1)}{2n^{1/2}\sigma}\left(\frac{2}{n}\sum_{i=1}^{n} g(z_i) + \binom{n}{2}^{-1}\sum_{i<j}\varphi(z_i, z_j)\right)\left(1 - \frac{1}{2n\sigma^2}\sum_{i=1}^{n} f(z_i)\right) \\
&= \Bigg[\frac{n-1}{2\sigma n^{1/2}}\binom{n}{2}^{-1}\sum_{i<j}\Bigg(h(z_i, z_j) - \frac{1}{2\sigma^2}\big(g(z_i)f(z_j) + g(z_j)f(z_i)\big) \\
&\qquad\qquad - \frac{1}{n\sigma^2}E\big(g(z_1)f(z_1)\big)\Bigg)\Bigg] \\
&\quad + \frac{n-1}{4n^{3/2}\sigma^3}\binom{n}{2}^{-1}\sum_{i<j}\big(g(z_i)f(z_j) + g(z_j)f(z_i)\big) \\
&\quad - \frac{n-1}{2n^{5/2}\sigma^3}\sum_{k=1}^{n}\big(g(z_i)f(z_i) - E(g(z_i)f(z_i))\big) \\
&\quad - \frac{n-1}{4n^{3/2}\sigma^3}\binom{n}{2}^{-1}\sum_{i<j}\big(\varphi(z_i, z_j)(f(z_i) + f(z_j))\big) \\
&\quad - \frac{(n-1)(n-2)}{4n^{5/2}\sigma^3}\sum_{i=1}^{n}\left(f(z_i)\binom{n-1}{2}^{-1}\sum_{\substack{k<m \\ k\neq i\neq m}}\varphi(z_k, z_m)\right) \\
&= \frac{n-1}{n}\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\left(\tilde{h}(z_i, z_j) - \frac{1}{n\sigma^2}E\big(g(z_1)f(z_1)\big)\right) \\
&\quad + \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} \quad \text{(say)},
\end{aligned}
$$

where

$$
\tilde{h}(z_i, z_j) = h(z_i, z_j) - \frac{1}{2\sigma^2}\big(g(z_i)f(z_j) + g(z_j)f(z_i)\big) \tag{A.2}
$$

and I, II, III and IV are defined accordingly. Since $Ef(z_i) = 0$ and $Eg(z_1) = 0$, it follows that

$$
E\left(\tilde{h}(z_i, z_j) \mid (z_i, \delta_i)\right) = g(z_i).
$$

Since $(n^{1/2}/2\sigma)\binom{n}{2}^{-1}\sum_{i<j}\tilde{h}(z_i, z_j)$ is a standardized U-statistic of degree 2, which satisfies all the conditions of Theorem 1.2 in Bickel, Götze and Van

Zwet (1986) [notice that under the assumption in (2.1), function g is bounded and the distribution of $g(z_1)$ is continuous], we thus have

$$P\left(\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\tilde{h}(z_i, z_j) \le x\right) = \Phi(x) - \frac{\kappa}{6}\phi(x)n^{-1/2}(x^2-1) + o(n^{-1/2})$$

uniformly in x, where

$$\begin{aligned}\kappa &= \sigma^{-3}\left[Eg^3(z_1) + 3E\left(g(z_1)g(z_2)\left(\varphi(z_1, z_2) - \frac{1}{2\sigma^2}g(z_1)f(z_2) - \frac{1}{2\sigma^2}g(z_2)f(z_1)\right)\right)\right] \\ &= \sigma^{-3}\big(Eg^3(z_1) + 3E(g(z_1)g(z_2)\varphi(z_1, z_2)) - 3E(g(z_1)f(z_1))\big).\end{aligned} \tag{A.3}$$

Therefore,

$$\begin{aligned}&P\left(\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\left(\tilde{h}(z_i, z_j) - \frac{1}{n\sigma^2}E(g(z_1)f(z_1))\right) \le x\right) \\ &\quad = P\left(\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j}\tilde{h}(z_i, z_j) \le x + \frac{E(g(z_1)f(z_1))}{2n^{1/2}\sigma^3}\right) \\ &\quad = \Phi(x) + n^{-1/2}\phi(x)\big(\kappa_1 x^2 + \kappa_2\big) + o(n^{-1/2})\end{aligned}$$

uniformly in x, where

$$\begin{aligned}\kappa_1 &= -\frac{1}{6\sigma^3}\big(Eg^3(z_1) + 3E(g(z_1)g(z_2)\varphi(z_1, z_2)) - 3E(g(z_1)f(z_1))\big), \\ \kappa_2 &= \frac{1}{6\sigma^3}\big(Eg^3(z_1) + 3E(g(z_1)g(z_2)\varphi(z_1, z_2))\big).\end{aligned} \tag{A.4}$$

To end the proof of this lemma, we need to show that the error terms I, II, III and IV are asymptotically negligible. First, the functions f, g and φ are all bounded, and $Ef(z_1) = Eg(z_1) = E\varphi(z_1, z_2) = 0$. It is easy to check that $E\mathrm{I}^2 = O(n^{-2})$, $E\mathrm{II}^2 = O(n^{-2})$ and $E\mathrm{III}^2 = O(n^{-2})$. The term $E\mathrm{IV}^2 = O(n^{-2})$ follows from a careful calculation similar to that given in Callaret and Veraverbeke (1981). Using the Chebyshev inequality,

$$P\big(|\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}| > (n\log n)^{-1/2}\big) \le \frac{n^{-2}}{(n\log n)^{-1}} = o(n^{-1/2}).$$

The proof of (A.1) is complete by applying the delta method. □

LEMMA 2. *Under the assumption in* (2.1), *one can write*

$$P\big(|\gamma_3| \geq (n \log n)^{-1/2}\big) = o(n^{-1/2}),$$

where γ_3 *is given in* (2.11).

PROOF. We can write

$$\text{(A.5)}\qquad \begin{aligned} \gamma_3 &= \frac{n-1}{n}\gamma_2 \frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j} h(z_i, z_j) + \frac{n^{1/2}\gamma}{\hat{\sigma}} \\ &= (\mathbf{i}) + \frac{n^{1/2}\gamma}{\hat{\sigma}} \quad \text{(say)} \end{aligned}$$

Choose a small ε, $0 < \varepsilon < 1/2$. First, since h is a bounded function, a direct application of Hoeffding's inequality shows

$$P\left(\frac{n^{1/2}}{2\sigma}\binom{n}{2}^{-1}\sum_{i<j} h(z_i, z_j) \geq (\log n)^{1/2+\varepsilon}\right) = o(n^{-1/2}).$$

It follows that

$$P\left(|(\mathbf{i})| \geq \frac{1}{2}(n \log n)^{-1/2}\right) \leq P\left(|\gamma_2| \geq \frac{1}{c} n^{-1/2}(\log n)^{-1-\varepsilon}\right) + o(n^{-1/2}),$$

where γ_2 is defined in (2.9). For the first term of (2.9), $-\gamma_1/(2\sigma^2)$, one may use the proof of Theorem 1 in Lo and Singh (1986) to obtain

$$\text{(A.6)}\qquad P\left(\left|-\frac{\gamma_1}{2\sigma^2}\right| \geq \frac{1}{2} n^{-1/2}(\log n)^{-1-\varepsilon}\right) = o(n^{-1/2}).$$

[One can also treat γ_1 as a U-statistic to show (A.6).] For the second term of (2.9), note that

$$\text{(A.7)}\qquad \begin{aligned} &P(|\hat{\sigma}^2 - \sigma^2| > n^{-1/2}\log n) \\ &\quad \leq P\big(|\gamma_1| > n^{-1/2}\log n\big) + P\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(z_i)\right| > n^{-1/2}\log n\right) \\ &\quad = o(n^{-1/2}) \end{aligned}$$

by (A.6) and Bernstein's inequality. For n large enough, we have

$$\begin{aligned} &P\left(\left|\frac{(\hat{\sigma}-\sigma)^2(\hat{\sigma}+2\sigma)}{2\hat{\sigma}\sigma^2}\right| > \frac{1}{2}n^{-1/2}(\log n)^{-1-\varepsilon}\right) \\ &\quad = P\left(\left|\frac{(\hat{\sigma}^2-\sigma^2)^2(\hat{\sigma}+2\sigma)}{2\hat{\sigma}\sigma^2(\hat{\sigma}+\sigma)^2}\right| > \frac{1}{2}n^{-1/2}(\log n)^{-1-\varepsilon}\right) \\ &\quad \leq P\big(|\hat{\sigma}^2 - \sigma^2| > \sigma^2 n^{-1/4}(\log n)^{-(1+\varepsilon)/2}\big) \\ &\quad = o(n^{-1/2}), \end{aligned}$$

where the last equality is due to (A.7). It then turns out that

$$P\left(|(\mathbf{i})| \geq \frac{1}{2}(n \log n)^{-1/2}\right) = o(n^{-1/2}).$$

Now it suffices to show

$$P\left(|n^{1/2}\gamma| \geq \tfrac{1}{2}(n \log n)^{-1/2}\right) = P\left(|\gamma| \geq \tfrac{1}{2}n^{-1}(\log n)^{-1/2}\right) = o(n^{-1/2}).$$

Recall the definition of γ given in (2.5). First, one can employ the method in the proof of Theorem 1 in Lo and Singh (1986) to show that the tail probability inequality holds for the last two terms of γ. Second, the first four terms of γ consist of the average of iid random variables with mean zero and bounded variance multiplied by a factor n^{-1}, and hence their second moments are of order n^{-3}. The proof is complete using again the Chebyshev inequality. □

LEMMA 3. *Under the assumption in* (2.1), *we have*

$$\sup_x \left|P^*(\theta^* < x) - \Phi(x) - \left(\kappa_1^* x^2 + \kappa_2^*\right)\phi(x) n^{-1/2}\right| = o(n^{-1/2}) \tag{A.8}$$

and

$$P^*\left(|\gamma_3^*| > (n \log n)^{-1/2}\right) = o(n^{-1/2}) \tag{A.9}$$

uniformly on $\mathscr{E}_n$, *where* $\mathscr{E}_n$ *is given in* (3.3), *and* θ^* *and* γ_3^* *are the bootstrap analogues of* θ *and* γ_3, *respectively.*

PROOF. Similar to the expression of θ in Lemma 1, we may write

$$\theta^* = \frac{n-1}{n} \cdot \frac{n^{1/2}}{2\hat{\sigma}} \binom{n}{2}^{-1} \sum_{i<j} \left(\tilde{h}^*(z_i^*, z_j^*) - \frac{1}{n\hat{\sigma}^2} E^*(g^*(z_1^*) f^*(z_1^*))\right) + \mathrm{I}^* + \mathrm{II}^* + \mathrm{III}^* + \mathrm{IV}^*.$$

We first notice that the second moments of I^*, II^*, III^* and IV^* all take form like those that appeared in the left-hand side of (3.6), and they are on the order of n^{-2}. To be more specific, consider, for example, the term I^*. Since (3.6) holds uniformly on $\mathscr{E}_n$, by some calculations we know $n^2 E^*(\mathrm{I}^*)^2 = n^2 E\mathrm{I}^2 + o(1)$ uniformly on $\mathscr{E}_n$. So for large n, $n^2 E^*(\mathrm{I}^*)^2$ can be bounded by $2n^2 E\mathrm{I}^2$ on $\mathscr{E}_n$. The terms II^*, III^* and IV^* can be treated similarly. Then we can use the Markov inequality to show

$$P^*\left(|\mathrm{I}^*| + |\mathrm{II}^*| + |\mathrm{III}^*| + |\mathrm{IV}^*| > (n \log n)^{-1/2}\right) = o(n^{-1/2})$$

uniformly on $\mathscr{E}_n$. To prove (A.8), it remains to show

$$P^*\left(\frac{n^{1/2}}{2\hat{\sigma}} \binom{n}{2}^{-1} \sum_{i<j} \tilde{h}^*(z_i^*, z_j^*) \leq x\right) = \Phi(x) - \frac{\kappa^*}{6}\phi(x)(x^2 - 1) + o(^{-1/2}) \tag{A.10}$$

uniformly on $\mathscr{E}_n$, where κ^* is the bootstrap analogue of κ given in (A.3). It follows from (3.4) that

$$\sup_{\substack{(z,\,\delta)\in\mathscr{S}\\(\tilde z,\,\tilde\delta)\in\mathscr{S}}} \left|\tilde h^*(z,\tilde z) - \tilde h(z,\tilde z)\right| = o(1)$$

uniformly on $\mathscr{E}_n$. Therefore $\tilde h^*$ is a uniformly bounded function on $\mathscr{E}_n$ over all n because $\tilde h$ is bounded. Since g is bounded and $g(z_1)$ is continuously distributed, we have $Ee^{isg(z_1)} < 1$ for any $s \neq 0$. Again from (3.4) we know, for any fixed $a > 0$,

$$\sup_{|s|\le a} \left|E^* \exp\bigl(isg^*(z_1^*)\bigr) - E\exp\bigl(isg(z_1)\bigr)\right| = o(1)$$

uniformly on $\mathscr{E}_n$. Hence we have, for any fixed positive constants $a_1 < a_2$,

$$\sup_{a_1\le s\le a_2} \left|E^* \exp\bigl(isg^*(z_1^*)\bigr)\right| < 1 - c < 1$$

on $\mathscr{E}_n$ for large n, where c is some constant depending on a_1, a_2 and the distribution of $g(z_1)$ only. Now we can conclude (A.10) based on the Edgeworth expansion for U-statistics. [For a proof of this, the readers are referred to, for example, Lemma 1.2 in Bickel, Götze and van Zwet (1986).] Hence the proof of (A.8) is complete.

The proof of (A.9) again relies on the uniform convergences on $\mathscr{E}_n$ given in (3.4)–(3.6). Thanks also to the results on the bootstrap of the K-M estimator in Lo and Singh (1986), the procedures in the proof of Lemma 2 can be followed step by step to show (A.9). The details are omitted. □

REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

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

ALTSHULER, B. (1970). Theory for the measurement of competing risks in animal experiments. *Math. Biosci.* **6** 1–11.

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1986). The Edgeworth expansion for U-statistics of degree two. *Ann. Statist.* **14** 1463–1484.

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.

CALLARET, H. and VERAVERBEKE, N. (1981). The order of the normal approximation for a Studentized U-statistic. *Ann. Statist.* **9** 194–200.

CHEN, K. and YING, Z. (1996). A counterexample to a conjecture on Hall–Wellner band. *Ann. Statist.* **24** 641–646.

CUZICK, J. (1985). Asymptotic properties of censored linear rank tests. *Ann. Statist.* **13** 133–141.

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

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

GREENWOOD, M. (1926). The natural duration of cancer. In *Report on Public Health and Medical Subjects* **33** 1–26. Her Majesty's Stationary Office, London.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.* **16** 927–981.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, Berlin.

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

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

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

LO, S-H. and SINGH, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Related Fields* **71** 455–465.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179.

DEPARTMENT OF MATHEMATICS
HKUST
KOWLOON, HONG KONG

DEPARTMENT OF STATISTICS
DIVISION OF BIOSTATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027