The Annals of Statistics
1991, Vol. 19, No. 4, 1786–1812

LARGE-SAMPLE PROPERTIES FOR A GENERAL ESTIMATOR OF THE TREATMENT EFFECT IN THE TWO-SAMPLE PROBLEM WITH RIGHT CENSORING

BY XIAO-LI MENG, YIANNIS BASSIAKOS AND SHAW-HWA LO

Harvard University, Northeastern University and Columbia University

The estimation of the treatment effect in the two-sample problem with right censoring is of interest in survival analysis. In this article we consider both the location shift model and the scale change model. We establish the large-sample properties of a generalized Hodges–Lehmann type estimator. The strong consistency is established under the minimal possible conditions. The asymptotic normality is also obtained without imposing any conditions on the censoring mechanisms. As a by-product, we also establish a result for the oscillation behavior of the Kaplan–Meier process, which extends the Bahadur result for the empirical process to the censored case.

0. Introduction. In the analysis of survival data, it is often necessary to estimate the effect of a treatment or the difference between two treatments. In this setting, the lifetimes of a treatment and a control group are compared. The desired result is to show that a specific treatment extends the life of the patient, either in the sense that it extends it by a certain amount, or in the sense that it multiplies it by a certain factor. The first is the location shift model and the second is the scale change model. If the data are fully observable (no censoring), these problems have been extensively studied in the literature [see, e.g., Lehmann (1975)].

Let (x_i^0, u_i), $i = 1, \dots, n$, and (y_j^0, v_j), $j = 1, \dots, m$, be i.i.d. random variables, respectively. It is assumed here that all random variables are mutually independent. In the two-sample random censorship problem, one observes $\{(x_i, \varepsilon_i), i = 1, \dots, n\}$ and $\{(y_j, \gamma_j), j = 1, \dots, m\}$, where $x_i = \min\{x_i^0, u_i\}$, $\varepsilon_i = 1_{(x_i^0 \le u_i)}$, $y_j = \min\{y_j^0, v_j\}$ and $\gamma_j = 1_{(y_j^0 \le v_j)}$. Let F, U, G and V be the (left-continuous) c.d.f. for x^0, u, y^0 and v, respectively. In the location shift model, we assume $G(t) = F(t - \Delta)$ for all t, where Δ is an unknown parameter. It is then desired to estimate Δ.

Under this model, it is easy to see that Δ is the median of the distribution of $y^0 - x^0$. Assuming the median is unique, we can write $\Delta = K^{-1}(\frac{1}{2})$, where

$$K(\delta) = P\{y^0 - x^0 < \delta\} = \int_{-\infty}^{\infty} G(t + \delta)\, dF(t) \tag{0.1}$$

is the c.d.f. of $y^0 - x^0$. This suggests $\hat{K}^{-1}(\frac{1}{2})$ as an estimator for Δ, if we can find a consistent estimator $\hat{K}(\delta)$ for $K(\delta)$. For example, in the absence of

Received September 1989; revised September 1990.
AMS 1980 *subject classifications*. 62G05, 62E20.
Key words and phrases. Treatment effect, censoring, two-sample problem, Kaplan–Meier estimators, Hodges–Lehmann estimators, oscillation of the Kaplan–Meier process.

censoring, a natural estimator of $K(\delta)$ will be its sample analog $K_{nm}(\delta)$, which is obtained by substituting the corresponding empirical distributions for F and G in (0.1), respectively. The resulting estimator for Δ, $K_{nm}^{-1}(\frac{1}{2})$, is simply the sample median of $y_j^0 - x_i^0$, $i = 1, \dots, n$, $j = 1, \dots, m$, which is the well-known Hodges–Lehmann (1963) estimator. The success of this estimator depends (implicitly) upon the fact that one can estimate F and G consistently on the whole real line.

In the presence of censoring, one can construct $\hat{K}(\delta)$ by substituting the corresponding Kaplan–Meier estimators [Kaplan and Meier (1958)] F_n and G_m for F and G in (0.1), respectively. The resulting estimator $\hat{K}^{-1}(\frac{1}{2})$ is equivalent to the one proposed by Padgett and Wei (1982). However, since $\hat{K}(\delta)$ is not a consistent estimator for $K(\delta)$ for all δ due to the fact that the Kaplan–Meier estimators are only consistent up to certain uncensored values, as a result, $\hat{K}^{-1}(\frac{1}{2})$ might not be consistent either. In order to avoid the inconsistency of this estimator, it is assumed there that the support of the censoring extends to ∞ and so does the support of the survival curve. But, in practice, this infinite support assumption is usually violated in the presence of heavy censoring. In these cases the existing method could lead to inconsistent estimates. A way to overcome this difficulty is presented in Wei and Gail (1983), but there they impose strong conditions on the survival distribution, especially in the presence of heavy censoring (see particularly Remark 2 of their Theorem 2).

To avoid these strong conditions, Akritas (1986) introduced a general method for quantile estimation for the distribution of the differences in the two-sample problem, by truncating the samples before the upper limit of the support of the survival. He suggested there that this method could be used to estimate the shift parameter. But in order to do that, one must know the proportion of differences smaller than the shift. This is highly unlikely, since this proportion is a function of the unknown underlying distribution, unless the support of the censoring extends to ∞, in which case this proportion is equal to 0.5.

In order to derive a procedure that has the widest possible applicability—without all the previous conditions—we considered a procedure which in essence estimates this proportion from the data, and then uses the corresponding quantile to estimate the shift. Because the argument in the quantile is random, Akritas's (1986) results (especially Corollary 3.2) cannot be applied in establishing the large-sample properties of our estimator.

The purpose of this article is to prove rigorously the strong consistency and asymptotic normality of our estimator. Before we proceed we give a brief description of this method [for details see Bassiakos, Meng and Lo (1990)]. We introduce the truncated version of the K function as follows. Let T_1 and T_2 be two preselected constants such that $T_1 < \sup(H_F)$ and $T_2 < \sup(H_G)$, where $\sup(H_F)$ and $\sup(H_G)$ are the upper limits of the supports of the distributions H_F of x and H_G of y, respectively. Setting

$$(0.2) \qquad K_1(\delta) = P\{y^0 - x^0 < \delta,\, x^0 < T_1\} = \int_{-\infty}^{T_1} G(t + \delta)\, dF(t),$$

it follows that

$$K_1(\Delta) = \int_{-\infty}^{T_1} F(t)\,dF(t) \equiv P_1 \quad \text{(say)}. \tag{0.3}$$

This provides an estimator of Δ, $\hat{K}_1^{-1}(\hat{P}_1)$, where $\hat{K}_1(\delta)$ and $\hat{P}_1$ are sample analogs of $K_1(\delta)$ and P_1, respectively. Unfortunately, this estimator may not be consistent when $\Delta > T_2 - T_1$, since $\hat{K}_1(\delta)$ may not be consistent for $K_1(\delta)$ when $\delta > T_2 - T_1$. In fact, we can only show that $\min\{\hat{K}_1^{-1}(\hat{P}_1), T_2 - T_1\}$ is a consistent estimator for $\min\{\Delta, T_2 - T_1\}$ regardless of the censoring mechanism. To solve this problem, we can introduce another function:

$$K_2(\delta) = 1 - P\{y^0 - x^0 > \delta, y^0 < T_2\} = 1 - \int_{-\infty}^{T_2} F(t-\delta)\,dG(t), \tag{0.4}$$

and then

$$K_2(\Delta) = 1 - \int_{-\infty}^{T_2} G(t)\,dG(t) \equiv P_2 \quad \text{(say)}. \tag{0.5}$$

It will be shown that $\max\{\hat{K}_2^{-1}(\hat{P}_2), T_2 - T_1\}$ is a consistent estimator of $\max\{\Delta, T_2 - T_1\}$, where $\hat{K}_2(\delta)$ and $\hat{P}_2$ are sample analogs of $K_2(\delta)$ and P_2, because $\hat{K}_2(\delta)$ is a consistent estimator of $K_2(\delta)$ when $\delta \geq T_2 - T_1$. Since

$$\min\{\Delta, T_2 - T_1\} + \max\{\Delta, T_2 - T_1\} = \Delta + T_2 - T_1, \tag{0.6}$$

we can construct our estimator for Δ as

$$\begin{aligned} \hat{\Delta}_{nm} = {} & \min\left\{\hat{K}_1^{-1}\left(\hat{P}_1\right), T_2 - T_1\right\} \\ & + \max\left\{\hat{K}_2^{-1}\left(\hat{P}_2\right), T_2 - T_1\right\} - (T_2 - T_1). \end{aligned} \tag{0.7}$$

Notice that since the Kaplan–Meier estimator is left-continuous, $\hat{K}_1(\delta)$ is left-continuous while $\hat{K}_2(\delta)$ is right-continuous, and both are nondecreasing. The inverses utilized in (0.7) are defined in the following way.

For a nondecreasing function $g(x)$,

$$g^{-1}(y) = \inf\{x : g(x) > y\} \quad \text{for all } y, \tag{0.8}$$

if $g(x)$ is left-continuous; and

$$g^{-1}(y) = \inf\{x : g(x) \geq y\} \quad \text{for all } y, \tag{0.9}$$

if $g(x)$ is right-continuous.

1. Main results. The following two theorems give the large-sample properties of our estimator $\hat{\Delta}_{nm}$. The proofs of Theorems 1 and 2 will be given in Sections 2 and 4, respectively. Let $\overline{M}(t) = 1 - M(t)$ if $M(t)$ is a c.d.f.

THEOREM 1 (Strong consistency). *Suppose Δ is the unique solution for $K_i(\Delta) = P_i$, $i = 1, 2$. Then*

$$\hat{\Delta}_{nm} \to \Delta \ a.s. \quad \text{as } n \to \infty \text{ and } m \to \infty.$$

THEOREM 2 (Asymptotic normality). *Suppose $F(t)$ is continuous and*

$$d(T_0) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{-\infty}^{T_0} (F(t+\varepsilon) - F(t))\, dF(t) \tag{1.1}$$

exists and is positive, where $T_0 = \min(T_1, T_2 - \Delta)$. *Assume*

$$\lim_{n, m \to \infty} \frac{n}{n+m} = \lambda, \qquad 0 < \lambda < 1.$$

Then we have

$$\sqrt{n+m}\left(\hat{\Delta}_{mn} - \Delta\right) \to_{\mathscr{L}} N\left(0, \left[\frac{1}{\lambda}\sigma_1^2(T_0) + \frac{1}{1-\lambda}\sigma_2^2(T_0)\right] \frac{1}{d^2(T_0)}\right),$$

where

$$\sigma_1^2(T_0) = \frac{1}{4} \int_{-\infty}^{T_0} \frac{\left(\bar{F}^2(t) - \bar{F}^2(T_0)\right)^2}{\bar{F}(t)\bar{H}_F(t)}\, dF(t) \tag{1.2}$$

and

$$\sigma_2^2(T_0) = \frac{1}{4} \int_{-\infty}^{T_0+\Delta} \frac{\left(\bar{G}^2(t) - \bar{G}^2(T_0+\Delta)\right)^2}{\bar{G}(t)\bar{H}_G(t)}\, dG(t). \tag{1.3}$$

COROLLARY 1. *If $F(t)$ has density $f(t)$ and $\int_{-\infty}^{T_0} f^2(t)\, dt < \infty$, then the conclusion of Theorem 2 holds with $d(T_0) = \int_{-\infty}^{T_0} f^2(t)\, dt$.*

REMARK 1. In the absence of censoring, we can take $T_1 = T_2 = \infty$ (define $T_2 - T_1 = 0$). The resulting estimator $\hat{\Delta}_{nm}$ is essentially [up to an order of $O(1/(n+m))$] the well-known Hodges–Lehmann estimator, that is, the sample median of $y_j^0 - x_i^0$, $i = 1, \dots, n$, $j = 1, \dots, m$. The corresponding asymptotic variance of $\sqrt{n+m}(\hat{\Delta}_{nm} - \Delta)$ in this case is [assuming $\int_{-\infty}^{\infty} f^2(t)\, dt < \infty$ and noticing that $T_0 = \infty$ and $\bar{U}(t) \equiv \bar{V}(t) \equiv 1$]

$$\left[12\lambda(1-\lambda)\left(\int_{-\infty}^{\infty} f^2(t)\, dt\right)^2\right]^{-1},$$

which is exactly the same one as shown in Hodges and Lehmann (1963).

REMARK 2. Estimating the variance of $\hat{\Delta}_{nm}$ is quite straightforward. We can consistently estimate $\sigma_1^2(T_0)$ and $\sigma_2^2(T_0)$, by substituting the Kaplan–Meier estimators for $\bar{F}$ and $\bar{G}$ and the empirical distributions for $\bar{H}_F$ and $\bar{H}_G$ in (1.2) and (1.3), respectively. Finally, $d(T_0)$ can be easily estimated using Parceval's equality, as in Chapter 4 of Rao (1983). The only difference is that the empirical distribution weights are replaced by their Kaplan–Meier counterparts. This method allows us to estimate consistently the variance of the estimator without having to estimate the density, as required by the existing methods. The details about the variance estimation and other practical issues can be found in Bassiakos, Meng and Lo (1990).

For the scale change model $G(t) = F(t/\Theta)$, we can first transform the data into log-scale, which translates the scale change model into a location shift model. Then we can apply the method mentioned above to estimate $\Delta = \log \Theta$ directly.

On the other hand, we can attack the scale change problem directly by appropriate modification in the proposed method for the location shift problem. In fact, we can define the "scale change" versions of the $K_i(\delta)$, $i = 1, 2$, functions of (0.2) and (0.4), as $\mathscr{K}_i(\theta)$, by changing $t + \delta$ to $t\theta$, $t - \delta$ to t/θ and integrating from 0 instead of $-\infty$. Accordingly let $\mathscr{P}_i = \mathscr{K}_i(\Theta)$. Then the corresponding estimator of Θ can be constructed as [compare this with (0.7)]

$$\hat{\Theta}_{nm} = \min\left\{\hat{\mathscr{K}}_1^{-1}\left(\hat{\mathscr{P}}_1\right), T_2/T_1\right\}\max\left\{\hat{\mathscr{K}}_2^{-1}\left(\hat{\mathscr{P}}_2\right), T_2/T_1\right\}T_1/T_2,$$

where $\hat{\mathscr{K}}_i$ and $\hat{\mathscr{P}}_i$ are the sample (Kaplan–Meier) analogs of $\mathscr{K}_i$ and $\mathscr{P}_i$, $i = 1, 2$, respectively. Notice that the two methods give identical estimates.

The following two theorems on the large-sample properties of $\hat{\Theta}_{nm}$ can be established either directly, by applying the same type arguments that we will develop in the rest of this article for the location shift model, or indirectly, by applying Taylor expansion to $\hat{\Theta}_{nm} - \Theta = e^{\hat{\Delta}_{nm}} - e^{\Delta}$, and then applying Theorems 1 and 2, respectively.

THEOREM 1′. *Suppose* Θ *is the unique solution for* $\mathscr{K}_i(\Theta) = \mathscr{P}_i$, $i = 1, 2$. *Then*

$$\hat{\Theta}_{nm} \to \Theta \ a.s. \quad \text{as } n \to \infty \text{ and } m \to \infty.$$

THEOREM 2′. *Suppose* $F(t)$ *is continuous and*

$$\tilde{d}\left(\tilde{T}_0\right) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^{\tilde{T}_0} t\left(F(t+\varepsilon) - F(t)\right) dF(t)$$

exists and is positive, where $\tilde{T}_0 = \min\{T_1, T_2/\Theta\}$. *Assume*

$$\lim_{n, m \to \infty} \frac{n}{n+m} = \lambda, \qquad 0 < \lambda < 1.$$

Then

$$\sqrt{n+m}\left(\hat{\Theta}_{nm} - \Theta\right) \to_{\mathscr{L}} N\left(0, \Theta^2\left[\frac{1}{\lambda}\tilde{\sigma}_1^2\left(\tilde{T}_0\right) + \frac{1}{1-\lambda}\tilde{\sigma}_2^2\left(\tilde{T}_0\right)\right]\frac{1}{\tilde{d}^2\left(\tilde{T}_0\right)}\right),$$

where

$$\tilde{\sigma}_1^2\left(\tilde{T}_0\right) = \frac{1}{4}\int_0^{\tilde{T}_0} \frac{\left(\bar{F}^2(t) - \bar{F}^2\left(\tilde{T}_0\right)\right)^2}{\bar{F}(t)\bar{H}_F(t)} \, dF(t)$$

and

$$\tilde{\sigma}_2^2(\tilde{T}_0) = \frac{1}{4}\int_0^{\tilde{T}_0\Theta} \frac{\left(\overline{G}^2(t) - \overline{G}^2(\tilde{T}_0\Theta)\right)^2}{\overline{G}(t)\overline{H}_G(t)}\, dG(t).$$

COROLLARY 1'. *If $F(t)$ has density $f(t)$ and $\int_0^{\tilde{T}_0} tf^2(t)\,dt < \infty$, then the conclusion of Theorem 2' holds with $\tilde{d}(\tilde{T}_0) = \int_0^{\tilde{T}_0} tf^2(t)\,dt$.*

As a by-product, we also establish the following result which describes the oscillation behavior of the Kaplan–Meier process, and therefore extends the Bahadur result for empirical process [Bahadur (1966)] to the censored case. Since this result itself may be of independent interest, we state it as a proposition. The proof of this proposition will be given in Section 3.

PROPOSITION. *Suppose $F(t)$ is continuous on $(-\infty, \sup(H_F))$ and $F_n(t)$ is the Kaplan–Meier estimator for $F(t)$. Let $\{a_n\}$ be a sequence of positive constants such that*

$$a_n \sim C_0 n^{-1/2}(\log n)^q, \qquad n \to \infty,$$

for some constants $C_0 > 0$, and $q \geq \frac{1}{2}$. For $0 < \alpha < 1$, put

$$H_n(\alpha) = \sup_{|\alpha-\beta|\leq a_n} \left|\left[F_n(F^{-1}(\alpha)) - F_n(F^{-1}(\beta))\right] - (\alpha - \beta)\right|. \tag{1.4}$$

Then with probability 1,

$$H_n^*(T) \equiv \sup_{0<\alpha\leq F(T)} H_n(\alpha) = O\left(n^{-3/4}(\log n)^{(1+q)/2}\right), \qquad n \to \infty, \tag{1.5}$$

for any $T < \sup(H_F)$. Furthermore, if $F(t)$ satisfies the Lipshitz condition on $(-\infty, \sup(H_F))$, that is, there exists a universal constant $C > 0$, such that

$$|F(t) - F(s)| \leq C|t-s| \quad \text{for all } t, s \in (-\infty, \sup(H_F)),$$

then with probability 1,

$$\begin{aligned} &\sup_{t\leq T}\sup_{|t-s|\leq a_n} \left|[F_n(t) - F_n(s)] - [F(t) - F(s)]\right| \\ &\qquad = O\left(n^{-3/4}(\log n)^{(1+q)/2}\right), \qquad n \to \infty, \end{aligned} \tag{1.6}$$

for any $T < \sup(H_F)$.

2. Strong consistency. Before we prove the strong consistency of $\hat{\Delta}_{nm}$, that is, Theorem 1, we first need to prove the following lemma, which establishes the strong consistency of $\hat{K}_i(\delta)$ and $\hat{P}_i$, $i = 1, 2$.

LEMMA 2.1. *With probability* 1, *as* $n, m \to \infty$,

$$\sup_{\delta \le T_2 - T_1} \left|\hat{K}_1(\delta) - K_1(\delta)\right| \to 0,$$

$$\sup_{\delta \ge T_2 - T_1} \left|\hat{K}_2(\delta) - K_2(\delta)\right| \to 0$$

and

$$\hat{P}_i - P_i \to 0, \qquad i = 1, 2,$$

where $K_i(\delta)$ *and* P_i *are defined in* (0.2)–(0.5), *and* $\hat{K}_i(\delta)$ *and* $\hat{P}_i$, $i = 1, 2$, *are their sample* (*Kaplan–Meier*) *analogs, respectively.*

PROOF. We prove the lemma by establishing the following inequalities. Let

$$F_n^*(t) = F_n(t) - F(t) \tag{2.1}$$

and

$$G_m^*(t) = G_m(t) - G(t), \tag{2.2}$$

then we have

$$\sup_{\delta \le T_2 - T_1} \left|\hat{K}_1(\delta) - K_1(\delta)\right| \le \sup_{t \le T_2} |G_m^*(t)| + \sup_{t \le T_1} |F_n^*(t)| + |F_n^*(T_1)|, \tag{2.3}$$

$$\sup_{\delta \ge T_2 - T_1} \left|\hat{K}_2(\delta) - K_2(\delta)\right| \le \sup_{t \le T_1} |F_n^*(t)| + \sup_{t \le T_2} |G_m^*(t)| + |G_m^*(T_2)| \tag{2.4}$$

and

$$\left|\hat{P}_1 - P_1\right| \le 2 \sup_{t \le T_1} |F_n^*(t)| + |F_n^*(T_1)|, \tag{2.5}$$

$$\left|\hat{P}_2 - P_2\right| \le 2 \sup_{t \le T_2} |G_m^*(t)| + |G_m^*(T_2)|. \tag{2.6}$$

Then the lemma follows immediately from the fact

$$\sup_{t \le T_1} |F_n^*(t)| \to 0 \text{ a.s.} \quad \text{as } n \to \infty$$

and

$$\sup_{t \le T_2} |G_m^*(t)| \to 0 \text{ a.s.} \quad \text{as } m \to \infty.$$

To prove (2.3), first write

$$\begin{aligned} \hat{K}_1(\delta) - K_1(\delta) &= \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF_n(t) + \int_{-\infty}^{T_1} G(t + \delta)\, dF_n^*(t) \\ &\equiv D_1(\delta) + D_2(\delta). \end{aligned} \tag{2.7}$$

For $D_1(\delta)$, we have

$$\sup_{\delta \le T_2 - T_1} |D_1(\delta)| \le \sup_{\delta \le T_2 - T_1} \sup_{t \le T_1} |G_m^*(t + \delta)| \le \sup_{t \le T_2} |G_m^*(t)|. \tag{2.8}$$

For $D_2(\delta)$, applying a general integration by parts formula [see, e.g., Lemma 18.7 of Liptser and Shiryayev (1978)], we get

$$\begin{aligned} |D_2(\delta)| &= \left| G(T_1 + \delta) F_n^*(T_1) - \int_{-\infty}^{T_1} F_n^*(t+)\, dG(t + \delta) \right| \\ &\le |F_n^*(T_1)| + \sup_{t \le T_1} |F_n^*(t)|. \end{aligned} \tag{2.9}$$

Thus (2.3) follows from (2.7)–(2.9). The rest of the inequalities can be established by similar arguments. □

Now we are ready to prove Theorem 1, by considering three different situations: (i) $\Delta < T_2 - T_1$, (ii) $\Delta > T_2 - T_1$ and (iii) $\Delta = T_2 - T_1$. The basic idea of the proof is to transform the problem from $\hat{\Delta}_{nm}$ to $\hat{K}_i^{-1}$ and then to $\hat{K}_i$, $i = 1, 2$, and finally apply Lemma 2.1.

PROOF OF THEOREM 1. (i) $\Delta < T_2 - T_1$. In this case we will show

$$P\left\{\hat{\Delta}_{nm} \ne \hat{K}_1^{-1}(\hat{P}_1), \text{i.o.}\right\} = 0. \tag{2.10}$$

First, by the uniqueness condition and the monotonicity of $K_i(\delta)$, $i = 1, 2$, we have

$$P_i = K_i(\Delta) < K_i(T_2 - T_1), \qquad i = 1, 2. \tag{2.11}$$

It was seen in Lemma 2.1 that $\hat{K}_i(T_2 - T_1) \to K_i(T_2 - T_1)$ a.s. and $\hat{P}_i \to P_i$ a.s., $i = 1, 2$. Hence (2.11) implies that

$$P\left\{\hat{P}_i \ge \hat{K}_i(T_2 - T_1), \text{i.o.}\right\} = 0 \quad \text{for } i = 1, 2.$$

Thus it is easy to see [see, e.g., Serfling (1980), page 3] that

$$P\left\{\hat{K}_i^{-1}(\hat{P}_i) > T_2 - T_1, \text{i.o.}\right\} \le P\left\{\hat{P}_i \ge \hat{K}_i(T_2 - T_1), \text{i.o.}\right\} = 0 \quad \text{for } i = 1, 2,$$

which implies (2.10). Now for any $0 < \varepsilon < (T_2 - T_1) - \Delta$, again by the uniqueness condition, we have

$$K_1(\Delta - \varepsilon) < P_1 = K_1(\Delta) < K_1(\Delta + \varepsilon). \tag{2.12}$$

Since $\Delta \pm \varepsilon < T_2 - T_1$, $\hat{K}_1(\Delta \pm \varepsilon) \to K_1(\Delta \pm \varepsilon)$ a.s. by Lemma 2.1, and hence (2.12) implies [again see Serfling (1980)]

$$P\left\{\hat{K}_1^{-1}(\hat{P}_1) > \Delta + \varepsilon, \text{i.o.}\right\} \le P\left\{\hat{P}_1 \ge \hat{K}_1(\Delta + \varepsilon), \text{i.o.}\right\} = 0, \tag{2.13}$$

$$P\left\{\hat{K}_1^{-1}(\hat{P}_1) < \Delta - \varepsilon, \text{i.o.}\right\} \le P\left\{\hat{P}_1 \le \hat{K}_1(\Delta - \varepsilon), \text{i.o.}\right\} = 0. \tag{2.14}$$

Thus, by (2.10), (2.13) and (2.14), we have

$$P\{|\hat{\Delta}_{mn} - \Delta| > \varepsilon, \text{i.o.}\} = P\{|\hat{K}_1^{-1}(\hat{P}_1) - \Delta| > \varepsilon, \text{i.o.}\} = 0,$$

that is,

$$\hat{\Delta}_{nm} \to \Delta \text{ a.s.} \quad \text{as } n, m \to \infty.$$

(ii) $\Delta > T_2 - T_1$. By applying the same type of arguments, we can establish

$$P\{\hat{\Delta}_{nm} \neq \hat{K}_2^{-1}(\hat{P}_2), \text{i.o.}\} = 0, \tag{2.15}$$

and for any $0 < \varepsilon < \Delta - (T_2 - T_1)$,

$$P\{|\hat{K}_2^{-1}(\hat{P}_2) - \Delta| > \varepsilon, \text{i.o.}\} = 0. \tag{2.16}$$

Then the assertion follows from (2.15) and (2.16) immediately.

(iii) $\Delta = T_2 - T_1$. In this case, it is easy to check

$$\min\{\hat{K}_1^{-1}(\hat{P}_1) - \Delta, 0\} \leq \hat{\Delta}_{mn} - \Delta \leq \max\{\hat{K}_2^{-1}(\hat{P}_2) - \Delta, 0\}. \tag{2.17}$$

Now for any $\varepsilon > 0$, since $\Delta - \varepsilon < T_2 - T_1$ and $\Delta + \varepsilon > T_2 - T_1$, Lemma 2.1 implies $\hat{K}_1(\Delta - \varepsilon) \to K_1(\Delta - \varepsilon)$ a.s. and $\hat{K}_2(\Delta + \varepsilon) \to K_2(\Delta + \varepsilon)$ a.s. Therefore, by applying the same arguments as in (i) and (ii), we have

$$P\{\hat{K}_1^{-1}(\hat{P}_1) < \Delta - \varepsilon, \text{i.o.}\} \leq P\{\hat{P}_1 \leq \hat{K}_1(\Delta - \varepsilon), \text{i.o.}\} = 0,$$

$$P\{\hat{K}_2^{-1}(\hat{P}_2) > \Delta + \varepsilon, \text{i.o.}\} \leq P\{\hat{P}_2 \geq \hat{K}_2(\Delta + \varepsilon), \text{i.o.}\} = 0.$$

Now, (2.17) implies

$$\begin{aligned} &P\{|\hat{\Delta}_{nm} - \Delta| > \varepsilon, \text{i.o.}\} \\ &\quad \leq P\{\hat{K}_1^{-1}(\hat{P}_1) < \Delta - \varepsilon, \text{i.o.}\} + P\{\hat{K}_2^{-1}(\hat{P}_2) > \Delta + \varepsilon, \text{i.o.}\} = 0, \end{aligned}$$

which completes our proof. □

REMARK. The condition given in Theorem 1 is the minimal possible one, in the sense that if it is violated, then the shift parameter Δ is not identifiable. In fact, as we will show below, this condition is equivalent to the underlying distribution F being discrete with isolated atoms, in which case any shift less than the minimum spacing will not be identifiable. Since we need this equivalence in the proof of Theorem 2, we state it as a lemma.

LEMMA 2.2. *The uniqueness condition in Theorem* 1 *is violated if and only if*

$$F(t) = \sum_j b_j 1_{(a_j < t)} \quad \textit{for } t \leq T_0, \tag{2.18}$$

where $b_j > 0$ *and* $\min_{i \neq j} |a_i - a_j| > 0$.

PROOF. Let $\rho(F, s) \equiv \int_{-\infty}^{T_0} |F(t) - F(t - s)|\, dF(t)$. Then it is easy to check that the uniqueness condition is violated if and only if $\rho(F, s) = 0$ for some $s \neq 0$. Notice that for any (left-continuous) c.d.f., the following decomposition holds:

$$F(t) = \alpha F_c(t) + (1 - \alpha) F_d(t),$$

where F_c is a continuous c.d.f. and $F_d(t) = \sum_j b_j 1_{(a_j < t)}$ is a (left-continuous) discrete c.d.f. By the monotonicity of a c.d.f., it is not hard to show for $s \neq 0$,

$$\text{(i)} \qquad \rho(F_c, s) = 0 \Leftrightarrow F_c(t) \equiv 0 \quad \text{for } t \leq T_0,$$

$$\text{(ii)} \qquad \rho(F_d, s) = 0 \Leftrightarrow 0 < s < \min_{i \neq j} |a_i - a_j|.$$

The "if" part is an immediate consequence of (ii). The "only if" part can be easily shown by combining (i) and (ii) and the fact $\rho(F, s) \geq \alpha^2 \rho(F_c, s) + (1 - \alpha)^2 \rho(F_d, s)$. □

3. Oscillation behavior of the Kaplan–Meier process. The result on the oscillation behavior of the Kaplan–Meier process will be utilized in proving the asymptotic normality. To establish this result itself, that is, the proposition, we need the following three lemmas. The first two lemmas are well known in the literature. The first lemma plays a key role in the present proof, as well as in proving the asymptotic normality (Theorem 2).

LEMMA 3.1 [Lo and Singh (1986)]. *Suppose $F(t)$ is continuous. Using the notation given in Section 0, we have the following representation for $F_n(t)$:*

$$F_n(t) - F(t) = \frac{1}{n} \sum_{i=1}^{n} \xi_F(x_i, \varepsilon_i, t) + r_n(t), \tag{3.1}$$

where

$$\xi_F(x, \varepsilon, t) = \bar{F}(t) \left[\int_{-\infty}^{x \wedge t} \frac{d\bar{F}(s)}{\bar{H}_F(s) \bar{F}(s)} + \frac{\varepsilon 1_{(x < t)}}{\bar{H}_F(x)} \right] \tag{3.2}$$

and

$$\sup_{t \leq T} |r_n(t)| = O\left(n^{-3/4} (\log n)^{3/4}\right) \ a.s. \quad \text{for any } T < \sup(H_F).$$

LEMMA 3.2 [Bernstein, from Uspensky (1937)]. *Let $Z_1, \dots, Z_L$ be independent random variables with mean 0 and variance σ_l^2, and satisfying $P\{|Z_l| \leq B\} = 1$ for each l, where $B < \infty$. Then for any $d > 0$,*

$$P\left\{ \left| \sum_{l=1}^{L} Z_l \right| \geq Ld \right\} \leq 2 \exp\left\{ - \frac{(Ld)^2}{2\sum_{l=1}^{L} \sigma_l^2 + \frac{2}{3} BLd} \right\}$$

for all $L = 1, 2, \dots$.

LEMMA 3.3.

$$P\left\{ \bigcup_{n=1}^{\infty} \sup_t \left| F_n(F^{-1}(F(t))) - F_n(t) \right| > 0 \right\} = 0.$$

PROOF. First notice that $F_n(t)$ depends on t only through $1_{(x_i^0 < t)}$ [not $1_{x_i < t)}$] for all i such that $\varepsilon_i = 1$. Since $F^{-1}(F(t)) \geq t$ and $F(t)$ is left-continuous, we have

$$1_{(x_i^0 < F^{-1}(F(t)))} = 1_{(x_i^0 < t)} + 1_{(t \leq x_i^0 < F^{-1}(F(t)))}. \tag{3.3}$$

For any $0 < \pi < 1$, let $I_\pi = \{t\colon F(t) = \pi\}$ and $\Pi_0 = \{\pi\colon I_\pi$ contains at least two points$\}$. By the monotonicity of F, I_π is an interval for $\pi \in \Pi_0$, and the I_π's are disjoint for different π's. Thus Π_0 is at most a countable set. For $\pi \in \Pi_0$, let $J_\pi = [t_\pi, F^{-1}(\pi))$, where $t_\pi = \inf\{t,\ t \in I_\pi\}$. (Notice that $-\infty < t_\pi < \infty$ for $0 < \pi < 1$.) From (3.3), it is easy to see that

$$\bigcup_{n=1}^{\infty} \left\{ \sup_t |F_n(F^{-1}(F(t))) - F_n(t)| > 0 \right\} \subset \bigcup_{n=1}^{\infty} \bigcup_{i=1}^{n} \bigcup_{\pi \in \Pi_0} \{x_i^0 \in J_\pi\}. \tag{3.4}$$

Since $F(F^{-1}(F(t))) = F(t)$ [see Serfling (1980)] for all t, hence (recall that F is left-continuous)

$$\begin{aligned} P\{x_i^0 \in J_\pi\} &= P\{t_\pi \leq x_i^0 < F^{-1}(F(t_\pi))\} \\ &= F(F^{-1}(F(t_\pi))) - F(t_\pi) = 0 \quad \text{for all } \pi. \end{aligned}$$

The lemma follows immediately from (3.4), since the right-hand side of (3.4) is a countable union of zero-measure sets. □

PROOF OF THE PROPOSITION. The proof consists of three steps. First, by an appropriate partition of a subinterval of $[0, 1]$ and applying the Lo–Singh representation (Lemma 3.1) to $F_n(t)$, we can establish

$$H_n^*(T) \leq 3 \max_{1 \leq j \leq [\sqrt{n}]} \max_{-b_n \leq r \leq b_n} \left| \frac{1}{n} \sum_{i=1}^{n} Z_i(j, r) \right| + R_n. \tag{3.5}$$

Here $H_n^*(T)$ is defined in (1.5), $b_n = O(n^{1/4}(\log n)^q)$, $R_n = O((\log n/n)^{3/4})$ a.s. and for given j and r, $\{Z_i(j, r),\ i = 1, \dots, n\}$ are i.i.d. bounded random variables with mean 0 and variance $\sigma_{j,r}^2 = O(n^{-1/2}(\log n)^q)$.

Then, by applying the Bernstein inequality (Lemma 3.2), we have

$$P\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} Z_i(j, r) \right| \geq d_n \right\} = O(n^{-4}), \tag{3.6}$$

uniformly for all j and r, where $d_n = C_1 n^{-3/4}((\log n)^{(q+1)/2}$ for some choice of constant C_1.

Finally, by applying the Bonferroni inequality twice, we obtain

$$P\left\{ \max_{1 \le j \le [\sqrt{n}]} \max_{-b_n \le r \le b_n} \left| \frac{1}{n} \sum_{i=1}^{n} Z_i(j, r) \right| \ge d_n \right\} = O(n^{-2}). \tag{3.7}$$

Then the assertion follows immediately from the Borel–Cantelli lemma and inequality (3.5). [Notice that $R_n = O(n^{-3/4}(\log n)^{(1+q)/2})$, since for $q \ge \frac{1}{2}$, $\frac{1}{2}(q+1) \ge \frac{3}{4}$.]

To prove (3.5), first we notice that since $F(t)$ is continuous, we can find a T^* such that $T < T^* < \sup(H_F)$ and $F(T) < F(T^*)$. Let n be large enough so that $a_n < F(T^*) - F(T)$. For $0 < \alpha < 1$, let $A_n(\alpha) = F_n(F^{-1}(\alpha)) - \alpha$, then we have

$$H_n(\alpha) = \sup_{|\alpha - \beta| \le a_n} |A_n(\alpha) - A_n(\beta)|, \tag{3.8}$$

where $H_n(\alpha)$ is defined in (1.4).

For $j = 0, 1, \dots, [\sqrt{n}]$, let $\alpha_j = (j/[\sqrt{n}])F(T)$ and $\alpha_{[\sqrt{n}]+1} = F(T^*)$. Now, notice that for any α, β, $0 < \alpha \le F(T)$ and $|\beta - \alpha| \le a_n$, if both α and β belong to the same interval $[\alpha_j, \alpha_{j+1}]$, then

(i) if $0 \le j \le [\sqrt{n}] - 1$, then $\alpha_{j+1} - \alpha_j = (1/[\sqrt{n}])F(T) < a_n$ for n large enough. Thus

$$\begin{aligned} |A_n(\alpha) - A_n(\beta)| &\le |A_n(\alpha) - A_n(\alpha_{j+1})| + |A_n(\alpha_{j+1}) - A_n(\beta)| \\ &\le 2 \max_{1 \le j \le [\sqrt{n}]} H_n(\alpha_j); \end{aligned} \tag{3.9}$$

(ii) if $j = [\sqrt{n}]$, then $\alpha = F(T) = \alpha_{[\sqrt{n}]}$ and hence

$$|A_n(\beta) - A_n(\alpha)| \le H_n(\alpha_{[\sqrt{n}]}) \le \max_{1 \le j \le [\sqrt{n}]} H_n(\alpha_j). \tag{3.10}$$

On the other hand, if α and β belong to two different intervals, say, $\alpha \in [\alpha_r, \alpha_{r+1}]$ and $\beta \in [\alpha_j, \alpha_{j+1}]$, where $1 \le j + 1 \le r \le [\sqrt{n}]$ (otherwise interchange j and r), then $|\alpha - \beta| \le a_n$ implies $\alpha_r - \alpha_{j+1} \le a_n$. Hence

$$\begin{aligned} |A_n(\alpha) - A_n(\beta)| &\le |A_n(\alpha) - A_n(\alpha_r)| + |A_n(\alpha_r) - A_n(\alpha_{j+1})| \\ &\quad + |A_n(\alpha_{j+1}) - A_n(\beta)| \\ &\le 3 \max_{1 \le j \le [\sqrt{n}]} H_n(\alpha_j). \end{aligned} \tag{3.11}$$

Thus, by (3.8)–(3.11), we have established

$$H_n^*(T) = \sup_{0 < \alpha \le F(T)} H_n(\alpha) \le 3 \max_{1 \le j \le [\sqrt{n}]} H_n(\alpha_j). \tag{3.12}$$

Now, for fixed j, $1 \le j \le [\sqrt{n}]$, let

$$\eta_{j,r} = \alpha_j + r\frac{a_n}{b_n} \quad \text{for } r = -b_n, \dots, b_n, \text{ where } b_n = \left[C_0 n^{1/4} (\log n)^q\right]$$

and

$$\phi_{j,r} = \left| A_n(\eta_{j,r}) - A_n(\alpha_j) \right|. \tag{3.13}$$

Notice that for any β, $|\beta - \alpha_j| \le a_n \Rightarrow \eta_{j,r} \le \beta \le \eta_{j,r+1}$ for some r, $-b_n \le r < b_n$. By the monotonicity of $F_n(F^{-1})$, we have

$$\begin{aligned} \left| A_n(\alpha_j) - A_n(\beta) \right| &= \left| \left[F_n\left(F^{-1}(\alpha_j)\right) - \alpha_j \right] - \left[F_n(F^{-1}(\beta)) - \beta \right] \right| \\ &\le \max\Big\{ \left| \left[F_n\left(F^{-1}(\alpha_j)\right) - \alpha_j \right] - \left[F_n\left(F^{-1}(\eta_{j,r})\right) - \eta_{j,r+1} \right] \right|, \\ &\qquad \left| \left[F_n\left(F^{-1}(\alpha_j)\right) - \alpha_j \right] - \left[F_n\left(F^{-1}(\eta_{j,r+1})\right) - \eta_{j,r} \right] \right| \Big\} \\ &\le \max\{\phi_{j,r}, \phi_{j,r+1}\} + |\eta_{j,r+1} - \eta_{j,r}| \\ &\le \max_{-b_n \le r \le b_n} \phi_{j,r} + \frac{a_n}{b_n}. \end{aligned} \tag{3.14}$$

Thus

$$H_n(\alpha_j) \le \max_{-b_n \le r \le b_n} \phi_{j,r} + \frac{a_n}{b_n} \quad \text{for } j = 1, 2, \dots, \left[\sqrt{n}\right]. \tag{3.15}$$

For fixed j and r, applying the Lo–Singh representation (Lemma 3.1), we have [notice that $F(F^{-1}(\alpha)) = \alpha$, since F is continuous]

$$\begin{aligned} \phi_{j,r} &= \left| A_n(\alpha_j) - A_n(\eta_{j,r}) \right| \\ &= \left| \left[F_n\left(F^{-1}(\alpha_j)\right) - \alpha_j \right] - \left[F_n\left(F^{-1}(\eta_{j,r})\right) - \eta_{j,r} \right] \right| \\ &= \Bigg| \frac{1}{n} \sum_{i=1}^{n} \left[\xi_i\left(F^{-1}(\alpha_j)\right) - \xi_i\left(F^{-1}(\eta_{j,r})\right) \right] \\ &\qquad + r_n\left(F^{-1}(\alpha_j)\right) + r_n\left(F^{-1}(\eta_{j,r})\right) \Bigg|, \end{aligned} \tag{3.16}$$

where $\xi_i(t)$ is an abbreviation of $\xi_F(x_i, \varepsilon_i, t)$, given in (3.2).

Let $Z_i(j,r) = \xi_i(F^{-1}(\alpha_j)) - \xi_i(F^{-1}(\eta_{j,r}))$. Notice that, since

$$\left| r_n\left(F^{-1}(\alpha_j)\right) + r_n\left(F^{-1}(\eta_{j,r})\right) \right| \le 2 \sup_{\alpha \le F(T^*)} \left| r_n\left(F^{-1}(\alpha)\right) \right| \le 2 \sup_{t \le T^*} |r_n(t)|,$$

we have from (3.16)

$$\phi_{j,r} \le \left| \frac{1}{n} \sum_{i=1}^{n} Z_i(j,r) \right| + 2 \sup_{t \le T^*} r_n(t). \tag{3.17}$$

Combining (3.12), (3.15) and (3.17), we establish (3.5), that is,

$$H_n^* \le 3 \max_{1 \le j \le [\sqrt{n}]} \max_{-b_n \le r \le b_n} \left| \frac{1}{n} \sum_{i=1}^{n} Z_i(j,r) \right| + R_n,$$

where, with probability 1,

$$R_n = 3\frac{a_n}{b_n} + 6\sup_{t \le T^*} r_n(t) = O(n^{-3/4}) + O\left(\left(\frac{\log n}{n}\right)^{3/4}\right) = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right).$$

Notice that for fixed j and r, $\{Z_i(j,r),\ i = 1,\dots,n\}$ are i.i.d. bounded [with the bound $B_0 = 2(\int_{-\infty}^{T^*} dF(t)/\bar{F}(t)\bar{H}_F(t) + 1/\bar{H}_F(T^*)) < \infty$] random variables with mean 0 and variance, say $\sigma_{j,r}^2$. Therefore, by applying the Bernstein inequality (Lemma 3.2), with the choice of $d = d_n = C_1 n^{-3/4}(\log n)^{(q+1)/2}$, where $C_1 > 0$ is a constant to be specified later, we obtain

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n Z_i(j,r)\right| \ge d_n\right\} \le 2e^{-\theta_{j,r}^{(n)}}, \tag{3.18}$$

where

$$\theta_{j,r}^{(n)} = \frac{nd_n^2}{2\sigma_{j,r}^2 + \frac{2}{3}B_0 d_n}. \tag{3.19}$$

Straightforward algebra shows (assume $\eta_{j,r} \le \alpha_j$, otherwise exchange $\eta_{j,r}$ and α_j).

$$\begin{aligned} |Z_i(j,r)| &= \left|\xi_i\big(F^{-1}(\alpha_j)\big) - \xi_i\big(F^{-1}(\eta_{j,r})\big)\right| \\ &= \left|(\eta_{j,r} - \alpha_j)\left[-\int_0^{\alpha_j} \frac{1_{(\alpha \le F(x_i))}}{(1-\alpha)\bar{H}_F(F^{-1}(\alpha))}\,d\alpha + \frac{\varepsilon_i 1_{(F(x_i^0) < \alpha_j)}}{\bar{H}_F(x_i)}\right]\right| \\ &\quad + \left|(1 - \eta_{j,r})\left[-\int_{\eta_{j,r}}^{\alpha_j} \frac{1_{(\alpha \le F(x_i))}}{(1-\alpha)\bar{H}_F(F^{-1}(\alpha))}\,d\alpha + \frac{\varepsilon_i 1_{(\eta_{j,r} \le F(x_i^0) < \alpha_j)}}{\bar{H}_F(x_i)}\right]\right| \\ &\le B_1|\eta_{j,r} - \alpha_j| + B_2 1_{(\eta_{j,r} \le \omega < \alpha_j)}, \end{aligned} \tag{3.20}$$

where $\omega = F(x_i^0)$ is a uniform random variable on $[0, 1]$,

$$B_1 = \frac{1}{\bar{H}_F(T^*)}\left(\frac{2}{\bar{F}(T^*)} + 1\right) \quad \text{and} \quad B_2 = \frac{1}{\bar{H}_F(T^*)}.$$

Therefore,

$$\begin{aligned} \sigma_{j,r}^2 = E\left[Z_i^2(j,r)\right] &\le 2\Big(B_1^2(\eta_{j,r} - \alpha_j)^2 + B_2^2 P\{\eta_{j,r} \le \omega < \alpha_j\}\Big) \\ &= 2\Big(B_1^2(\alpha_j - \eta_{j,r})^2 + B_2^2|\alpha_j - \eta_{j,r}|\Big), \end{aligned}$$

which implies that $\sigma_{j,r}^2 \le C_2 a_n$ for some universal constant C_2 and n large enough, since $|\alpha_j - \eta_{j,r}| \le a_n$. Thus from (3.19) we have

$$\theta_{j,r}^{(n)} \ge \frac{nd_n^2}{2C_2 a_n + \frac{2}{3}B_0 d_n} \sim \frac{C_1^2(\log n)}{2C_2 C_0 + \frac{2}{3}B_0\, C_1\left[(\log n)^{(1-q)/2}/n^{1/4}\right]}.$$

Therefore, if we choose C_1 large enough so that $C_1^2/4C_2C_0 > 4$, then for all n sufficiently large $\theta_{j,r}^{(n)} \ge C_1^2 \log n/4C_0C_2 > 4(\log n)$. It then follows from (3.18) that there exists N^* such that

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z_i(j,r)\right| \geq d_n\right\} \leq 4n^{-4}$$

for all $1 \leq j \leq [\sqrt{n}\,]$, $|r| \leq b_n$ and $n > N^*$. Thus we have established (3.6). The expression (3.7) is an immediate consequence of (3.6) and the Bonferroni inequality. This completes the proof of the first part of the proposition.

In order to prove the second part of the proposition, first notice that $|t - s| \leq a_n$ implies $|F(t) - F(s)| \leq C|t - s| \leq Ca_n \equiv a_n^*$. Therefore,

$$\begin{aligned}
&\sup_{t\leq T}\sup_{|t-s|\leq a_n} \left|[F_n(t) - F_n(s)] - [F(t) - F(s)]\right| \\
&\qquad \leq \sup_{t\leq T}\sup_{|F(t)-F(s)|\leq a_n^*} \left|[F_n(t) - F_n(s)] - [F(t) - F(s)]\right|.
\end{aligned}$$

By Lemma 3.3, with probability 1, the right-hand side above is equal to

$$\begin{aligned}
&\sup_{t\leq T}\sup_{|F(t)-F(s)|\leq a_n^*} \left|[F_n(F^{-1}(F(t))) - F_n(F^{-1}(F(s)))] - [F(t) - F(s)]\right| \\
&\quad \leq \sup_{0<\alpha\leq F(T)}\sup_{|\alpha-\beta|\leq a_n^*} \left|[F_n(F^{-1}(\beta)) - F_n(F^{-1}(\alpha))] - (\alpha - \beta)\right| \\
&\quad = O\left(n^{-3/4}(\log n)^{(1+q)/2}\right),
\end{aligned}$$

by the first part of the proposition (obviously F is continuous here), which completes our proof. □

4. Asymptotic normality.

The following four lemmas are needed in proving Theorem 2. The first two are well known in the literature. The third one is a consequence of the oscillation behavior of the Kaplan–Meier process we proved in Section 3. The fourth lemma gives the i.i.d. sums representation of $\hat{K}_i(\delta) - K_i(\delta)$ and $\hat{P}_i - P_i$, $i = 1, 2$, which will play an important role in proving Theorem 2.

LEMMA 4.1 [Berry and Esséen, from Serfling (1980)]. *Let $\{Z_l,\ l = 1, \dots, L\}$ be independent random variables with mean μ_l and variance $\sigma_l^2 > 0$. Let $S_L = \sum_{l=1}^{L}(Z_l - \mu_l)/(\sum_{l=1}^{L}\sigma_l^2)^{1/2}$, $\Phi_L(t) = P\{S_L \leq t\}$. Then*

$$\sup_t |\Phi_L(t) - \Phi(t)| \leq \frac{C\sum_{l=1}^{L} E|Z_l - \mu_l|^3}{\left(\sum_{l=1}^{L}\sigma_l^2\right)^{3/2}},$$

where C is a universal constant and $\Phi(t)$ is the standard normal c.d.f.

LEMMA 4.2 [Lo and Singh (1986)]. *If F is continuous, then with probability 1,*

$$\sup_{t\leq T} |F_n(t) - F(t)| = O\left(\left(\frac{\log n}{n}\right)^{1/2}\right) \quad a.s., \qquad n \to \infty,$$

for any $T < \sup(H_F)$.

LEMMA 4.3. *Suppose $F(t)$ is continuous and $0 < \lim_{n,m\to\infty}(n/L) = \lambda < 1$, where $L = n + m$. Then with probability* 1,

$$\sup_{\delta \le T_2 - T_1} \left| \int_{-\infty}^{T_1} (G_m(t+\delta) - G(t+\delta))\, d(F_n(t) - F(t)) \right| = O\left(\left(\frac{\log L}{L} \right)^{3/4} \right), \tag{4.1}$$

$$\sup_{\delta \ge T_2 - T_1} \left| \int_{-\infty}^{T_2} (F_n(t-\delta) - F(t-\delta))\, d(G_m(t) - G(t)) \right| = O\left(\left(\frac{\log L}{L} \right)^{3/4} \right), \tag{4.2}$$

$$\left| \int_{-\infty}^{T_1} (F_n(t) - F(t))\, d(F_n(t) - F(t)) \right| = O\left(\left(\frac{\log n}{n} \right)^{3/4} \right), \tag{4.3}$$

$$\left| \int_{-\infty}^{T_2} (G_m(t) - G(t))\, d(G_m(t) - G(t)) \right| = O\left(\left(\frac{\log m}{m} \right)^{3/4} \right). \tag{4.4}$$

PROOF. We prove these results by establishing upper bounds for the left-hand sides of (4.1)–(4.4). Since $F(t)$ is continuous [so is $G(t) = F(t - \Delta)$], we can find a collection of $\{t_i, 1 \le i \le I_n\}$ such that $-\infty < t_1 < t_2 < \cdots < t_{I_n} = T_1$ and $F(t_i) = (i/I_n)F(T_1)$, where $I_n = [(n/\log n)^{1/2}] + 1$. Similarly, for given δ, we can find $-\infty < s_1 < s_2 < \cdots < s_{J_m} = T_1$, where $J_m = [(m/\log m)^{1/2}] + 1$, such that $G(s_j + \delta) = (j/J_m)G(T_1 + \delta)$. Denote $t_0 = s_0 = -\infty$. Let $-\infty = r_0 \le r_1 \le \cdots \le r_{K_{nm}}$ be the collection of $\{t_i, 0 \le i \le I_n\}$ and $\{s_j, 0 \le j \le I_m\}$ after ordering, where $K_{nm} = I_m + J_n + 1$. Because $\{r_k, 0 \le k \le K_{nm}\}$ is a finer partition of $(-\infty, T_1)$ than $\{t_i, 0 \le i \le I_n\}$ and $\{s_j, 0 \le j \le J_m\}$, by the monotonicity of F and G, we have

$$F(r_k) - F(r_{k-1}) \le \frac{1}{I_n} \le \left(\frac{\log n}{n} \right)^{1/2}$$

and

$$G(r_k + \delta) - G(r_{k-1} + \delta) \le \frac{1}{J_M} \le \left(\frac{\log m}{m} \right)^{1/2} \quad \text{for } k = 1, \dots, K_{nm}.$$

Therefore, by applying the simple equality

$$\int_a^b f(t)\, dg(t) = \int_a^b (f(t) - f(b))\, dg(t) + f(b)(g(b) - g(a)),$$

we obtain

$$\begin{aligned}
&\sup_{\delta \le T_2 - T_1} \left| \int_{-\infty}^{T_1} (G_m(t+\delta) - G(t+\delta))\, d(F_n(t) - F(t)) \right| \\
&\quad \le \sup_{\delta \le T_2 - T_1} \left| \sum_{k=1}^{K_{nm}} \int_{r_{k-1}}^{r_k} [G_m^*(t+\delta) - G_m^*(r_k+\delta)]\, dF_n^*(t) \right| \\
&\qquad + \sup_{\delta \le T_2 - T_1} \left| \sum_{k=1}^{K_{nm}} G_m^*(r_k+\delta)[F_n^*(r_k) - F_n^*(r_{k-1})] \right| \\
&\quad \le 2 \sup_{\delta \le T_2 - T_1} \max_{1 \le k \le K_{nm}} \sup_{r_{k-1} \le t \le r_k} |G_m^*(t+\delta) - G_m^*(r_k+\delta)| \\
&\qquad + K_{nm} \sup_{\delta \le T_2 - T_1} \sup_{t \le T_1} |G_m^*(t+\delta)| \max_{1 \le k \le K_{nm}} |F_n^*(r_k) - F_n^*(r_{k-1})| \\
&\quad \le 2 \sup_{t \le T_2} \sup_{\substack{|G(s) - G(t)| \\ \le (\log m/m)^{1/2}}} |G_m^*(s) - G_m^*(t)| \\
&\qquad + K_{nm} \sup_{t \le T_2} |G_m^*(t)| \sup_{t \le T_1} \sup_{\substack{|F(s) - F(t)| \\ \le (\log n/n)^{1/2}}} |F_n^*(s) - F_n^*(t)|,
\end{aligned}$$

where $F_n^*(t)$ and $G_m^*(t)$ are defined in (2.1) and (2.2), respectively.

By Lemma 3.3, with probability 1, the right-hand-side above is equal to

$$2H_m^*(G, T_2) + K_{nm} \sup_{t \le T_2} |G_m^*(t)| H_n^*(F, T_1),$$

where

$$\begin{aligned}
&H_n^*(F, T_1) \\
&\quad = \sup_{0 \le \alpha < F(T_1)} \sup_{\substack{|\alpha - \beta| \\ \le (\log n/n)^{1/2}}} \left| [F_n(F^{-1}(\alpha)) - F_n(F^{-1}(\beta))] - (\alpha - \beta) \right|
\end{aligned}$$

and

$$\begin{aligned}
&H_m^*(G, T_2) \\
&\quad = \sup_{0 \le \alpha < G(T_2)} \sup_{\substack{|\alpha - \beta| \\ \le (\log m/m)^{1/2}}} \left| [G_m(G^{-1}(\alpha)) - G_m(G^{-1}(\beta))] - (\alpha - \beta) \right|.
\end{aligned}$$

Thus we have established (with probability 1)

$$\begin{aligned}
&\sup_{\delta \le T_2 - T_1} \left| \int_{-\infty}^{T_1} (G_m(t+\delta) - G(t+\delta))\, d(F_n(t) - F(t)) \right| \\
&\qquad \le 2H_m^*(G, T_2) + K_{nm} \sup_{t \le T_2} |G_m^*(t)| H_n^*(F, T_1).
\end{aligned} \tag{4.5}$$

By similar arguments as above, we can establish (with probability 1)

$$\begin{aligned}(4.6)\qquad &\sup_{\delta \ge T_2 - T_1}\left|\int_{-\infty}^{T_2}(F_n(t-\delta) - F(t-\delta))\,d(G_m(t) - G(t))\right| \\ &\quad\le 2H_n^*(F, T_1) + K_{nm}\sup_{t\le T_1}|F_n^*(t)|H_m^*(G, T_2).\end{aligned}$$

By the proposition on oscillation behavior of the Kaplan–Meier process (for $q = \frac{1}{2}$), we have

$$H_n^*(F, T_1) = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right) \quad \text{and} \quad H_m^*(G, T_2) = O\left(\left(\frac{\log m}{m}\right)^{3/4}\right).$$

Since $0 < \lambda < 1$, by Lemma 4.2,

$$K_{nm}\sup_{t\le T_1}|F_n^*(t)| = O(1) \quad \text{and} \quad K_{nm}\sup_{t\le T_2}|G_m^*(t)| = O(1).$$

Thus (4.1) and (4.2) follow immediately from (4.5) and (4.6), respectively.
Similarly (and more easily, in fact), we can establish (with probability 1)

$$\begin{aligned}(4.7)\qquad &\left|\int_{-\infty}^{T_1}(F_n(t) - F(t))\,d(F_n(t) - F(t))\right| \\ &\quad\le \left\{2 + I_n\sup_{t\le T_1}|F_n^*(t)|\right\}H_n^*(F, T_1)\end{aligned}$$

and

$$\begin{aligned}(4.8)\qquad &\left|\int_{-\infty}^{T_2}(G_m(t) - G(t))\,d(G_m(t) - G(t))\right| \\ &\quad\le \left\{2 + J_m\sup_{t\le T_2}|G_m^*(t)|\right\}H_m^*(G, T_2).\end{aligned}$$

By applying the same arguments as above, (4.3) and (4.4) follow from (4.7) and (4.8), respectively. □

The Lo–Singh representation of the Kaplan–Meier estimator (Lemma 3.1) is critical in proving the following lemma.

LEMMA 4.4. *Under the conditions of Lemma 4.3, we have the following representations*:

$$\begin{aligned}(4.9)\qquad \hat{K}_1(\delta) - K_1(\delta) &= \frac{1}{m}\sum_{j=1}^{m}\eta_j^{(G)}(T_1, \delta) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\left[\eta_i^{(F)}(T_1, \delta) - G(T_1 + \delta)\xi_i^{(F)}(T_1)\right] + R_1(\delta),\end{aligned}$$

with

$$\sup_{\delta \le T_1 - T_1} |R_1(\delta)| = O\left(\left(\frac{\log L}{L}\right)^{3/4}\right) \quad a.s.,$$

$$(4.10) \qquad \begin{aligned} \hat{K}_2(\delta) - K_2(\delta) = & -\frac{1}{n}\sum_{i=1}^{n} \eta_i^{(F)}(T_2 - \delta, \delta) \\ & + \frac{1}{m}\sum_{j=1}^{m} \left[\eta_j^{(G)}(T_2 - \delta, \delta) - F(T_2 - \delta)\xi_j^{(G)}(T_2)\right] + R_2(\delta), \end{aligned}$$

with

$$\sup_{\delta \ge T_2 - T_1} |R_2(\delta)| = O\left(\left(\frac{\log L}{L}\right)^{3/4}\right) \quad a.s.,$$

and

$$(4.11) \qquad \hat{P}_1 - P_1 = \frac{F(T_1)}{n}\sum_{i=1}^{n} \xi_i^{(F)}(T_1) + O\left(\left(\frac{\log n}{n}\right)^{3/4}\right),$$

$$(4.12) \qquad \hat{P}_2 - P_2 = -\frac{G(T_2)}{m}\sum_{j=1}^{m} \xi_j^{(G)}(T_2) + O\left(\left(\frac{\log m}{m}\right)^{3/4}\right),$$

where $\xi_i^{(F)}(t)(\xi_j^{(G)}(t))$ *is an abbreviation of* $\xi_F(x_i, \varepsilon_i, t)(\xi_G(y_j, \gamma_j, t))$ *of* (3.2), *and*

$$(4.13) \qquad \eta_i^{(F)}(a, \delta) = \int_{-\infty}^{a} \xi_i^{(F)}(t)\, dG(t + \delta),$$

$$(4.14) \qquad \eta_j^{(G)}(a, \delta) = \int_{-\infty}^{a} \xi_j^{(G)}(t + \delta)\, dF(t).$$

PROOF. Integration by parts gives

$$(4.15) \qquad \begin{aligned} \hat{K}_1(\delta) - K_1(\delta) &= \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF(t) + \int_{-\infty}^{T_1} G(t + \delta)\, dF_n^*(t) \\ &\quad + \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF_n^*(t) \\ &= \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF(t) + G(T_1 + \delta)F_n^*(T_1) \\ &\quad - \int_{-\infty}^{T_1} F_n^*(t)\, dG(t + \delta) + \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF_n^*(t). \end{aligned}$$

By Lemma 3.1, for $t \le T_1$, $\delta \le T_2 - T_1$ (hence $\delta + t \le T_2$), one can write

$$F_n^*(t) = \frac{1}{n} \sum_{i=1}^{n} \xi_F(x_i, \varepsilon_i, t) + r_n^{(F)}(t), \tag{4.16}$$

$$G_m^*(t + \delta) = \frac{1}{m} \sum_{j=1}^{m} \xi_G(y_j, \gamma_j, t + \delta) + r_m^{(G)}(t + \delta), \tag{4.17}$$

where, with probability 1,

$$\sup_{t \le T_1} |r_n^{(F)}(t)| = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right) \tag{4.18}$$

and

$$\sup_{t \le T_2} |r_m^{(G)}(t)| = O\left(\left(\frac{\log m}{m}\right)^{3/4}\right). \tag{4.19}$$

Combining (4.15)–(4.17), we obtain (4.9), with

$$\begin{aligned}\sup_{\delta \le T_2 - T_1} |R_1(\delta)| &= \sup_{\delta \le T_2 - T_1} \left| \int_{-\infty}^{T_1} r_m^{(G)}(t + \delta)\, dF(t) - \int_{-\infty}^{T_1} r_n^{(F)}(t)\, dG(t + \delta) \right. \\ &\qquad \left. + G(T_1 + \delta) r_n^{(F)}(T_1) + \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF_n^*(t) \right| \\ &\le 2 \sup_{t \le T_1} |r_n^{(F)}(t)| + \sup_{t \le T_2} |r_m^{(G)}(t)| \\ &\qquad + \sup_{\delta \le T_2 - T_1} \left| \int_{-\infty}^{T_1} G_m^*(t + \delta)\, dF_n^*(t) \right| \\ &= O\left(\left(\frac{\log L}{L}\right)^{3/4}\right),\end{aligned}$$

by (4.18), (4.19) and (4.1) of Lemma 4.3. The expression (4.10) can be established in the same way after a variable transformation ($t - \delta \to t$) in the integrals. Representations (4.11) and (4.12) are immediate consequences of Lemmas 3.1 and 4.3 [(4.3) and (4.4)] and the following two identities:

$$\begin{aligned}\hat{P}_1 - P_1 &= F(T_1)(F_n(T_1) - F(T_1)) \\ &\qquad + \int_{-\infty}^{T_1} (F_n(t) - F(t))\, d(F_n(t) - F(t))\end{aligned}$$

and

$$\begin{aligned}\hat{P}_2 - P_2 &= -G(T_2)(G_m(T_2) - G(T_2)) \\ &\qquad + \int_{-\infty}^{T_2} (G_m(t) - G(t))\, d(G_m(t) - G(t)).\end{aligned}$$

□

Now we have the tools to prove Theorem 2. Again the main idea of the proof is to transform the problem from $\hat{\Delta}_{nm}$ to $\hat{K}_i^{-1}$ and then to $\hat{K}_i$, $i = 1, 2$, and apply Lemma 4.4 and the Berry–Esséen bound (Lemma 4.1). As in the proof of consistency, we pursue the proof by considering three different cases: (i) $\Delta < T_2 - T_1$, (ii) $\Delta > T_2 - T_1$ and (iii) $\Delta = T_2 - T_1$.

PROOF OF THEOREM 2. (i) $\Delta < T_2 - T_1$. Since F is continuous, by Lemma 2.2, the uniqueness condition of Theorem 1 holds. Therefore, $P\{\hat{\Delta}_{nm} \neq \hat{K}_1^{-1}(\hat{P}_1), \text{ i.o.}\} = 0$ as showed in the proof of Theorem 1. Hence we only need to establish the asymptotic normality for $\sqrt{n+m}\,(\hat{K}_1^{-1}(\hat{P}_1) - \Delta)$. Let $L = n + m$ and $\lambda_L = n/L$. Fix t. Let A be a normalizing constant to be specified later and put

$$\varphi_L^{(1)}(t) \equiv P\left\{ \frac{\sqrt{L}\left(\hat{K}_1^{-1}(\hat{P}_1) - \Delta\right)}{A} < t \right\}.$$

Then, using Serfling's lemma [Serfling (1980), page 3], we have

$$\begin{aligned}\varphi_L^{(1)}(t) &= P\left\{\hat{P}_1 < \hat{K}_1(\Delta + tAL^{-1/2})\right\}\\ &= P\Big\{-\left[K_1(\Delta + tAL^{-1/2}) - K_1(\Delta)\right]\\ &\qquad < \left[\hat{K}_1(\Delta + tAL^{-1/2}) - K_1(\Delta + tAL^{-1/2})\right] - \left(\hat{P}_1 - P_1\right)\Big\}.\end{aligned}$$

For L large enough so that $\Delta + tAL^{-1/2} < T_2 - T_1$, by Lemma 4.4 [(4.9) and (4.11)], we have

$$\varphi_L^{(1)}(t) = P\left\{ -C_L(t) < S_L + O\left(\frac{(\log L)^{3/4}}{L^{1/4}} \right) \right\},$$

where

$$\begin{aligned}C_L(t) &= \frac{L\left[K_1(\Delta + tAL^{-1/2}) - K_1(\Delta)\right]}{\left[\sum_{l=1}^L V(Z_l)\right]^{1/2}},\\ S_L &= \frac{\sum_{l=1}^L Z_l}{\left[\sum_{l=1}^L V(Z_l)\right]^{1/2}},\\ Z_l &= \begin{cases} \dfrac{-\eta_{i,L}^{(F)} + \Omega_L \xi_i^{(F)}(T_1)}{\lambda_L}, & l = i,\ i = 1, \dots, n,\\[2ex] \dfrac{\eta_{j,L}^{(G)}}{1 - \lambda_L}, & l = j + n,\ j = 1, \dots, m,\end{cases}\end{aligned}$$

and

$$\begin{aligned}\eta_{i,L}^{(F)} &= \eta_i^{(F)}(T_1, \Delta + tAL^{-1/2}),\\ \eta_{j,L}^{(G)} &= \eta_j^{(G)}(T_1, \Delta + tAL^{-1/2})\end{aligned}$$

[then $\eta_i^{(F)}(\alpha,\delta)$ and $\eta_j^{(G)}(\alpha,\delta)$ are defined in (4.13) and (4.14), respectively] and

$$\Omega_L = G(T_1 + \Delta + tAL^{-1/2}) - F(T_1) = F(T_1 + tAL^{-1/2}) - F(T_1).$$

Since $O((\log L)^{3/4}/L^{1/4}) \to 0$ a.s., by Slutsky's theorem, in order to prove $\varphi_L^{(1)}(t) \to \Phi(t)$, it is enough to show

$$\tilde{\varphi}_L^{(1)}(t) \equiv P\{-C_L(t) < S_L\} \to \Phi(t).$$

It is easy to check that

$$|\tilde{\varphi}_L^{(1)}(t) - \Phi(t)| \le |P\{S_L \le -C_L(t)\} - \Phi(-C_L(t))| + |\Phi(t) - \Phi(C_L(t))|.$$

By the Berry–Esséen bound (Lemma 4.1), we have

$$|P\{S_L \le -C_L(t)\} - \Phi(-C_L(t))| \le C\frac{\sum_{l=1}^L E|Z_l|^3}{\left(\sum_{l=1}^L V(Z_l)\right)^{3/2}} = O(L^{-1/2}), \tag{4.21}$$

the last equation above is true because Z_l is a bounded random variable for all $l = 1,\dots,L$. It remains to investigate whether $C_L(t) \to t$. First, notice that since F is continuous, we have $\Omega_L = F(T_1 + tAL^{-1/2}) - F(T_1) \to 0$ as $L \to \infty$. Thus we have

$$\begin{aligned}\frac{1}{L}\sum_{l=1}^L V(Z_l) &= \frac{1}{L}\left[\frac{n}{\lambda_L^2}V(\eta_1^{(F)} - \Omega_L\xi_1) + \frac{m}{(1-\lambda_L)^2}V(\eta_1^{(G)})\right]\\ &\to \frac{1}{\lambda}V(\eta_1^{(F)}) + \frac{1}{1-\lambda}V(\eta_1^{(G)}),\end{aligned} \tag{4.22}$$

where $\eta_1^{(F)} = \eta_1^{(F)}(T_1,\Delta)$ and $\eta_1^{(G)} = \eta_1^{(G)}(T_1,\Delta)$. Now writing

$$\begin{aligned}C_L(t) &= tA\frac{K_1(\Delta + tAL^{-1/2}) - K_1(\Delta)}{tAL^{-1/2}}\,\frac{1}{\left((1/L)\sum_{l=1}^L V(Z_l)\right)^{1/2}}\\ &\to tAK_1'(\Delta)\frac{1}{\left[(1/\lambda)V(\eta_1^{(F)}) + (1/(1-\lambda))V(\eta_1^{(G)})\right]^{1/2}} \quad \text{as } L \to \infty,\end{aligned}$$

we see, if we choose

$$A = \frac{\left[(1/\lambda)V(\eta_1^{(F)}) + (1/(1-\lambda))V(\eta_1^{(G)})\right]^{1/2}}{K_1'(\Delta)},$$

then

$$C_L(t) \to t \quad \text{as } L \to \infty.$$

Hence

$$|\Phi(C_L(t)) - \Phi(t)| \to 0 \quad \text{for all } t \text{ as } L \to \infty.$$

It is easy to verify that [note $T_0 = \min(T_1, T_2 - \Delta) = T_1$], $V(\eta_1^{(F)}) = \sigma_1^2(T_0)$, $V(\eta_1^{(G)}) = \sigma_2^2(T_0)$ and $K_1'(\Delta) = d(T_0)$, where $\sigma_i^2(T_0)$, $i = 1, 2$, are given by (1.2) and (1.3), and $d(T_0)$ is defined in (1.1). For example, $V(\eta_1^{(F)})$ is obtained below, following similar work by Efron (1967), page 847]:

$$\begin{aligned}
V(\eta_1^{(F)}) &= \int_{-\infty}^{T_0}\int_{-\infty}^{T_0}\Gamma(s,t)\,dG(t+\Delta)\,dG(t+\Delta)\\
&= 2\int_{z_0}^{1}\int_{z_0}^{z}\int_{z_0}^{s}\frac{st}{z^2\overline{U}(\overline{F}^{-1}(z))}\,dt\,ds\,dz\\
&= \frac{1}{4}\int_{z_0}^{1}\frac{(z^2-z_0^2)^2}{z^2\overline{U}(\overline{F}^{-1}(z))}\,dz\\
&= \frac{1}{4}\int_{-\infty}^{T_0}\frac{(\overline{F}^2(t)-\overline{F}^2(T_0))^2}{\overline{F}^2(t)\overline{U}(t)}\,dF(t) = \sigma_1^2(T_0),
\end{aligned}$$

where $\Gamma(s,t)$ is the covariance structure of the $\xi_i^{(F)}(t)$'s [see Lo and Singh (1986)] and $z_0 = \overline{F}(T_0)$. This completes our proof for the case $\Delta < T_2 - T_1$.

(ii) $\Delta > T_2 - T_1$. In this case we have

$$P\left\{\hat{\Delta}_{nm} \neq \hat{K}_2^{-1}(\hat{P}_2), \text{i.o.}\right\} = 0.$$

Fix t. For L large enough so that $\Delta + tAL^{-1/2} > T_2 - T_1$ (again, A is a positive constant to be specified), by Lemma 4.4 [(4.10) and (4.12)], we obtain

$$\begin{aligned}
\varphi_L^{(2)}(t) &\equiv P\left\{\frac{\sqrt{L}\left(\hat{K}_2^{-1}(\hat{P}_2)-\Delta\right)}{A} \le t\right\}\\
&= 1 - P\left\{\hat{K}_2^{-1}(\hat{P}_2) > \Delta + tAL^{-1/2}\right\}\\
&= 1 - P\left\{\hat{P}_2 > \hat{K}_2(\Delta + tAL^{-1/2})\right\}\\
&= 1 - P\left\{-C_L^*(t) > \frac{\sum_{l=1}^{L} Z_l^*}{\left(\sum_{l=1}^{L} V(Z_l^*)\right)^{1/2}} + O\left(\frac{(\log L)^{3/4}}{L^{1/4}}\right)\right\},
\end{aligned}$$

where

$$Z_l^* = \begin{cases}
\dfrac{-\eta_i^{(F)}\left(T_2 - (\Delta + tAL^{-1/2}), \Delta + tAL^{-1/2}\right)}{\lambda_L}, & l = i,\ i = 1,\dots,n,\\[2ex]
\dfrac{\eta_j^{(G)}\left(T_2 - (\Delta + tAL^{-1/2}), \Delta + tAL^{-1/2}\right) - \Omega_L^*\xi_j^{(G)}(T_2)}{1-\lambda_L}, & l = j+n,\ j = 1,\dots,m,
\end{cases}$$

$$C_L^*(t) = \frac{L\left[K_2(\Delta + tAL^{-1/2}) - K_2(\Delta)\right]}{\left[\sum_{l=1}^{L} V(Z_l^*)\right]^{1/2}}$$

and

$$\Omega_L^* = F\left(T_2 - \Delta - tAL^{-1/2}\right) - G(T_2)$$
$$= F\left(T_2 - \Delta - tAL^{-1/2}\right) - F(T_2 - \Delta).$$

By applying the same arguments as in (i), we only need to choose

$$A = \frac{\left[(1/\lambda)V\left(\eta_1^{(F)}(T_2 - \Delta, \Delta)\right) + (1/(1-\lambda))V\left(\eta_1^{(G)}(T_2 - \Delta, \Delta)\right)\right]^{1/2}}{K_2'(\Delta)},$$

which will guarantee $C_L^*(t) \to t$ for all t. Again, we can verify that (recall that $T_2 - \Delta < T_1$ implies $T_2 - \Delta = T_0$), $V(\eta_1^{(F)}(T_2 - \Delta, \Delta)) = \sigma_1^2(T_0)$, $V(\eta_1^{(G)}(T_2 - \Delta, \Delta)) = \sigma_2^2(T_0)$ and $K_2'(\Delta) = d(T_0)$.

(iii) $\Delta = T_2 - T_1$. In this case we have

$$(4.23)\quad \hat{\Delta}_{nm} - \Delta = \min\left\{\hat{K}_1^{-1}(\hat{P}_1) - \Delta, 0\right\} + \max\left\{\hat{K}_2^{-1}(\hat{P}_2) - \Delta, 0\right\}.$$

From this expression it should be clear that $\hat{\Delta}_{nm} - \Delta$ is either positive, if the maximum in (4.23) is $\hat{K}_2^{-1}(\hat{P}_2) - \Delta$ and the minimum is 0, or negative, if the minimum in (4.23) is $\hat{K}_1^{-1}(\hat{P}_1) - \Delta$ and the maximum is 0. So, if we let $S_{nm} = (\hat{K}_1^{-1}(\hat{P}_1) - \Delta)(\hat{K}_2^{-1}(\hat{P}_2) - \Delta)$, it is easy to see that for $t \le 0$,

$$(4.24)\quad \left\{\hat{\Delta}_{nm} - \Delta < t,\ S_{nm} \ge 0\right\} = \left\{\hat{K}_1^{-1}(\hat{P}_1) - \Delta < t,\ S_{nm} \ge 0\right\};$$

and for $t \ge 0$,

$$(4.25)\quad \left\{\hat{\Delta}_{nm} - \Delta > t,\ S_{nm} \ge 0\right\} = \left\{\hat{K}_2^{-1}(\hat{P}_2) - \Delta > t,\ S_{nm} \ge 0\right\}.$$

It is possible that $S_{nm} < 0$ [which corresponds to $\hat{\Delta}_{nm} = \hat{K}_1^{-1}(\hat{P}_1) + \hat{K}_2^{-1}(\hat{P}_2) - (T_2 - T_1)$ or $\hat{\Delta}_{nm} = T_2 - T_1$]. We intend to prove that this event has asymptotically probability 0. It is easy to see [Serfling (1980), page 3] that

$$(4.26)\qquad P\{S_{nm} < 0\} \le P\left\{\left(\hat{K}_1(\Delta) - \hat{P}_1\right)\left(\hat{K}_2(\Delta) - \hat{P}_2\right) \le 0\right\}.$$

But, since $T_1 = T_2 - \Delta$,

$$\hat{K}_1(\Delta) - \hat{P}_1 = \hat{K}_2(\Delta) - \hat{P}_2 + \int_{-\infty}^{T_1} (G_m(t+\Delta) - F_n(t))\, d(G_m(t+\Delta) - F_n(t)).$$

Since the second term of the right-hand side above is of smaller order than the first, $\hat{K}_2(\Delta) - \hat{P}_2$ is asymptotically equal to $\hat{K}_1(\Delta) - \hat{P}_1$, so their product is nonnegative. (A more rigorous argument will follow at the end of this section.)

Now notice that if $t \le 0$, then for any $A > 0$, $\Delta + tAL^{-1/2} \le \Delta = T_2 - T_1$, and for $t \ge 0$, $\Delta + tAL^{-1/2} \ge \Delta = T_2 - T_1$. Thus by the same arguments as in

(i) and (ii), we can prove

$$P\left\{\frac{\sqrt{L}\left(\hat{K}_1^{-1}(\hat{P}_1)-\Delta\right)}{A}<t\right\}\to\Phi(t)\quad\text{for } t\le 0$$

and

$$P\left\{\frac{\sqrt{L}\left(\hat{K}_2^{-1}(\hat{P}_2)-\Delta\right)}{A}\le t\right\}\to\Phi(t)\quad\text{for } t\ge 0,$$

respectively, where

$$A=\frac{\left((1/\lambda)\sigma_1^2(T_0)+(1/(1-\lambda))\sigma_2^2(T_0)\right)^{1/2}}{d(T_0)}.$$

(Notice in the present case, $T_0 = T_1 = T_2 - \Delta$, so this A works for both cases.) Therefore, in view of (4.24), (4.25), and the fact that $P\{S_{nm} < 0\} \to 0$ as $n, m \to \infty$, the proof of the theorem is completed. To establish $P\{S_{nm} < 0\} \to 0$ more rigorously, since $T_1 = T_2 - \Delta$, by Lemma 4.4, we have (with probability 1)

$$\hat{K}_1(\Delta)-\hat{P}_1=\frac{1}{L}\sum_{l=1}^{L}W_l+O\left(\left(\frac{\log L}{L}\right)^{3/4}\right), \tag{4.27}$$

$$\hat{K}_2(\Delta)-\hat{P}_2=\frac{1}{L}\sum_{l=1}^{L}W_l+O\left(\left(\frac{\log L}{L}\right)^{3/4}\right), \tag{4.28}$$

where

$$W_l=\begin{cases}-\dfrac{1}{\lambda_L}\eta_i^{(F)}(T_1,\Delta), & l=i,\ i=1,\dots,n,\\[2ex] \dfrac{1}{1-\lambda_L}\eta_j^{(G)}(T_1,\Delta), & l=j+n,\ j=1,\dots,m.\end{cases}$$

Notice that W_l, $l = 1, 2, \dots, L$, are independent variables with mean 0 and finite variance. Therefore, (4.27) and (4.28) implies

$$\begin{aligned}S_{nm}^{*} &\equiv \frac{L^2\left(\hat{K}_1(\Delta)-\hat{P}_1\right)\left(\hat{K}_2(\Delta)-\hat{P}_2\right)}{\Sigma_l^L V(W_l)}\\ &=\frac{\left(\Sigma_{l=1}^L W_l\right)^2}{\Sigma_{l=1}^L V(W_l)}+O\left(\frac{(\log L)^{3/4}}{L^{1/4}}\right)\\ &\to_{\mathscr{L}} N^2(0,1)=\chi_1^2\quad\text{as } L\to\infty,\end{aligned}$$

where χ_1^2 is a chi-square random variable with one degree of freedom. Thus, from (4.26), we have

$$P\{S_{nm} < 0\} \le P\{S_{nm}^* \le 0\} \to P\{\chi_1^2 \le 0\} = 0 \quad \text{as } m, n \to \infty.$$

This completes our entire proof. □

NOTE ADDED IN PROOF. As one of the referees pointed out, there are two recent papers that deal with the regression analysis with censored data. Lai and Ying (1991) establish the large-sample properties of a modified Buckley–James estimator. Tsiatis (1990) uses linear rank tests to estimate the regression parameters. The equivalency between these two estimators is presented in another recent paper by Ritov (1990). Both methods could be applied to the two-sample problem, since the latter can be formulated as a regression problem. But in order to guarantee the consistency and the asymptotic normality of their estimators they need strong conditions, since they are addressing the more general problem. Their asymptotic variances also contain the derivative of the survival density which is hard to estimate well in the presence of censoring [Wei, Ying and Lin (1990)]. Since we concentrate on the two-sample problem, we were able to avoid all these conditions (as we had intended) and our asymptotic variance does not involve the derivative of the density and can be easily estimated from the data, as presented in the article.

REFERENCES

AKRITAS, M. G. (1986). Empirical processes associated with V-statistics and a class of estimators under random censoring. *Ann. Statist.* **14** 619–637.

BAHADUR, R. R. (1966). A note on quantities in large samples. *Ann. Math. Statist.* **37** 577–580.

BASSIAKOS, Y. C., MENG, X. L. and LO, S. H. (1990). A general estimator of the treatment effect, when the data are heavily censored. Technical Report, Dept. Statistics, Harvard Univ.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. California Press, Berkeley.

HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

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

LAI, T. L. and YING, Z. (1991). Large sample theory of a modified Buckley–James estimator in regression analysis with censored data. *Ann. Statist.* **19** 1370–1402.

LEHMANN, E. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes* **2**. Springer, New York.

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

PADGETT, W. and WEI, L. J. (1982). Estimation of the ratio of scale parameters in the two sample problem with arbitrary right censorship. *Biometrika* **69** 252–256.

RAO, B. S. L. P. (1983). *Nonparametric Functional Estimation*. Academic, London.

RITOV, Y. (1990). Estimation in the linear regression model with censored data. *Ann. Statist.* **18** 303–328.

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

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.

WEI, L. J. and GAIL, M. H. (1983). Nonparametric estimation for a scale-change with censored observations. *J. Amer. Statist. Assoc.* **78** 382–388.

WEI, L. J., YING, Z. and LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77** 845–851.

XIAO LI MENG
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138

YIANNIS BASSIAKOS
NORTHEASTERN UNIVERSITY
BOSTON, MASSACHUSETTS 02115

SHAW-HWA LO
DIVISION OF BIOSTATISTICS
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
600 WEST 168TH STREET
NEW YORK, NEW YORK 10032-3799