The Annals of Statistics
1992, Vol. 20, No. 3, 1361–1390

MINIMUM HELLINGER-TYPE DISTANCE ESTIMATION FOR CENSORED DATA[1]

BY ZHILIANG YING

University of Illinois, Urbana-Champaign

A Hellinger-type distance for hazard rate functions is defined. It is used to obtain a class of minimum distance estimators for data that are subject to a possible right censorship. The corresponding score process is shown to be approximated by a martingale, which is exploited to obtain the asymptotic normality under considerably weaker conditions than those normally assumed for minimum Hellinger distance estimators. It is also shown that under the parametric assumption the estimators are asymptotically as efficient as the maximum likelihood estimators.

1. Introduction. Let $X_1, \dots, X_n$ be a sequence of independent and identically distributed random variables with a common distribution function F. Let f denote its density function. A usual parametric approach is to assume that f belongs to some parametric family of density functions $\{f_\theta : \theta \in \Theta\}$ and to try to estimate, by the maximum likelihood method, the parameter θ_0 for which $f_{\theta_0} = f$. Although under the parametric assumption the maximum likelihood estimator is usually asymptotically efficient, it may behave poorly if this assumption is slightly violated. On the other hand, the total abandonment of the parametric assumption and the use of standard nonparametric procedures avoid making the parametric assumption but, in the mean time, may reduce efficiency.

An alternative approach that retains the efficiency of the maximum likelihood estimator when the parametric assumption is satisfied and remains somewhat robust when the assumption is slightly violated is due to Beran (1977), who introduced the now well-known minimum Hellinger distance estimator (MHDE). Let $\hat{f}_n$ be some kernel density estimator from the sample $X_1, \dots, X_n$. Beran's MHDE $\hat{\theta}_n^b$ is defined as a solution of

$$(1.1)\qquad \int \left[\hat{f}_n^{1/2}(x) - f_{\hat{\theta}_n^b}^{1/2}(x) \right]^2 dx = \min_{\theta \in \Theta} \int \left[\hat{f}_n^{1/2}(x) - f_\theta^{1/2}(x) \right]^2 dx.$$

It is shown in Beran (1977) that if $f = f_{\theta_0}$, then under certain conditions and with an appropriate choice of the kernel density estimator $\hat{f}_n$, $\sqrt{n}(\hat{\theta}_n^b - \theta_0)$ converges in distribution to a normal random variable with mean zero and variance the inverse of the Fisher information. Therefore the MHDE $\hat{\theta}_n^b$ is asymptotically efficient under the parametric assumption. Moreover, it is

Received October 1989; revised October 1991.
[1]Research supported in part by AFOSR and NSF Grant DMS-86-09819.
AMS 1980 *subject classifications*. Primary 62F10, 62E20; secondary 60F05.
Key words and phrases. Hellinger-type distance, hazard function, censored data, kernel estimator, martingale.

intuitively clear from (1.1) and indeed is shown in Beran (1977) that $\hat{\theta}_n^b$ possesses certain robustness properties. Extensions to the multivariate setting are studied in Tamura and Boos (1986), while the corresponding MHDE for counting data can be found in Simpson (1987).

In this article we are interested in developing minimum Hellinger-type distance estimators for data that are subject to a possible right censorship. Censored data often occur in survival analysis. Here by a right censorship we shall mean that, in addition to the i.i.d. random variables $X_1, \dots, X_n$ with the common density f, there exist censoring variables $C_1, \dots, C_n$, which are independent of each other and independent of $X_1, \dots, X_n$, such that only

$$(1.2) \qquad Z_i = \min\{X_i, C_i\} \quad \text{and} \quad \delta_i = I_{\{X_i \le C_i\}}, \qquad i = 1, \dots, n,$$

are observed. The censoring variables may take the value ∞ but not $-\infty$. Moreover, let G_i, which may vary with different i, be the censoring distribution for C_i. This will include the fixed censorship model, by taking G_i to be one-step jump functions, the random censorship model, by taking $G_i \equiv G$ for all i, and the uncensored model, by taking $G_i(x) = 0$, $x < \infty$ and $\Delta G_i(\infty) = 1$ for all i. Since there are already many viable ways to estimate the density function f for the censored data, an obvious extension of Beran's MHDE seems to be a solution of (1.1) with $\hat{f}_n$ being some estimate of f from the censored data $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$. Unfortunately, this naive approach turns out to be inappropriate, since we can show that it does not provide an asymptotically efficient estimator when the parametric assumption is satisfied.

Recently, Yang (1991) introduced an extension of the MHDE to censored data by considering the joint density function of Z_i, δ_i, the observed data, with respect to some measure involving G. There a random censorship model $G_i \equiv G$ is assumed. The Hellinger distance for the joint density function is used to define the estimator. By using similar arguments as those of Beran (1977) in conjunction with the weak convergence of the Kaplan–Meier estimate of the survival distribution [cf. Gill (1983)], asymptotic results are obtained.

The thesis of this paper is to introduce a new Hellinger-type distance estimator that seems to incorporate censorship in a natural way and to investigate asymptotic properties of it. Our approach differs from those of Beran (1977) and Yang (1991) in that we use a Hellinger-type distance for the hazard rate function rather than for the density function. The use of the hazard rate function stems from our understanding that in dealing with survival data, the hazard function often exhibits certain advantages, both conceptually and technically. A particular advantage here is that we can handle non-i.i.d. censoring, since the corresponding estimating equation has a natural martingale integral approximation. Another advantage, also as a result of martingale integral approximation, is that we have the flexibility to trim off the "tail" data without introducing a significant bias. This is especially useful in survival analysis where statistical inference based on observations within a period of time is needed. We can also explore this property to dampen possible erratic tail behavior so that stringent conditions assumed by Beran (1977) as well as Yang (1991) can be removed. Furthermore, the approach can easily be extended to handle the multiplicative counting process models; cf. Borgan (1984).

The paper is organized as follows. Section 2 introduces a class of minimum Hellinger-type distance estimators, along with some notation and definitions. Then, in Section 3, we derive consistency and asymptotic normality of the restricted minimum Hellinger-type distance estimator, which uses only observations falling into a fixed finite interval. By placing an appropriate weight function into the Hellinger-type distance, we shall, in Section 4, show that the resulting minimum Hellinger-type distance estimator is again consistent and asymptotically normal under certain regularity conditions, which are satisfied by usual parametric families, and that the estimator is asymptotically as efficient as the maximum likelihood estimator should the parametric assumption hold. The paper is concluded in Section 5 with discussions on computation, optimality and robustness and extensions to the truncated data.

2. A Hellinger-type distance for hazard rate functions. Throughout the rest of this paper, we denote $X_1, \dots, X_n$ to be a sequence of independent and identically distributed random variables with a common density function f. Denote F to be the cumulative distribution function and let $\lambda = f/(1-F)$ and $\Lambda = -\log(1-F)$ be the corresponding hazard rate and the cumulative hazard functions. As pointed out in the preceding section, it is assumed that the censoring variables $C_1, \dots, C_n$, with subdistribution functions $G_1, \dots, G_n$, are independent of each other and independent of $X_1, \dots, X_n$. The case in which $G_i(x) = 0$ for all $x < \infty$ and $i = 1, \dots, n$, corresponds to the uncensored model. With Z_i and δ_i defined by (1.2) let

$$\begin{aligned} N_n(x) &= \sum_{i=1}^{n} I_{\{Z_i \le x,\, \delta_i = 1\}}, & \bar{N}_n(x) &= N_n(x)/n, \\ Y_n(x) &= \sum_{i=1}^{n} I_{\{Z_i \ge x\}}, & \bar{Y}_n(x) &= Y_n(x)/n. \end{aligned} \tag{2.1}$$

These empirical processes are often used in survival analysis, and are associated with certain martingale integral representations. For each t, define the σ field $\mathscr{F}_t = \sigma\{N_n(x), Y_n(x), x \le t\} \vee \mathscr{N}$, where $\mathscr{N}$ denotes the family of all null sets. It is well known that the process $N_n(t) - \int_{-\infty}^{t} Y_n(x)\, d\Lambda(x)$, $t > -\infty$, is a continuous time martingale with respect to the σ filtration $\{\mathscr{F}_t\}$; compare with Gill (1980). We shall use this fact to derive the asymptotic distribution of the minimum Hellinger-type distance estimator to be introduced.

As in Beran (1977), let $\{f_\theta \colon \theta \in \Theta\}$, where $\Theta \subset R^p$, be a parametric family of density functions. Accordingly, let F_θ, λ_θ and Λ_θ be the corresponding cumulative distribution, hazard rate and cumulative hazard functions. We now define our minimum Hellinger-type distance functionals. Let $a < b$ be fixed numbers, let μ be a hazard rate function and let $1 - H$ be a (sub)distribution function. Then the minimum Hellinger-type distance functional $\xi(\mu, H, a, b)$ is defined implicitly as a solution satisfying

$$(2.2)\quad \begin{aligned}&\int_a^b \left[\lambda^{1/2}_{\xi(\mu, H, a, b)}(x) - \mu^{1/2}(x)\right]^2 H(x)\,dx\\ &\quad = \min_{\theta\in\Theta}\int_a^b \left[\lambda_\theta^{1/2}(x) - \mu^{1/2}(x)\right]^2 H(x)\,dx.\end{aligned}$$

Here confining the integration to the interval $[a, b]$ serves two purposes: to avoid certain technical difficulties related to the tail behavior and to get time dependent functionals such as $\xi(\mu, H, 0, t)$ which are of interest in survival analysis. Now let $\hat{\lambda}_n$ be a kernel estimator, satisfying certain properties that will be specified later, of the underlying hazard rate function λ. We define our minimum Hellinger-type distance estimator, restricted to $[a, b]$, by $\hat{\theta}_n(a, b) = \xi(\hat{\lambda}_n, \overline{Y}_n, a, b)$. In other words, $\hat{\theta}_n(a, b)$ satisfies

$$(2.3)\quad \begin{aligned}&\int_a^b \left[\lambda^{1/2}_{\hat{\theta}_n(a, b)}(x) - \hat{\lambda}_n^{1/2}(x)\right]^2 \overline{Y}_n(x)\,dx\\ &\quad = \min_{\theta\in\Theta}\int_a^b \left[\lambda_\theta^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right]^2 \overline{Y}_n(x)\,dx.\end{aligned}$$

Let Γ_n be defined by

$$(2.4)\quad \Gamma_n(x) = \frac{1}{n}\sum_{i=1}^n [1 - G_i(x)].$$

Then $\hat{\theta}_n(a, b)$ should estimate $\xi(\lambda, \Gamma_n(1 - F), a, b)$ since $\hat{\lambda}_n \sim \lambda$ and $\overline{Y}_n \sim \Gamma_n(1 - F)$. Also the restriction of the integration in (2.2) to $[a, b]$ generally makes the estimator less efficient unless similar conditions as those introduced in Beran (1977) or Yang (1991) are satisfied; that is, $(\partial/\partial\theta)\lambda_\theta$ has a compact support $\subset [a, b]$ and λ is bounded away from zero on $[a, b]$. To define an estimator that will be asymptotically as efficient as the maximum likelihood estimator for the usual parametric families, we put a weight function w_n into the Hellinger-type distance and define $\hat{\theta}_n$ as a solution satisfying

$$(2.5)\quad \begin{aligned}&\int_{-\infty}^{\infty} \left[\lambda^{1/2}_{\hat{\theta}_n}(x) - \hat{\lambda}_n^{1/2}(x)\right]^2 \overline{Y}_n(x) w_n(x)\,dx\\ &\quad = \min_{\theta\in\Theta}\int_{-\infty}^{\infty} \left[\lambda_\theta^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right]^2 \overline{Y}_n(x) w_n(x)\,dx.\end{aligned}$$

The weight function w_n typically approximates an indicator function which dampens the tail effect, making the estimator $\hat{\theta}_n$ stable and yet asymptotically efficient.

Similar to Beran (1977) and Yang (1991), $\hat{\theta}_n$ is also approximating the maximum likelihood estimator of θ if $f \in \{f_\theta: \theta \in \Theta\}$. We give here a heuristic argument. Since the likelihood function is proportional to $\prod_{i=1}^n f_\theta^{\delta_i}(Z_i)(1 - F_\theta(Z_i))^{1-\delta_i}$, it follows that the m.l.e. is to maximize $\int \log\lambda_\theta(x)\,dN_n(x) - \int \lambda_\theta(x) Y_n(x)\,dx$. Because $dN_n(x)/Y_n(x) \approx \hat{\lambda}_n(x)\,dx$, this maximization is asymptotically equivalent to minimizing

$$-\int\left[\log\frac{\lambda_\theta(x)}{\hat\lambda_n(x)}\right]\hat\lambda_n(x)Y_n(x)\,dx+\int\left[\lambda_\theta(x)-\hat\lambda_n(x)\right]Y_n(x)\,dx$$

$$\approx-\int\left[\frac{\lambda_\theta(x)}{\hat\lambda_n(x)}-1-\frac{1}{2}\left(\frac{\lambda_\theta(x)}{\hat\lambda_n(x)}-1\right)^2\right]\hat\lambda_n(x)Y_n(x)\,dx$$

$$+\int\left[\lambda_\theta(x)-\hat\lambda_n(x)\right]Y_n(x)\,dx$$

$$=\frac{1}{2}\int\left[\frac{\lambda_\theta(x)-\hat\lambda_n(x)}{\sqrt{\hat\lambda_n(x)}}\right]^2Y_n(x)\,dx$$

$$\approx 2\int\left[\lambda_\theta^{1/2}(x)-\hat\lambda_n^{1/2}(x)\right]^2Y_n(x)\,dx,$$

where the last approximation follows by using the identity [cf. Beran (1977)]

$$u^{1/2}-v^{1/2}=\frac{u-v}{2u^{1/2}}+\frac{(u-v)^2}{2u^{1/2}(u^{1/2}+v^{1/2})^2}\tag{2.6}$$

with $u=\hat\lambda_n$ and $v=\lambda_\theta$, and by omitting the second order term. Note that we did not include the weight function w_n in our derivation since it is only to stabilize the tail behavior of λ. Indeed, it will be shown in Section 4 that $\sqrt{n}(\hat\theta_n-\theta_0)$ converges weakly to a normal random variable with mean zero and variance the inverse of the Fisher information, thus making $\hat\theta_n$ as efficient as the maximum likelihood estimator.

3. Consistency and asymptotic normality of $\hat\theta_n(a,b)$. This section is devoted to studying the asymptotic behavior of $\hat\theta_n(a,b)$ defined by (2.3). Following Beran (1977), we first establish the existence and the continuity of the minimum Hellinger distance functional $\xi(\mu,H,a,b)$. From this and construction of the consistent hazard rate function estimator, the consistency of $\hat\theta_n(a,b)$ follows. By making some further assumptions on the choice of bandwidth used in the kernel estimator $\hat\lambda_n$, and by developing certain approximations of the estimating equations, we shall derive weak convergence of $\hat\theta_n(a,b)$.

Throughout the sequel it will be assumed that for some function Γ,

$$\sup_{-\infty<x<\infty}|\Gamma_n(x)-\Gamma(x)|\to 0,\tag{3.1}$$

recalling that $\Gamma_n(x)=n^{-1}\Sigma_{i=1}^n(1-G_i(x))$. Therefore, by a Glivenko–Cantelli-type argument,

$$\sup_{-\infty<x<\infty}\left|\bar Y_n(x)-(1-F(x))\Gamma(x)\right|\to 0\quad\text{a.s.}\tag{3.2}$$

Note that for random censorship models, $\Gamma_n(x)=\Gamma(x)=1-G(x)$. Let τ_0, τ_1

and τ denote the "endpoints" of F and $(1 - F)\Gamma$; that is,

$$\tau_0 = \inf\{x: F(x) > 0\}, \qquad \tau_1 = \sup\{x: 1 - F(x) > 0\},$$
$$\tau = \sup\{x: (1 - F(x))\Gamma(x) > 0\}. \tag{3.3}$$

To ensure that the minimum values in (2.2), (2.3) and (2.5) are attainable, we assume that Θ is a compact subset of R^p. As is pointed out in Beran (1977), for a noncompact Θ, it is often possible to embed it into a compact set and to show that the minimization can only occur within Θ. In these situations, all the results here remain valid. In practice, when the minimizer does not exist for a particular sample realization or is difficult to compute, one may wish to use a suboptimal solution instead. Asymptotic properties of the suboptimal solutions may be established along the line of Beran and Millar [(1987), Lemma 5.1].

The next two lemmas deal with existence and uniqueness of the minimum distance functional defined by (2.2), and will be used throughout the rest of the paper.

LEMMA 1. *Suppose that $\lambda_\theta(u)$ is a bivariate continuous function on $\Theta \times [a, b]$. Then $\xi(\mu, H, a, b)$ defined by* (2.2) *exists provided $\int_a^b \mu(x)\,dx < \infty$. If, in addition, $\xi(\mu, H, a, b)$ is unique, then for any μ_n and H_n, $\int_a^b(\mu_n^{1/2}(x) - \mu^{1/2}(x))^2\,dx \to 0$ and $\sup_{x\in[a,b]}|H_n(x) - H(x)| \to 0$ imply $\xi(\mu_n, H_n, a, b) \to \xi(\mu, H, a, b)$.*

PROOF. From the assumption that $\lambda_\theta(u)$ is continuous on $\Theta \times [a, b]$ it is clear that the trivariate functional

$$\Phi(\theta, \mu^{1/2}, H) = \int_a^b \left[\lambda_\theta^{1/2}(x) - \mu^{1/2}(x)\right]^2 H(x)\,dx \tag{3.4}$$

is continuous with the Euclidean norm for θ, the $L^2[a, b]$ norm for $\mu^{1/2}$ and the sup-norm for H. In particular, for fixed μ and H, $\Phi(\theta, \mu^{1/2}, H)$ is a continuous function of θ on the compact set Θ and therefore the minimization (2.2) is attainable; that is, $\xi(\mu, H, a, b)$ is well defined. To show $\xi(\mu_n, H_n, a, b) \to \xi(\mu, H, a, b)$, we note that if otherwise, then there exists a subsequence n_k such that $\xi(\mu_{n_k}, H_{n_k}, a, b) \to \xi_0 \neq \xi(\mu, H, a, b)$ as $k \to \infty$. But by the continuity of Φ and the definition of the functional ξ,

$$\begin{aligned}\Phi(\xi_0, \mu^{1/2}, H) &= \lim_{k\to\infty} \Phi\left(\xi(\mu_{n_k}, H_{n_k}, a, b), \mu_{n_k}^{1/2}, H_{n_k}\right)\\ &\leq \lim_{k\to\infty} \Phi\left(\xi(\mu, H, a, b), \mu_{n_k}^{1/2}, H_{n_k}\right)\\ &= \Phi(\xi(\mu, H, a, b), \mu^{1/2}, H).\end{aligned}$$

However, since $\xi_0 \neq \xi(\mu, H, a, b)$, the above inequality contradicts (2.2) and the uniqueness assumption. □

LEMMA 2. *Let* $\tau_H = \sup\{x\colon H(x) > 0\}$. *Suppose that* $\lambda = \lambda_{\theta_0}$ *for some* $\theta_0 \in \Theta$ *and that for each pair* $\theta_1 \neq \theta_2$ *the Lebesgue measure of* $\{x \in [a, b \wedge \tau_H]\colon \lambda_{\theta_1}(x) \neq \lambda_{\theta_2}(x)\}$ *is nonzero, where* $b \wedge \tau_H$ *denotes* $\min\{b, \tau_H\}$. *Then* $\xi(\lambda, H, a, b) = \theta_0$.

PROOF. Since $\lambda = \lambda_{\theta_0}$, the definition of ξ implies that

$$0 = \int_a^b \left[\lambda^{1/2}_{\xi(\lambda, H, a, b)}(x) - \lambda^{1/2}_{\theta_0}(x)\right]^2 H(x)\,dx$$

$$\geq \frac{1}{n}\int_a^{b\wedge\tau_H} I_{\{[\lambda^{1/2}_{\xi(\lambda, H, a, b)}(x) - \lambda^{1/2}_{\theta_0}(x)]^2 \geq n^{-1}\}} H(x)\,dx.$$

But $H(x) > 0$ on $[a, b \wedge \tau_H)$. Therefore, for every n, the set $\{x \in [a, b \wedge \tau_H]\colon [\lambda^{1/2}_{\xi(\lambda, H, a, b)}(x) - \lambda^{1/2}_{\theta_0}(x)]^2 \geq n^{-1}\}$ is a null set. Thus $\{x \in [a, b \wedge \tau_H]\colon \lambda_{\xi(\lambda, H, a, b)}(x) \neq \lambda_{\theta_0}(x)\}$ is also a null set, implying that $\xi(\lambda, H, a, b) = \theta_0$ by the assumption. □

The rest of this section is to study the limiting behavior of $\hat{\theta}_n(a, b)$ defined by (2.3). The hazard rate function estimator $\hat{\lambda}_n$, used in (2.3), will be a kernel estimator of the form

$$\text{(3.5)} \quad \hat{\lambda}_n(x) = \frac{1}{d_n}\int_{-\infty}^{\infty} K\left(\frac{x-u}{d_n}\right) d\hat{\Lambda}_n(u) \quad \text{with } \hat{\Lambda}_n(x) = \int_{-\infty}^{x} \frac{dN_n(u)}{Y_n(u)},$$

where K is a kernel function, that is, $K \geq 0$ and $\int K = 1$, and d_n is positive and converges to 0 at a certain rate. We refer to Ramlau-Hansen (1983) for a detailed discussion of such estimators.

LEMMA 3. *Let* $\hat{\lambda}_n$ *be the kernel estimator of* λ *defined by* (3.5) *with K being of bounded variation and having a compact support* $\subset [-c, c]$ *for some* $c > 0$.

(i) *Suppose that there exists an open interval* $(x_0, x_1) \supset [a, b]$ *with* $x_1 < \tau$ *such that* λ *is continuous in* (x_0, x_1). *Suppose also that the bandwidths* d_n *are so chosen that* $d_n \to_P 0$ *and* $nd_n^2 \to_P \infty$. *Then*

$$\text{(3.6)} \qquad \sup_{x\in[a, b]} \left|\hat{\lambda}_n(x) - \lambda(x)\right| \to_P 0.$$

(ii) *Let the bandwidths* $d_n \to 0$ *be nonrandom. Define* $\tilde{\lambda}_n$ *by*

$$\text{(3.7)} \qquad \tilde{\lambda}_n(x) = \frac{1}{d_n}\int_{-\infty}^{\infty} K\left(\frac{x-u}{d_n}\right)\lambda(u)\,du.$$

Suppose that for a sequence of constants $k_n \to \infty$, $P\{Y_n(b + d_n c) \geq k_n\} \to 1$. *Then there exist nonnegative random variables* Q_n *and* R_n *such that*

$$\text{(3.8)} \qquad \left[\hat{\lambda}_n(x) - \tilde{\lambda}_n(x)\right]^2 \leq Q_n(x) + R_n(x),$$

where $P\{\sup_{x\in[a,b]}Q_n(x)=0\}\to 1$, *as* $n\to\infty$, *and*

$$\sup_{x\in[a,b]} ER_n(x) \le 2\int_{-\infty}^{\infty} K^2(s)\,ds(d_n k_n)^{-1} \sup_{x\in[a-d_nc,\,b+d_nc]} \lambda(x).$$

In addition, if K *is symmetric and* λ *is twice continuously differentiable on* $[a-d_nc, b+d_nc]$, *then*

$$\sup_{x\in[a,b]} \left|\tilde\lambda_n(x)-\lambda(x)\right| \le d_n^2 \sup_{x\in[a-d_nc,\,b+d_nc]} |\lambda''(x)| \int_{-\infty}^{\infty} s^2K(s)\,ds. \tag{3.9}$$

PROOF. Part (i) is essentially a special case of Theorem 4.1.2 of Ramlau-Hansen (1983) and we refer to it for the proof of (3.6).

The proof of (ii) is also similar to Ramlau-Hansen (1983), except we need to consider those s for which $Y_n(s)/n$ is small. Let

$$\hat\lambda_n^*(x) = \frac{1}{d_n}\int_{-\infty}^{\infty} K\left(\frac{x-s}{d_n}\right) I_{\{Y_n(s)\ge k_n\}}\,d\hat\Lambda_n(s),$$

$$\tilde\lambda_n^*(x) = \frac{1}{d_n}\int_{-\infty}^{\infty} K\left(\frac{x-s}{d_n}\right) I_{\{Y_n(s)\ge k_n\}}\lambda(s)\,ds.$$

Since

$$\left\{\hat\lambda_n(x)\ne\hat\lambda_n^*(x) \text{ for some } x\in[a,b]\right\}\cup\left\{\tilde\lambda_n(x)\ne\tilde\lambda_n^*(x) \text{ for some } x\in[a,b]\right\}$$
$$\subset\left\{Y_n(b+d_nc)<k_n\right\},$$

it follows from the assumption on Y_n that $Q_n(x)=2[\hat\lambda_n(x)-\tilde\lambda_n(x)-\hat\lambda_n^*(x)+\tilde\lambda_n^*(x)]^2$ satisfies

$$P\left\{\sup_{x\in[a,b]} Q_n(x)=0\right\}\to 1.$$

Furthermore, since $\int_{-\infty}^{x} I_{\{Y_n(s)\ge k_n\}}d[\hat\Lambda_n(s)-\Lambda(s)]$ is a martingale,

$$\begin{aligned}
E\left[\hat\lambda_n^*(x)-\tilde\lambda_n^*(x)\right]^2 &= \frac{1}{d_n^2}E\int_{-\infty}^{\infty} K^2\left(\frac{x-s}{d_n}\right) I_{\{Y_n(s)\ge k_n\}}\frac{\lambda(s)\,ds}{Y_n(s)}\\
&\le \frac{1}{d_n^2k_n}\int_{-\infty}^{\infty} K^2\left(\frac{x-s}{d_n}\right)\lambda(s)\,ds \qquad (3.10)\\
&\le \frac{1}{d_nk_n}\sup_{u\in[x-d_nc,\,x+d_nc]}\lambda(u)\int_{-\infty}^{\infty} K^2(s)\,ds.
\end{aligned}$$

Taking $\sup_{x\in[a,b]}$ on both sides of (3.10) we have

$$\sup_{x\in[a,b]} ER_n(x) \le 2\int_{-\infty}^{\infty} K^2(s)\,ds(d_nk_n)^{-1} \sup_{x\in[a-d_nc,\,b+d_nc]} \lambda(x),$$

where $R_n(x) = 2[\hat{\lambda}_n^*(x) - \tilde{\lambda}_n^*(x)]^2$. (3.8) then follows by noting that

$$\left[\hat{\lambda}_n(x) - \tilde{\lambda}_n(x)\right]^2 \leq 2\left[\hat{\lambda}_n(x) - \tilde{\lambda}_n(x) - \hat{\lambda}_n^*(x) + \tilde{\lambda}_n^*(x)\right]^2 + 2\left[\hat{\lambda}_n^*(x) - \tilde{\lambda}_n^*(x)\right]^2.$$

Finally, (3.9) may be seen by the inequality

$$\left|\tilde{\lambda}_n(x) - \lambda(x)\right| = \left|\int_{-c}^{c} K(s)[\lambda(x - d_n s) - \lambda(x)]\, ds\right| \leq \frac{d_n^2}{2} \sup_{u \in [x - d_n c, x + d_n c]} |\lambda''(u)| \int_{-\infty}^{\infty} K(s) s^2\, ds,$$

noting that $\lambda(x - d_n s) - \lambda(x) = \lambda'(x)\, d_n s + 2^{-1}\lambda''(x^*)(d_n s)^2$ and that $\int sK(s)\, ds = 0$. □

Theorem 1. *Let $[a, b]$ be a finite interval on the real line with $b < \tau$. Let $\hat{\lambda}_n$ be defined by (3.5) with K being of bounded variation and having a compact support and with $d_n \to_P 0$ and $nd_n^2 \to_P \infty$. Suppose that $\lambda_\theta(x)$ is continuous in $\Theta \times [a, b]$ and $\theta_\lambda(a, b) = \xi(\lambda, (1 - F)\Gamma, a, b)$ is unique. Suppose also that λ is continuous in an open interval containing $[a, b]$. Then $\hat{\theta}_n(a, b) \to_P \theta_\lambda(a, b)$.*

Remark. In particular, if $\lambda = \lambda_{\theta_0}$ for some θ_0, then $\hat{\theta}_n(a, b) \to_P \theta_0$.

Proof. From Lemma 3(i), we have $\int_a^b [\hat{\lambda}_n^{1/2}(x) - \lambda^{1/2}(x)]^2\, dx \to_P 0$, which implies, by Lemma 3.3.2 of Chow and Teicher (1978), that for any subsequence $\{n_k\}$ of $\{n\}$ there is still a subsequence $\{n_{kl}\} \subset \{n_k\}$ such that $\int_a^b [\hat{\lambda}_{n_{kl}}^{1/2}(x) - \lambda^{1/2}(x)]^2\, dx \to 0$, a.s. Thus from Lemma 1 and (3.2) we have $\hat{\theta}_{n_{kl}}[a, b] \to \theta_\lambda[a, b]$ a.s. Since $\{n_k\}$ is arbitrary, $\hat{\theta}_n(a, b) \to_P \theta_\lambda(a, b)$, again by Lemma 3.3.2 of Chow and Teicher (1978). □

Theorem 2. *Let $\hat{\lambda}_n$ be defined by (3.5) with K being symmetric and satisfying the same conditions of Theorem 1 and d_n being nonrandom and satisfying $nd_n^2 \to \infty$ and $nd_n^4 \to 0$. Suppose that λ is continuous and strictly positive in an open interval containing $[a, b]$ and $\theta_\lambda(a, b) = \xi(\lambda, (1 - F)\Gamma, a, b)$ is unique and lies in the interior of Θ. Also suppose that $\lambda_\theta(x), \dot{\lambda}_\theta(x) = (\partial/\partial\theta)\lambda_\theta(x)$ and $\ddot{\lambda}_\theta(x) = (\partial^2/\partial\theta^2)\lambda_\theta(x)$ are continuous on $\Theta \times [a, b]$ with $\lambda_{\theta_\lambda(a,b)}(x) > 0$ for all $x \in [a, b]$ and that $A(\theta_\lambda(a, b), a, b)$ is nonsingular, where*

$$A(\theta, a, b) = \int_a^b \dot{\lambda}_\theta(x)\dot{\lambda}_\theta^T(x)\left[\frac{\lambda(x)}{\lambda_\theta(x)}\right]^{1/2} \frac{\Gamma(x)\, dF(x)}{\lambda_\theta(x)\lambda(x)} + 2\int_a^b \ddot{\lambda}_\theta(x)\left\{1 - \left[\frac{\lambda(x)}{\lambda_\theta(x)}\right]^{1/2}\right\} \frac{\Gamma(x)}{\lambda(x)}\, dF(x).$$

Then for $\hat{\theta}_n(a,b) = \xi(\hat{\lambda}_n, \bar{Y}_n, a, b)$ *and* $\theta_n(a,b) = \xi(\lambda, \bar{Y}_n, a, b)$,

$$\sqrt{n}\left(\hat{\theta}_n(a,b) - \theta_n(a,b)\right) \to_{\mathscr{D}} N\left(0, \textstyle\sum(\theta_\lambda(a,b), a, b)\right), \tag{3.11}$$

where

$$\sum(\theta, a, b) = A^{-1}(\theta, a, b)\int_a^b \dot{\lambda}_\theta(x)\dot{\lambda}_\theta^T(x)\frac{\Gamma(x)\,dF(x)}{\lambda_\theta(x)\lambda(x)}A^{-1}(\theta, a, b).$$

COROLLARY 1. *If, in addition to the assumptions of Theorem* 2, *we also assume that conditions of Lemma* 2 *are satisfied, then*

$$\sqrt{n}\left(\hat{\theta}_n(a,b) - \theta_0\right) \to_{\mathscr{D}} N\left(0, I_{\theta_0}^{-1}(a,b)\right), \tag{3.12}$$

where

$$I_\theta(a,b) = \int_a^b \dot{\lambda}_\theta(x)\dot{\lambda}_\theta^T(x)\frac{\Gamma(x)}{\lambda_\theta^2(x)}\,dF(x). \tag{3.13}$$

Moreover, if there is $b_0 \in (a,b]$ *such that the identifiability conditions of Lemma* 2 *are satisfied with* b *replaced by* b_0 *and that* $I_{\theta_0}(a, b_0)$ *is positive definite, then*

$$\sqrt{n}\left(\hat{\theta}_n(a,\cdot) - \theta_0\right) \to_{\mathscr{D}[b_0,b]} Z(\cdot), \tag{3.14}$$

where Z *is a multivariate Gaussian process with mean zero and covariance matrix function* $E[Z(t)Z^T(s)] = I_{\theta_0}^{-1}(a,t)I_{\theta_0}(a, t\wedge s)I_{\theta_0}^{-1}(a,s)$. *Here and in the sequel,* $\mathscr{D}[u,v]$ *denotes the space, equipped with the Skorokhod topology, of functions on* $[u,v]$ *that are right continuous with left limits.*

REMARK 1. In general, $\theta_n(a,b)$ are random vectors because $\bar{Y}_n$, which may be regarded as weights of the Hellinger distance, are random. However, as is stated in the preceding corollary, under the parametric assumption, the $\theta_n(a,b)$ essentially take the same value θ_0, which neither depends on n nor is random. We may also derandomize θ_n by using $\eta_n(a,b) = \xi(\lambda, E\bar{Y}_n, a, b)$ and consider instead the convergence of $\sqrt{n}(\hat{\theta}_n(a,b) - \eta_n(a,b))$ as will be discussed in Corollary 2.

REMARK 2. Some regularity conditions introduced here in terms of the hazard rate function are related to their counterparts in terms of the density function. The requirement that λ be bounded away from zero in $[a,b]$ is equivalent to the condition that its corresponding density function f is bounded away in the same interval. Also an identifiability condition related to that of Lemma 2 is that the set $\{x \in (-\infty, b]: \lambda_{\theta_1}(x) \neq \lambda_{\theta_2}(x)\}$ has a positive Lebesgue measure. It is readily seen that this statement is the same as saying that the set $\{x \in (-\infty, b]: f_{\theta_1}(x) \neq f_{\theta_2}(x)\}$ has a positive Lebesgue measure. However, replacing $-\infty$ by a may cause the equivalence to be invalid.

REMARK 3. If we impose a further requirement that $\dot{\lambda}_\theta$ has a compact support contained in $[a, b]$, which is also equivalent to the condition that $\dot{f}_\theta$ has a compact support in $[a, b]$, as is assumed in Beran (1977), then $I_\theta(a, b) = I_\theta = \int_{-\infty}^{\infty} [\dot{\lambda}_\theta(t)/\lambda_\theta(t)]^2 \Gamma(t)\, dF(t)$, which can be shown to be the average (since we do not assume an i.i.d. censorship) Fisher information. The Fisher information representation in terms of hazard rate function has also been discussed in Borgan (1984) and Efron and Johnstone (1990). In this case, $\hat{\theta}_n(a, b)$ provides an asymptotically efficient estimator if the parametric assumption holds.

REMARK 4. Let $C_i(t) = C_i \wedge t$ and suppose that X_i are censored by $C_i(t)$. Then the average Fisher information becomes $I_\theta(t) = \int_{-\infty}^{t} [\dot{\lambda}_\theta(u)/\lambda_\theta(u)]^2 \Gamma(u)\, dF(u)$. In survival analysis, $I_\theta(t)$ may be interpreted as the information accumulated up to time t. This shows that $I_\theta(a, b)$ is the Fisher information collected during the time interval (a, b). Thus the estimator $\hat{\theta}_n(a, b)$ utilizes efficiently the localized information observed in (a, b).

PROOF OF THEOREM 2 AND COROLLARY 1. From Theorem 1 and Lemma 1, both $\hat{\theta}_n(a, b)$ and $\theta_n(a, b)$ converge to $\theta_\lambda(a, b)$ in probability. But $\hat{\theta}_n(a, b)$ and $\theta_n(a, b)$ are minimizers of functions $\Phi(\cdot, \hat{\lambda}_n^{1/2}, \overline{Y}_n)$ and $\Phi(\cdot, \lambda^{1/2}, \overline{Y}_n)$, respectively, defined by (3.4). Thus it follows from the existence and the continuity of $\dot{\lambda}_\theta(x)$, the positivity of $\lambda_{\theta_\lambda(a,b)}(x)$ and the assumption that $\theta_\lambda(a, b)$ belongs to the interior of Θ that, as $n \to \infty$ and with probability approaching to 1,

$$\int_a^b \left[\lambda_{\hat{\theta}_n(a,b)}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right] \lambda_{\hat{\theta}_n(a,b)}^{-1/2}(x) \dot{\lambda}_{\hat{\theta}_n(a,b)}(x) \overline{Y}_n(x)\, dx = 0, \tag{3.15}$$

$$\int_a^b \left[\lambda_{\theta_n(a,b)}^{1/2}(x) - \lambda^{1/2}(x)\right] \lambda_{\theta_n(a,b)}^{-1/2}(x) \dot{\lambda}_{\theta_n(a,b)}(x) \overline{Y}_n(x)\, dx = 0. \tag{3.16}$$

The assumption on the existence and the continuity of $\ddot{\lambda}_\theta(x)$ also enables us to take the Taylor expansion of (3.15) at $\theta_n(a, b)$ to get

$$\begin{aligned} &\int_a^b \frac{\lambda_{\theta_n(a,b)}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n(a,b)}^{1/2}(x)} \dot{\lambda}_{\theta_n(a,b)}(x) \overline{Y}_n(x)\, dx \\ &\quad + \frac{1}{2}\left[\int_a^b \frac{\hat{\lambda}_n^{1/2}}{\lambda_{\theta_n^*(a,b)}^{3/2}(x)} \dot{\lambda}_{\theta_n^*(a,b)}(x) \dot{\lambda}_{\theta_n^*(a,b)}^T(x) \overline{Y}_n(x)\, dx \right. \\ &\quad \left. + 2\int_a^b \frac{\lambda_{\theta_n^*(a,b)}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*(a,b)}^{1/2}(x)} \ddot{\lambda}_{\theta_n^*(a,b)}(x) \overline{Y}_n(x)\, dx\right] \\ &\quad \times \left[\hat{\theta}_n(a, b) - \theta_n(a, b)\right] = 0. \end{aligned} \tag{3.17}$$

Here in order to simplify our notation we have used the same symbol $\theta_n^*(a, b)$ to represent p different vectors, each of which lies on the line segment connecting $\hat{\theta}_n(a, b)$ and $\theta_n(a, b)$, in the above Taylor expansions of the set of p equations. Since $\hat{\theta}_n(a, b)$ and $\theta_n(a, b)$ converge to $\theta_\lambda(a, b)$ and

$\sup_{x\in[a,b]}|\hat{\lambda}_n(x)-\lambda(x)|\to_P 0$, the continuity assumptions on $\lambda_\theta(x)$, $\dot{\lambda}_\theta(x)$ and $\ddot{\lambda}_\theta(x)$ together with (3.2) imply that

$$\begin{aligned}(3.18)\qquad &\int_a^b \frac{\hat{\lambda}_n^{1/2}(x)}{\lambda^{3/2}_{\theta_n^*(a,b)}(x)}\dot{\lambda}_{\theta_n^*(a,b)}(x)\dot{\lambda}^T_{\theta_n^*(a,b)}(x)\bar{Y}_n(x)\,dx\\ &\quad+2\int_a^b \frac{\lambda^{1/2}_{\theta_n^*(a,b)}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda^{1/2}_{\theta_n^*(a,b)}(x)}\ddot{\lambda}_{\theta_n^*(a,b)}(x)\bar{Y}_n(x)\,dx\\ &\to_P A(\theta_\lambda(a,b),a,b).\end{aligned}$$

On the other hand, from (3.16) and (2.6),

$$\begin{aligned}&\int_a^b \frac{\lambda^{1/2}_{\theta_n(a,b)}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda^{1/2}_{\theta_n(a,b)}(x)}\dot{\lambda}_{\theta_n(a,b)}(x)\bar{Y}_n(x)\,dx\\ &\quad=\int_a^b \frac{\lambda^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda^{1/2}_{\theta_n(a,b)}(x)}\dot{\lambda}_{\theta_n(a,b)}(x)\bar{Y}_n(x)\,dx\\ (3.19)\qquad &\quad=\frac{1}{2}\int_a^b\left[\lambda(x)-\hat{\lambda}_n(x)\right]\frac{\dot{\lambda}_{\theta_n(a,b)}(x)\bar{Y}_n(x)\,dx}{\lambda^{1/2}(x)\lambda^{1/2}_{\theta_n(a,b)}(x)}\\ &\qquad+\frac{1}{2}\int_a^b\left[\lambda(x)-\hat{\lambda}_n(x)\right]^2\frac{\dot{\lambda}_{\theta_n(a,b)}(x)\bar{Y}_n(x)\,dx}{\lambda^{1/2}(x)\left[\lambda^{1/2}(x)+\hat{\lambda}_n^{1/2}(x)\right]^2\lambda^{1/2}_{\theta_n(a,b)}(x)}.\end{aligned}$$

Now Lemma 3(ii) and the assumptions on the choice of d_n, that is, $nd_n^2\to\infty$ and $nd_n^4\to 0$, imply that $\sup_{x\in[a,b]}|\lambda(x)-\hat{\lambda}_n(x)|=o_p(n^{-1/4})$. Therefore from the positivity and differentiability assumptions on λ and $\lambda_\theta(x)$, it follows that

$$\begin{aligned}(3.20)\qquad &\int_a^b\left[\lambda(x)-\hat{\lambda}_n(x)\right]^2\frac{\dot{\lambda}_{\theta_n(a,b)}(x)\bar{Y}_n(x)\,dx}{\lambda^{1/2}(x)\left[\lambda^{1/2}(x)+\hat{\lambda}_n^{1/2}(x)\right]^2\lambda^{1/2}_{\theta_n(a,b)}(x)}\\ &\quad=o_p(n^{-1/2}).\end{aligned}$$

In view of (3.17)–(3.20) we see that (3.11) will hold if it can be shown that

$$\begin{aligned}(3.21)\qquad &\sqrt{n}\int_a^b\left[\lambda(x)-\hat{\lambda}_n(x)\right]\frac{\dot{\lambda}_{\theta_n(a,b)}(x)\bar{Y}_n(x)\,dx}{\lambda^{1/2}(x)\lambda^{1/2}_{\theta_n(a,b)}(x)}\\ &\quad\to_{\mathscr{D}} N\left(0,\int_a^b\dot{\lambda}_{\theta_\lambda(a,b)}(x)\dot{\lambda}^T_{\theta_\lambda(a,b)}(x)\frac{\Gamma(x)\,dF(x)}{\lambda_{\theta_\lambda(a,b)}(x)\lambda(x)}\right).\end{aligned}$$

We shall show that the left-hand side of (3.21) may be approximated by a martingale integral and then apply a martingale central limit theorem to prove (3.21). The first step to do so is to show that $\theta_n(a,b)$ on the left-hand side of (3.21) can be replaced by the nonrandom quantity

$$\eta_n(a,b) = \xi\big(\lambda, E\overline{Y}_n, a, b\big). \tag{3.22}$$

Since $\eta_n(a,b) \to \theta_\lambda(a,b)$ and $\lambda_{\eta_n(a,b)}(x) > 0$ for all $x \in [a,b]$ and large n, we can, similar to (3.17), take the Taylor expansion of $\int_a^b[\lambda^{1/2}_{\theta_n(a,b)}(x) - \lambda^{1/2}(x)]\lambda^{-1/2}_{\theta_n(a,b)}(x)\dot{\lambda}_{\theta_n(a,b)}(x)\overline{Y}_n\,dx$ at $\eta_n(a,b)$ and get

$$\begin{aligned}
&\|\theta_n(a,b) - \eta_n(a,b)\|\\
&\quad= O_p\left(\left\|\int_a^b\left[\lambda^{1/2}_{\eta_n(a,b)}(x) - \lambda^{1/2}(x)\right]\lambda^{-1/2}_{\eta_n(a,b)}(x)\dot{\lambda}_{\eta_n(a,b)}(x)\overline{Y}_n(x)\,dx\right\|\right)\\
&\quad= O_p\left(\left\|\int_a^b\left[\lambda^{1/2}_{\eta_n(a,b)}(x) - \lambda^{1/2}(x)\right]\lambda^{-1/2}_{\eta_n(a,b)}(x)\dot{\lambda}_{\eta_n(a,b)}(x)\right.\right.\\
&\qquad\qquad\left.\left.\times\left(\overline{Y}_n(x) - E\overline{Y}_n(x)\right)dx\right\|\right)\\
&\quad= O_p\left(\int_a^b\left|\overline{Y}_n(x) - E\overline{Y}_n(x)\right|dx\right) = o_p(n^{-1/2+\epsilon})
\end{aligned}$$

for any $\varepsilon > 0$, noting that

$$\int_a^b\left[\lambda^{1/2}_{\eta_n(a,b)}(x) - \lambda^{1/2}(x)\right]\lambda^{-1/2}_{\eta_n(a,b)}(x)\dot{\lambda}_{\eta_n(a,b)}(x)E\overline{Y}_n(x)\,dx = 0$$

and that $E\int_a^b[\overline{Y}_n(x) - E\overline{Y}_n(x)]^2\,dx \le n^{-1}(b-a)$. Letting $0 < \varepsilon < 1/4$, it follows that $\|\theta_n(a,b) - \eta_n(a,b)\| = o_p(n^{-1/4})$, which together with the smoothness and the positivity assumptions on λ_θ and λ implies

$$\begin{aligned}
&\int_a^b\left[\lambda(x) - \hat{\lambda}_n(x)\right]\frac{\dot{\lambda}_{\theta_n(a,b)}(x)\overline{Y}_n(x)\,dx}{\lambda^{1/2}(x)\lambda^{1/2}_{\theta_n(a,b)}(x)}\\
&\quad-\int_a^b\left[\lambda(x) - \hat{\lambda}_n(x)\right]\frac{\dot{\lambda}_{\eta_n(a,b)}(x)\overline{Y}_n(x)\,dx}{\lambda^{1/2}(x)\lambda^{1/2}_{\eta_n(a,b)}(x)} = o_p(n^{-1/2}),
\end{aligned}\tag{3.23}$$

noting that $\sup_{x\in[a,b]}|\hat{\lambda}_n(x) - \lambda(x)| = o_p(n^{-1/4})$ by Lemma 3(ii). Next by (3.9) we have

$$\sup_{x\in[a,b]}\left|\tilde{\lambda}_n(x) - \lambda(x)\right| = o_p(n^{-1/2}), \tag{3.24}$$

which implies that

$$\int_a^b \hat{\lambda}_n(x) \frac{\dot{\lambda}_{\eta_n(a,b)}(x)\bar{Y}_n(x)\,dx}{\lambda^{1/2}(x)\lambda^{1/2}_{\eta_n(a,b)}(x)} - \int_a^b \frac{\dot{\lambda}_{\eta_n(a,b)}(x)\bar{Y}_n(x)}{\lambda^{1/2}(x)\lambda^{1/2}_{\eta_n(a,b)}(x)}\,d\hat{\Lambda}_n(x)$$

$$(3.25)\quad = \int_a^b \frac{\dot{\lambda}_{\eta_n(a,b)}(x)\bar{Y}_n(x)}{\lambda^{1/2}(x)\lambda^{1/2}_{\eta_n(a,b)}(x)} \times \left\{\frac{1}{d_n}\int_{-\infty}^{\infty} K\left(\frac{x-s}{d_n}\right) d\big[\hat{\Lambda}_n(s) - \Lambda(s)\big]\,dx - d\big[\hat{\Lambda}_n(x) - \Lambda(x)\big]\right\} + o_p(n^{-1/2}).$$

Clearly, the total variation $\int_a^b |d[\dot{\lambda}_{\eta_n(a,b)}(x)\bar{Y}_n(x)/(\lambda^{1/2}(x)\lambda^{1/2}_{\eta_n(a,b)}(x))]| = O_p(1)$. Thus (3.25) is $o_p(n^{-1/2})$ if we can show that

$$(3.26)\quad \sup_{x\in[a,b]}\left|\int_{-\infty}^{x}\frac{1}{d_n}\int_{-\infty}^{\infty}K\left(\frac{u-s}{d_n}\right)d\big[\hat{\Lambda}_n(s)-\Lambda(s)\big]\,du - \big[\hat{\Lambda}_n(x)-\Lambda(x)\big]\right| = o_p(n^{-1/2}).$$

To do so we first note that from the integration by parts formula we have

$$\frac{1}{d_n}\int_{-\infty}^{\infty}K\left(\frac{u-s}{d_n}\right)d\big[\hat{\Lambda}_n(s)-\Lambda(s)\big] = \int_{-\infty}^{\infty}\big[\hat{\Lambda}_n(s)-\Lambda(s)\big]\frac{1}{d_n^2}K'\left(\frac{u-s}{d_n}\right)ds.$$

Therefore,

$$\int_{-\infty}^{x}\frac{1}{d_n}\int_{-\infty}^{\infty}K\left(\frac{u-s}{d_n}\right)d\big[\hat{\Lambda}_n(s)-\Lambda(s)\big]\,du - \big[\hat{\Lambda}_n(x)-\Lambda(x)\big]$$

$$(3.27)\quad = \int_{-\infty}^{\infty}\big[\hat{\Lambda}_n(s)-\Lambda(s)\big]\frac{1}{d_n}K\left(\frac{x-s}{d_n}\right)ds - \big[\hat{\Lambda}_n(x)-\Lambda(x)\big]$$

$$= \int_{-\infty}^{\infty}\big[\hat{\Lambda}_n(x-d_nv)-\Lambda(x-d_nv)-(\hat{\Lambda}_n(x)-\Lambda(x))\big]K(v)\,dv.$$

Since Λ is continuous and the jumps of $\hat{\Lambda}_n$ are of $O(1/n)$, it follows that $\sqrt{n}\,\sup_{x\in[a,b]}|\hat{\Lambda}_n(x-d_nv)-\Lambda(x-d_nv)-(\hat{\Lambda}_n(x)-\Lambda(x))|$ is of the order $o_p(1)$. Therefore, (3.26), which implies (3.25), holds. Now Rebolledo's martingale central limit theorem [cf. Gill (1980)] implies that

$$(3.28)\quad \sqrt{n}\int_a^t \frac{\dot{\lambda}_{\eta_n(a,b)}(x)\bar{Y}_n(x)}{\lambda^{1/2}(x)\lambda^{1/2}_{\eta_n(a,b)}(x)}d\big[\hat{\Lambda}_n(x)-\Lambda(x)\big] \to_{\mathscr{D}[a,b]} \int_a^t \frac{\dot{\lambda}_{\theta_\lambda(a,b)}(x)(1-F(x))\Gamma(x)}{\lambda^{1/2}_{\theta_\lambda(a,b)}(x)\lambda^{1/2}(x)}\,dW(x),$$

where W is a zero-mean Gaussian martingale with covariance function $\int_{-\infty}^{t} d\Lambda(u)/[(1-F(u))\Gamma(u)]$, $t \in [a,b]$. Since with t replaced by b, the right-hand side of (3.28) is a normal random vector with mean zero and covariance matrix

$$\int_a^b \dot{\lambda}_{\theta_\lambda(a,b)}(x)\dot{\lambda}^T_{\theta_\lambda(a,b)}(x)\left[\lambda_{\theta_\lambda(a,b)}(x)\lambda(x)\right]^{-1}\Gamma(x)\,dF(x),$$

(3.23), (3.25) and (3.28) imply (3.21), which further implies (3.11).

To prove Corollary 1, we note that under the assumptions of Lemma 2, $\theta_n(a,b)=\theta_0$ for all large n and

$$A(\theta_0,a,b)=\int_a^b \dot{\lambda}_{\theta_0}(x)\dot{\lambda}^T_{\theta_0}(x)\lambda_{\theta_0}^{-2}(x)\Gamma(x)\,dF(x)=I_{\theta_0}(a,b),$$

implying that $\Sigma(\theta_0,a,b)=I_{\theta_0}^{-1}$. Hence, (3.11) becomes (3.12). Now since $I_{\theta_0}(a,b_0)$ is positive definite, $\hat{\lambda}_n \to \lambda_{\theta_0}$ and $\int_a^t[\lambda_\theta^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)]^2\bar{Y}_n(x)\,dx$ is continuous in θ and t, it follows that for all large n, $\hat{\theta}_n(a,t)$ is continuous in $[b_0,b]$. Moreover, it is easily seen that all the previous approximations are valid with b replaced by t and uniformly in $t\in[b_0,b]$. Therefore by using (3.28) we can easily get (3.14). □

COROLLARY 2. *Let $\eta_n(a,b)=\xi(\lambda, E\bar{Y}_n, a, b)$, the same as* (3.22), *and let all the assumptions of Theorem* 2 *be satisfied. Assume*

$$\Gamma(x,y)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(1-G_i(x))(1-G_i(y)) \tag{3.29}$$

exists for all x and y. Therefore, $\Gamma(x)=\Gamma(x,-\infty)$. Denote

$$\begin{aligned} B(\theta,a,b) &= \int_a^b \dot{\lambda}_\theta(x)\dot{\lambda}^T_\theta(x)\frac{\Gamma(x)\,dF(x)}{\lambda_\theta(x)\lambda(x)} \\ &\quad + 2\int_a^b\int_a^x\left[U_\theta(x)V_\theta^T(y)+V_\theta(x)U_\theta^T(x)\right] \\ &\quad \times(1-F(x))\Gamma(x)\frac{dF(y)}{1-F(y)}\,dx \\ &\quad + 4\int_a^b\int_a^b U_\theta(x)U_\theta^T(y)[(1-F(x\vee y))\Gamma(x\vee y) \\ &\qquad\qquad -(1-F(x))(1-F(y))\Gamma(x,y)]\,dx\,dy, \end{aligned} \tag{3.30}$$

where

$$U_\theta(x)=\left(1-\frac{\lambda^{1/2}(x)}{\lambda_\theta^{1/2}(x)}\right)\dot{\lambda}_\theta(x),\qquad V_\theta(x)=\frac{\dot{\lambda}_\theta(x)}{\lambda^{1/2}(x)\lambda_\theta^{1/2}(x)}.$$

Then

$$\sqrt{n}\left(\hat{\theta}_n(a,b) - \eta_n(a,b)\right)$$
$$\to_{\mathscr{D}} N\left(0, A^{-1}(\lambda_\lambda(a,b),a,b)B(\theta_\lambda(a,b),a,b)A^{-1}(\theta_\lambda(a,b),a,b)\right).$$

REMARK. $\eta_n(a,b)$ depends upon n because we do not wish to assume the censoring distributions to be the same. However if this is the case, that is, $G_i = G$, then $\eta_n(a,b) = \eta(a,b) = \xi(\lambda,(1-F)(1-G),a,b)$.

The proof of Corollary 2 is very much the same as that of Theorem 1, except we need to take into account an extra variation due to the replacement of $\overline{Y}_n$ by $E\overline{Y}_n$. Therefore the details are omitted.

4. Construction and asymptotic properties of $\hat{\theta}_n$. This section is devoted to the study of the asymptotic behavior of $\hat{\theta}_n$ defined by (2.5) with an appropriate weight function w_n. Extending the integral in the definition of the Hellinger-type distance from $[a,b]$ to the whole line is necessary in order to utilize all the essential information in the data so as to make the estimator asymptotically efficient when the parametric assumption holds. A main technical difficulty arising from this extension of $[a,b]$ to the whole line is caused by tail fluctuations of relevant hazard rate functions. The same difficulty also occurs in the usual minimum Hellinger distance estimation when density functions are involved. However, unlike the latter case, here we can arbitrarily dampen the tail influence by putting a weight function in the Hellinger-type distance for hazard rate functions without introducing a significant bias into the estimator. Although there are many possible ways to choose the weight function w_n, we shall discuss here only a specific type defined by (4.2). The consistency and the asymptotic normality of the estimator $\hat{\theta}_n$ with this type of weight functions will then be shown.

Let τ_0, τ_1 and τ be defined by (3.3). Denote θ_λ to be a solution of

$$\int_{\tau_0}^{\tau}\left(\lambda_{\theta_\lambda}^{1/2}(x) - \lambda^{1/2}(x)\right)^2(1-F(x))\Gamma(x)\,dx$$
$$(4.1)\qquad = \min_{\theta\in\Theta}\int_{\tau_0}^{\tau}\left(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x)\right)^2(1-F(x))\Gamma(x)\,dx.$$

In Theorem 3 we shall give conditions that guarantee the existence of θ_λ. Let α_i and c_i, $i = 1,2$, be some positive constants and let $c > 0$ be so chosen that the kernel function K has a compact support contained in $[-c,c]$. Define the weight function as

$$(4.2)\qquad w_n(x) = \begin{cases} 1, & \text{if } \overline{N}_n(x - d_n c) \ge c_1(\log n)^{-\alpha_1} \\ & \quad \text{and } \overline{Y}_n(x + d_n c) \ge c_2(\log n)^{-\alpha_2}, \\ 0, & \text{otherwise.} \end{cases}$$

The weight function w_n is basically an adaptive version of the interval $[a, b]$ introduced earlier. It confines us to a time interval that is an open subinterval of and approaching to $[\tau_0, \tau]$. Thus, if $[a, b] \subset (\tau_0, \tau)$, then for all large n, $w_n(x) = 1$, $\forall\ x \in [a, b]$. Its main advantage is that it enables us to avoid certain technical difficulties due to a possible discontinuity of the hazard rate function at τ_0 and to the well-known tail instability of $\hat{\Lambda}_n$ at τ. The use of $\log n$ in the definition of the weight function here is that it does not affect certain "algebraic bounds" that will be used in the technical development. The choice of α_i and c_i, $i = 1, 2$, should be based on a proper balance between efficiency and stability.

The following lemma shows that on the set $w_n(x) = 1$, $\overline{Y}_n(x)$ and $E\overline{Y}_n(x) = (1 - F(x))\Gamma_n(x)$ are essentially the same, so are $\overline{N}_n(x)$ and $E\overline{N}_n(x)$.

LEMMA 4. *On an event with probability* 1, *for any* $\varepsilon > 0$, *there exists an* n_ε, *depending on the sample point, such that for all* $n \geq n_\varepsilon$, $w_n(x) = 1$ *implies*

$$(4.3)\quad (1 - \varepsilon)(1 - F(x))\Gamma_n(x) \leq \overline{Y}_n(x) \leq (1 + \varepsilon)(1 - F(x))\Gamma_n(x),$$

$$(4.4)\quad \begin{aligned} (1 - \varepsilon)\Gamma_n(x)F(x) &\leq (1 - \varepsilon)\int_{-\infty}^{x} \Gamma_n(s)\, dF(s) \leq \overline{N}_n(x) \\ &\leq (1 + \varepsilon)\int_{-\infty}^{x} \Gamma_n(s)\, dF(s) \leq (1 + \varepsilon)F(x). \end{aligned}$$

PROOF. From an exponential inequality of Bretagnolle [cf. Shorack and Wellner (1986), page 797, Inequality 2], it follows that for every $\varepsilon > 0$,

$$P\left\{\sup_x \left|\overline{Y}_n(x) - (1 - F(x))\Gamma_n(x)\right| \geq n^{-1/2+\varepsilon}\right\} \leq D\exp\{-2n^{-2\varepsilon}\}$$

for some constant $D > 0$ and all n. Therefore, the Borel–Cantelli lemma can be used to show that for any $\gamma > 0$,

$$(4.5)\quad \sup_x \left|\overline{Y}_n(x) - E\overline{Y}_n(x)\right| = o(n^{-1/2+\gamma}) \quad \text{a.s.}$$

Since $w_n(x) = 1$ implies $\overline{Y}_n(x) \geq c_2(\log n)^{-\alpha_2}$, (4.3) follows from (4.5).

Now let $\overline{M}_n(x) = \overline{N}_n(x) - \int_{-\infty}^{x}\overline{Y}_n(s)(1 - F(s))^{-1}\, dF(s)$. Then

$$M_n^*(x) = \int_{-\infty}^{x} I_{\{1-F(s)\geq n^{-\gamma/2}\}}\, d\overline{M}_n(s)$$

is a martingale satisfying $\sup_x|\Delta M_n^*(x)| \leq 1/n$ and

$$\langle M_n^*\rangle(x) = \frac{1}{n}\int_{-\infty}^{x} I_{\{1-F(s)\geq n^{-\gamma/2}\}}\frac{\overline{Y}_n(s)\, dF(s)}{1 - F(s)} \leq n^{-(1-\gamma/2)}.$$

Here and in the sequel, $\Delta g(t)$ of a right continuous function g denotes $g(t) - g(t-)$. Thus from a martingale exponential inequality [cf. Shorack and

Wellner (1986), page 899, Inequality 1] we have, for $\gamma < 1/2$,

$$P\Big(\sup_x |M_n^*(x)| \geq n^{-1/2+\gamma}\Big) \leq 2\exp\left\{-\frac{n^{-1+2\gamma}}{n^{-1+\gamma/2}}\psi\left(\frac{n^{-1/2+\gamma}n^{-1}}{n^{-1+\gamma/2}}\right)\right\}$$
$$= o(e^{-n^{\gamma}}) \quad \text{a.s.},$$

where ψ is a continuous function on $[0, \infty)$ with $\psi(0) = 1$. Therefore, the Borel–Cantelli lemma together with the fact, for all large n, $M_n^*(x) = \overline{M}_n(x)$ if $w_n(x) = 1$, implies that

$$\sup_{x:\, w_n(x)=1} \left| \overline{N}_n(x) - \int_{-\infty}^{x} \overline{Y}_n(s)(1 - F(s))^{-1}\, dF(s) \right| = O(n^{-1/2+\gamma}) \quad \text{a.s.}$$

From the above bound, (4.5) and (4.3) we have

$$\sup_{x:\, w_n(x)=1} \left| \overline{N}_n(x) - \int_{-\infty}^{x} \Gamma_n(s)\, dF(s) \right| = o(n^{-1/2+2\gamma}) \quad \text{a.s.},$$

which implies (4.4) because $\overline{N}_n(x) \geq c_1(\log n)^{-\alpha_1}$ on $w_n(x) = 1$. □

THEOREM 3. *Suppose that $\lambda_\theta(x)$ is continuous in $\Theta \times (\tau_0, \tau)$ and that λ is continuously differentiable in (τ_0, τ). Assume that*

$$(1 - F(\tau))\Gamma_n(\tau +) = o(n^{-\varepsilon}) \tag{4.6a}$$

for some $\varepsilon > 0$ and that

$$\lim_{\substack{b \uparrow \tau \\ a \downarrow \tau_0}} \sup_{\theta \in \Theta} \left\{ \int_{\tau_0}^{a} \left(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x)\right)^2 (1 - F(x))\, dx + \int_{b}^{\tau} \left(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x)\right)^2 (1 - F(x))\, dx \right\} = 0. \tag{4.6b}$$

Then θ_λ exists. In addition, suppose that θ_λ is unique,

$$|x| + |\lambda'(x)| = O\big((1 - F(x))^{-k}\big), \quad \textit{as } x \uparrow \tau \tag{4.7}$$

for some $k > 0$, and that $\hat{\lambda}_n$ is defined by (3.5) *with its kernel function K being continuously differentiable and having a compact support contained in $[-c, c]$, and with nonrandom bandwidth d_n satisfying $nd_n^2 \to \infty$ and $d_n = o(n^{-\alpha})$ for some $\alpha > 0$. Then $\hat{\theta}_n \to_P \theta_\lambda$.*

REMARK. Condition (4.6a) says that, for large n, the risk set size process Y_n beyond the "last point" τ is at most of the order $n^{1-\varepsilon}$ for some $\varepsilon > 0$. It is imposed to guarantee that the Hellinger-type distance used in (2.5) approaches that used in (4.1). This condition is certainly satisfied if $\tau = \tau_1$ or if the censoring variables are i.i.d.

PROOF. Choose $a_n \downarrow \tau_0$ and $b_n \uparrow \tau$. From Lemma 1, for each n, the function $\Psi_n(\theta) = \int_{a_n}^{b_n}(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x))^2(1 - F(x))\Gamma(x)\, dx$ is continuous on Θ. But

(4.6b) implies that $\Psi_n(\theta)$ converges uniformly to $\Psi(\theta) = \int_{\tau_0}^{\tau}(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x))^2(1 - F(x))\Gamma(x)\,dx$. Therefore, Ψ is also continuous on Θ, whose compactness implies that θ_λ exists. To show $\hat{\theta}_n \to \theta_\lambda$, we need the convergence result

$$\int_{\tau_0}^{\tau}\left(\lambda_\theta^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right)^2 \overline{Y}_n(x) w_n(x)\,dx \tag{4.8}$$
$$\to_P \int_{\tau_0}^{\tau}\left(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x)\right)^2 (1 - F(x))\Gamma(x)\,dx$$

for every $\theta \in \Theta$, which will be shown at the end of the proof. Now for any fixed $[a, b] \subset (\tau_0, \tau)$, the definition of $\hat{\theta}_n$ implies that, for all large n,

$$\int_{\tau_0}^{\tau}\left(\lambda_\theta^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right)^2 \overline{Y}_n(x) w_n(x)\,dx$$
$$\geq \int_a^b\left(\lambda_{\hat{\theta}_n}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right)^2 \overline{Y}_n(x) w_n(x)\,dx \tag{4.9}$$
$$= \int_a^b\left(\lambda_{\hat{\theta}_n}^{1/2}(x) - \lambda^{1/2}(x)\right)^2 (1 - F(x))\Gamma(x)\,dx + o_p(1),$$

where the last equality follows from Lemma 3 and (4.5), noting that for all large n, $w_n(x) = 0$ for $x \geq \tau$ by (4.6a) and $w_n(x) = 1$ for $x \in [a, b]$. Suppose that $\hat{\theta}_n$ does not converge to θ_λ. Then there exist $\varepsilon_0 > 0$, $\delta_0 > 0$ and a subsequence n_k such that $P\{|\hat{\theta}_{n_k} - \theta_\lambda| \geq \varepsilon_0\} \geq \delta_0$, for all large k. However, the uniqueness of θ_λ and (4.6) imply that there exists $\varepsilon_1 > 0$ and we can choose $a > \tau_0$ and $b < \tau$ such that

$$\inf_{|\theta - \theta_\lambda| \geq \varepsilon_0} \int_a^b\left(\lambda_\theta^{1/2}(x) - \lambda^{1/2}(x)\right)^2 (1 - F(x))\Gamma(x)\,dx \tag{4.10}$$
$$\geq \int_{\tau_0}^{\tau}\left(\lambda_{\theta_\lambda}^{1/2}(x) - \lambda^{1/2}(x)\right)^2 (1 - F(x))\Gamma(x)\,dx + \varepsilon_1.$$

Combining together (4.8)–(4.10) and $P\{|\hat{\theta}_{n_k} - \theta_\lambda| \geq \varepsilon_0\} \geq \delta_0$ clearly leads to a contradiction, showing that $\hat{\theta}_n \to_P \theta_\lambda$. Thus it remains to show (4.8). We observe that (4.8) holds trivially if τ_0 and τ are replaced by a and b. In view of Lemma 4 and (4.6) and the Chebyshev' inequality, it suffices to show

$$\lim_{a \downarrow \tau_0} \limsup_{n \to \infty} \int_{\tau_0}^{a} E\hat{\lambda}_n(x)(1 - F(x))\Gamma_n(x) \times I_{\{F(x - d_n c) \geq c_1/2(\log n)^{-\alpha_1}\}}\,dx = 0, \tag{4.11a}$$

$$\lim_{b \uparrow \tau} \limsup_{n \to \infty} \int_{b}^{\tau} E\hat{\lambda}_n(x)(1 - F(x))\Gamma_n(x) \times I_{\{1 - F(x + d_n c) \geq c_2/2(\log n)^{-\alpha_2}\}}\,dx = 0, \tag{4.11b}$$

noting that $\lambda(1 - F) = f \in L^1(\tau_0, \tau)$. The equation (4.11a) is easily seen from

$$\int_{\tau_0}^{a} E\hat{\lambda}_n(x)(1 - F(x))\Gamma_n(x) I_{\{F(x - d_n c) \geq c_1/2(\log n)^{-\alpha_1}\}}\, dx$$

$$\leq \int_{\tau_0 + d_n c}^{a} \int_{-\infty}^{\infty} d_n^{-1} K\left(\frac{x - s}{d_n}\right)\lambda(s)\, ds\, dx \leq \int_{\tau_0}^{a + d_n c} \lambda(s)\, ds \to 0$$

as $n \to \infty$ and $a \downarrow \tau_0$. Note that the argument holds regardless of whether $\tau_0 = -\infty$ or $\tau_0 > -\infty$. For (4.11b), we have

$$\int_b^{\tau} E\hat{\lambda}_n(x)(1 - F(x))\Gamma_n(x) I_{\{1 - F(x + d_n c) \geq c_2/2(\log n)^{-\alpha_2}\}}\, dx$$

$$\leq \int_b^{\tau - d_n c} \int_{-\infty}^{\infty} d_n^{-1} K\left(\frac{x - s}{d_n}\right)\lambda(s)\, ds (1 - F(x))\Gamma_n(x) \times I_{\{1 - F(x + d_n c) \geq c_2/2(\log n)^{-\alpha_2}\}}\, dx$$

$$\leq \int_b^{\tau - d_n c} d_n c \sup_{x^*: |x^* - x| \leq d_n c} \{|\lambda'(x^*)|\}(1 - F(x))\Gamma_n(x) \times I_{\{1 - F(x + d_n c) \geq c_2/2(\log n)^{-\alpha_2}\}}\, dx$$

$$+ \int_b^{\tau - d_n c} \lambda(x)(1 - F(x))\, dx.$$

The second term on the right-hand side of the preceding equation approaches 0 trivially. From (4.7), since $d_n = o(n^{-\alpha})$, we have, on $1 - F(x + d_n c) \geq c_2/2(\log n)^{-\alpha_2}$, for some constant $B > 0$,

$$d_n \sup_{x^*: |x^* - x| \leq d_n c} \{|\lambda'(x^*)|\} \leq B d_n (1 - F(x + d_n c))^{-k}$$

$$\leq B^2 (1 - F(x + d_n c))^{k+1}.$$

Since (4.7) implies $\int_b^{\infty}(1 - F(x))^{k+1}\, dx < \infty$, the first term also converges to 0. Hence (4.11b) holds. □

Throughout the rest of this section we shall always assume that θ_λ exists and is unique. Before introducing the next theorem we shall define θ_n and η_n, similar to $\theta_n(a, b)$ and $\eta_n(a, b)$ of the previous section, by

$$\int \left(\lambda_{\theta_n}^{1/2}(x) - \lambda^{1/2}(x)\right)^2 \overline{Y}_n(x) w_n(x)\, dx = \min_{\theta \in \Theta} \int \left(\lambda_{\theta}^{1/2}(x) - \lambda^{1/2}(x)\right)^2 \overline{Y}_n(x) w_n(x)\, dx, \tag{4.12}$$

$$\int \left(\lambda_{\eta_n}^{1/2}(x) - \lambda^{1/2}(x)\right)^2 E\overline{Y}_n(x) \tilde{w}_n(x)\, dx = \min_{\theta \in \Theta} \int \left(\lambda_{\theta}^{1/2}(x) - \lambda^{1/2}(x)\right)^2 E\overline{Y}_n(x) \tilde{w}_n(x)\, dx, \tag{4.13}$$

where $\tilde{w}_n$ is the same as w_n defined by (4.3) except $\overline{N}_n$ and $\overline{Y}_n$ are replaced by $E\overline{N}_n$ and $E\overline{Y}_n$. Moreover, we list below some regularity conditions that will be used in establishing weak convergence of $\hat{\theta}_n$.

(4.14) For some $k_1 > 0$, λ is twice continuously differentiable in (τ_0, τ) and

$$\lambda^{-1}(x) + |\lambda'(x)| + |\lambda''(x)| = O\big((1 - F(x))^{-k_1} + F^{-k_1}(x)\big).$$

(4.15) There exists a constant $\delta_1 > 0$ such that

$$\lim_{\substack{b \uparrow \tau \\ a \downarrow \tau_0}} \sup_{\|\theta - \theta_\lambda\| \le \delta_1} \left\{ \int_{\tau_0}^{a} \|\dot{\lambda}_\theta(x)\|^2 \frac{\lambda^{1/2}(x)}{\lambda_\theta^{3/2}(x)} (1 - F(x))\, dx + \int_b^{\tau} \|\dot{\lambda}_\theta(x)\|^2 \frac{\lambda^{1/2}(x)}{\lambda_\theta^{3/2}(x)} (1 - F(x))\, dx \right\} = 0,$$

$$\lim_{\substack{b \uparrow \tau \\ a \downarrow \tau_0}} \sup_{\|\theta - \theta_\lambda\| \le \delta_1} \left\{ \int_{\tau_0}^{a} \|\ddot{\lambda}_\theta(x)\|^2 \left(\frac{\lambda^{1/2}(x)}{\lambda_\theta^{1/2}(x)} + 1 \right) (1 - F(x))\, dx + \int_b^{\tau} \|\ddot{\lambda}_\theta(x)\|^2 \left(\frac{\lambda^{1/2}(x)}{\lambda_\theta^{1/2}(x)} + 1 \right) (1 - F(x))\, dx \right\} = 0.$$

(4.16) $\lambda_\theta(x)$, $\dot{\lambda}_\theta(x)$ and $\ddot{\lambda}_\theta(x)$ are continuous in $\Theta \times (\tau_0, \tau)$ and for some $k_2 > 0$ and $\delta_2 > 0$,

$$\sup_{\|\theta - \theta_\lambda\| \le \delta_2} \left\{ \lambda_\theta(x) + \frac{\|\dot{\lambda}_\theta(x)\|^4}{\lambda_\theta^3(x)} + \frac{\|\ddot{\lambda}_\theta(x)\|^2}{\lambda_\theta(x)} \right\} = O\big((1 - F(x))^{-k_2} + (F(x))^{-k_2}\big).$$

THEOREM 4. *Suppose that the assumptions of Theorem 3 together with (4.14)–(4.16) are satisfied and that θ_λ lies in the interior of Θ. In addition, suppose that the kernel function K is symmetric and that the bandwidth sequence $\{d_n\}$ satisfies $d_n^2 = O(n^{-1/2-\beta_1})$ and $d_n^{-1/2} = O(n^{1/4-\beta_2})$ for some $\beta_i > 0$, $i = 1, 2$. Denote $A(\theta) = A(\theta, \tau_0, \tau)$ and $B(\theta) = B(\theta, \tau_0, \tau)$, where $A(\theta, a, b)$ and $B(\theta, a, b)$ are defined as in Theorem 2 and*

$$D(\theta) = \int_{\tau_0}^{\tau} \dot{\lambda}_\theta(x) \dot{\lambda}_\theta^T(x) \frac{\Gamma(x)\, dF(x)}{\lambda_\theta(x)\lambda(x)}. \tag{4.17}$$

Then

$$\sqrt{n}\left(\hat{\theta}_n - \theta_n\right) \to_{\mathscr{D}} N\left(0, A^{-1}(\theta_\lambda) D(\theta_\lambda) A^{-1}(\theta_\lambda)\right), \tag{4.18}$$

$$\sqrt{n}\left(\hat{\theta}_n - \eta_n\right) \to_{\mathscr{D}} N\left(0, A^{-1}(\theta_\lambda) B(\theta_\lambda) A^{-1}(\theta_\lambda)\right). \tag{4.19}$$

In particular, if the parametric and the identifiability assumptions of Lemma 2 are satisfied for some $[a, b] \subset (\tau_0, \tau)$, *then* $\theta_n = \eta_n = \theta_0$ *for all large n and*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \to_{\mathscr{D}} N\left(0, I^{-1}(\theta_0)\right), \tag{4.20}$$

where $I(\theta) = \int_{\tau_0}^{\tau} \dot{\lambda}_\theta(x) \dot{\lambda}_\theta^T(x) (\Gamma(x)/\lambda_\theta^2(x))\, dF(x)$.

REMARK 1. The regularity conditions (4.14)–(4.16) are satisfied by the usual parametric families when Θ are taken to be compact subsets of their corresponding natural parameter spaces. For example, it can be easily verified that they hold for the normal, the exponential and the Cauchy families with compact Θ.

REMARK 2. As we mentioned in the previous section, $I(\theta)$ is the Fisher information matrix. Therefore, (4.20) shows that under the parametric assumption $\hat{\theta}_n$ is asymptotically as efficient as the maximum likelihood estimator.

REMARK 3. The two conditions on d_n imply that $nd_n^4 = O(n^{-2\beta_1}) \to 0$ and $(nd_n^2)^{-1} = O(n^{-4\beta_2}) \to 0$. It can be shown then that β_ι are restricted to $\beta_1 < 1/2$ and $\beta_2 < 1/8$. Moreover, it can also be shown that if we choose $d_n = n^{-\alpha}$ with $1/4 < \alpha < 1/2$, then the conditions are satisfied.

We preface the proof with the following two lemmas.

LEMMA 5. *Suppose the condition* (4.14) *is satisfied. Let* $\hat{\lambda}_n$ *and* $\tilde{\lambda}_n$ *be defined by* (3.5) *and* (3.7) *with the kernel function K and with the bandwidth* d_n *satisfying the same assumptions as specified in Theorem* 4. *Then*

$$\sup_x \left\{ w_n(x) \left| \tilde{\lambda}_n(x) - \lambda(x) \right| \right\} = o_p(n^{-1/2-\varepsilon_1}) \quad \text{for every } \varepsilon_1 < \beta_1, \tag{4.21}$$

$$\sup_x \left\{ w_n(x) \left| \hat{\lambda}_n(x) - \lambda(x) \right| \right\} = o_p(n^{-1/4-\varepsilon_2}) \quad \text{for every } \varepsilon_2 < \beta_2. \tag{4.22}$$

PROOF. From Lemma 4 and the definition of w_n, for all large n, $w_n(x) > 0$ implies $F(x - d_n c) \geq c_1 (\log n)^{-\alpha_1}/2$ and $1 - F(x + d_n c) \geq c_2 (\log n)^{-\alpha_2}/2$. Consequently, by (3.9), (4.14) and the assumption $d_n^2 = O(n^{-1/2-\beta_1})$, for some $k > 0$,

$$\begin{aligned}\sup_x \left\{ w_n(x) \left| \tilde{\lambda}_n(x) - \lambda(x) \right| \right\} &= O_p\left(d_n^2 (\log n)^k\right) \\ &= o_p(n^{-1/2-\varepsilon_1}) \quad \text{for every } \varepsilon_1 < \beta_1;\end{aligned}$$

that is, (4.21) holds. Now applying the integration by parts formula and the fact that $\int dK = 0$, it is easily seen that

$$\begin{aligned}
&\left|\hat{\lambda}_n(x) - \tilde{\lambda}_n(x)\right| \\
&= \frac{1}{d_n}\left|\int_{x-d_n c}^{x+d_n c}\left[\hat{\Lambda}_n(s) - \Lambda(s) - \hat{\Lambda}_n(x) + \Lambda(x)\right] dX\left(\frac{x-s}{d_n}\right)\right| \\
&\le \frac{1}{\sqrt{n}\, d_n \bar{Y}_n(x + d_n c)} \\
&\quad \times \sup_{s\in[x-d_n c,\, x+d_n c]}\left|\sqrt{n}\,\bar{Y}_n(s)\left[\hat{\Lambda}_n(s) - \Lambda(s) - \hat{\Lambda}_n(x) + \Lambda(x)\right]\right| \\
&\quad \times \int_{-\infty}^{\infty}|K'(u)|\, du,
\end{aligned}$$

which together with Lemma 6 and the definition of w_n implies

$$\sup_x\left\{w_n(x)\left|\hat{\lambda}_n(x) - \tilde{\lambda}_n(x)\right|\right\} = o_p\left(\frac{(\log n)^k n^{\varepsilon}}{\sqrt{n}\, d_n^{1/2}}\right) \tag{4.23}$$

for every $\varepsilon > 0$ and some $k > 0$. Since $d_n^{-1/2} = O(n^{1/4-\beta_2})$, (4.23) and (4.21) imply (4.22). □

LEMMA 6. *Let $\hat{\Lambda}_n$ be Nelson's estimator as in* (3.5). *Suppose that* (4.14) *is satisfied. Let h_n satisfy $n^{-\gamma_2} \le h_n \le n^{-\gamma_1}$ for some $0 < \gamma_1 < \gamma_2 < 1$. Then for every $\varepsilon > 0$,*

$$\sup_{\substack{|x-y|\le h_n,\\ w_n(x)=1}}\left|\hat{\Lambda}_n(x) - \Lambda(x) - \hat{\Lambda}_n(y) - \Lambda(y)\right| = o_p\left(n^{-1/2+\varepsilon}h_n^{1/2}\right).$$

PROOF. First choose $x_1 < \cdots < x_{n^4}$ in such a way that $F(x_i) = i/n^4$. Then

$$P\left\{X_k \in [x_{i-1}, x_i] \text{ and } X_l \in [x_{i-1}, x_i] \text{ for some } 1 < i \le n^4 \text{ and } k \ne l\right\} \le \frac{1}{n^2},$$

implying that for all large n, there can only be at most one survival time X_k in each interval $[x_{i-1}, x_i]$. Therefore, for all large n, N_n can only have at most one jump within each $[x_{i-1}, x_i]$. In view of this and the fact that

$$\sup_{i,\, 1-F(x_i)\ge(\log n)^{-2\alpha_2}}\left[\Lambda(x_i) - \Lambda(x_{i-1})\right] \le \frac{(\log n)^{2\alpha_2}}{n^4} = o\left(n^{-1/2}h_n^{1/2}\right),$$

it suffices to prove that

$$\sup_{\substack{0<x_i-x_j\le h_n,\\ 1-F(x_i)\ge(\log n)^{-2\alpha_2},\\ w_n(x_i)=1}} \left|\hat{\Lambda}_n(x_i) - \Lambda(x_i) - \hat{\Lambda}_n(x_j) + \Lambda(x_j)\right| = o_p\left(n^{-1/2+\varepsilon}h_n^{1/2}\right) \tag{4.24}$$

for every $\varepsilon > 0$. Now for each fixed pair (j, i) with $0 < x_i - x_j \le h_n$,

$$V_{ij}(t) = \int_{x_j}^{t} I_{\{\bar{Y}_n(s)\ge c_2(\log n)^{-\alpha_2},\, 1-F(s)\ge(\log n)^{-2\alpha_2}\}} d\left[\hat{\Lambda}_n(s) - \Lambda(s)\right], \qquad t \in [x_j, x_i],$$

is a martingale with

$$\Delta V_{ij}(t) \le (\log n)^{\alpha_2}(c_2 n)^{-1} \quad \text{and} \quad \langle V_{ij}\rangle_t \le (\log n)^{3\alpha_2}(c_2 n)^{-1} h_n.$$

Thus the martingale exponential inequality [cf. Shorack and Wellner (1986), page 899, Inequality 1] can be used to get for every $1/2 > \varepsilon > 0$,

$$P\left\{\sup_{x_j\le t\le x_i} \left|V_{ij}(t)\right| \ge n^{-1/2+\varepsilon}h_n^{1/2}\right\} \le 2\exp\left\{-\frac{\left[n^{-1/2+\varepsilon}h_n^{1/2}\right]^2}{2(\log n)^{3\alpha_2}(c_2 n)^{-1}h_n}\zeta_n\right\} \le e^{-n^{\varepsilon}} \tag{4.25}$$

for all large n, where $\zeta_n \to 1$. Since there are less than n^8 pairs of (i, j), (4.25) and the Borel–Cantelli lemma can be used to get $\sup_{i,j}|V_{ij}(x_i)| = O(n^{-1/2+\varepsilon}h_n^{1/2})$, which in view of Lemma 4 implies (4.24). □

Proof of Theorem 4. For the sake of simplicity, we shall only prove the case $p = 1$, that is, $\Theta \subset R$, since the proof for $p > 1$ differs only in notation.

From Theorem 3, $\hat{\theta}_n \to_P \theta_\lambda$. Furthermore, since $\bar{Y}_n(x)w_n(x) \to_P (1 - F(x))\Gamma(x)$ and $E\bar{Y}_n(x)\tilde{w}_n(x) \to (1 - F(x))\Gamma(x)$, for all $x \in (\tau_0, \tau)$, we can apply the same argument as in the proof of Theorem 3 to show that $\theta_n \to_P \theta_\lambda$ and $\eta_n \to \theta_\lambda$. To prove (4.18), we note that analogous to (3.15) and (3.16),

$$\int_{\tau_0}^{\tau}\left[\lambda_{\hat{\theta}_n}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)\right]\lambda_{\hat{\theta}_n}^{-1/2}(x)\dot{\lambda}_{\hat{\theta}_n}(x)\bar{Y}_n(x)w_n(x)\,dx = 0, \tag{4.26}$$

$$-\int_{\tau_0}^{\tau}\left[\lambda_{\theta_n}^{1/2}(x) - \lambda^{1/2}(x)\right]\lambda_{\theta_n}^{-1/2}(x)\dot{\lambda}_{\theta_n}(x)\bar{Y}_n(x)w_n(x)\,dx = 0. \tag{4.27}$$

Summing together the left-hand sides of (4.26) and (4.27) and then taking the

Taylor expansion of it (as a function of $\hat{\theta}_n$) at θ_n, we get

$$
\begin{aligned}
0 = &\int_{\tau_0}^{\tau} \frac{\lambda^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n}^{1/2}(x)} \dot{\lambda}_{\theta_n}(x)\overline{Y}_n(x)w_n(x)\,dx \\
&+ \frac{1}{2}\left\{\int_{\tau_0}^{\tau} \frac{\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*}^{3/2}(x)} \dot{\lambda}_{\theta_n^*}^2(x)\overline{Y}_n(x)w_n(x)\,dx \right. \\
&\left. + 2\int_{\tau_0}^{\tau} \frac{\lambda_{\theta_n^*}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*}^{1/2}(x)} \ddot{\lambda}_{\theta_n^*}(x)\overline{Y}_n(x)w_n(x)\,dx\right\}\left(\hat{\theta}_n - \theta_n\right).
\end{aligned} \tag{4.28}
$$

For each $[a, b] \subset (\tau_0, \tau)$, we can apply an argument analogous to that leading to (3.18) to get

$$
\begin{aligned}
&\int_{a}^{b} \frac{\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*}^{3/2}(x)} \dot{\lambda}_{\theta_n^*}^2(x)\overline{Y}_n(x)w_n(x)\,dx \\
&\quad + 2\int_{a}^{b} \frac{\lambda_{\theta_n^*}^{1/2}(x) - \hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*}^{1/2}(x)} \ddot{\lambda}_{\theta_n^*}(x)\overline{Y}_n(x)w_n(x)\,dx \to_P A(\theta_\lambda, a, b).
\end{aligned} \tag{4.29}
$$

From Lemma 4 and (4.15),

$$
\sup_{|\theta - \theta_\lambda| \le \delta_1} \int_{b}^{\tau} \frac{\lambda^{1/2}(x)}{\lambda_{\theta}^{3/2}(x)} \dot{\lambda}_{\theta}^2(x)\overline{Y}_n(x)w_n(x)\,dx = o_p(1) \tag{4.30}
$$

as $n \to \infty$ and $b \uparrow \tau$.

Moreover, from Lemmas 4 and 5,

$$
\begin{aligned}
\sup_{|\theta - \theta_\lambda| \le \delta_1} &\int_{b}^{\tau} \frac{\left|\hat{\lambda}_n^{1/2}(x) - \lambda^{1/2}(x)\right|}{\lambda_{\theta}^{3/2}(x)} \dot{\lambda}_{\theta}^2(x)\overline{Y}_n(x)w_n(x)\,dx \\
&= o_p\left(\int_{b}^{\tau} n^{-1/9}[1 - F(x)]^{-k_2/2}\overline{Y}_n(x)w_n(x)\,dx\right) \\
&= o_p\left(\int_{b}^{\tau} [1 - F(x)]^{1+k}\,dx\right) = o_p(1)
\end{aligned} \tag{4.31}
$$

as $b \uparrow \tau$ with k chosen to be the same as that in (4.7), noting that $|\hat{\lambda}_n^{1/2}(x) - \lambda^{1/2}(x)| \le |\hat{\lambda}_n(x) - \lambda(x)|^{1/2}$ and that $\sup_x\{w_n(x)n^{-1/9}[1 - F(x)]^{-\tilde{k}}\} \to 0$ for any $\tilde{k} > 0$. Combining (4.30) and (4.31) we get as $n \to \infty$ and $b \uparrow \tau$,

$$
\sup_{|\theta - \theta_\lambda| \le \delta_1} \int_{b}^{\tau} \frac{\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta}^{3/2}(x)} \dot{\lambda}_{\theta}^2(x)\overline{Y}_n(x)w_n(x)\,dx \to_P 0. \tag{4.32}
$$

A similar argument leads to

$$
\sup_{|\theta - \theta_\lambda| \le \delta_1} \int_{\tau_0}^{a} \frac{\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta}^{3/2}(x)} \dot{\lambda}_{\theta}^2(x)\overline{Y}_n(x)w_n(x)\,dx \to_P 0 \tag{4.33}
$$

as $n \to \infty$ and $a \downarrow \tau_0$. Likewise, it can also be shown using the second part of

(4.15) and Lemmas 4 and 5 that as $n \to \infty$, $a \downarrow \tau_0$ and $b \uparrow \tau$,

$$\sup_{|\theta-\theta_\lambda|\le\delta_2}\left\{\int_{\tau_0}^{a}+\int_{b}^{\tau}\right\}\left|\frac{\lambda_\theta^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda_\theta^{1/2}(x)}\right|\left|\ddot{\lambda}_\theta(x)\right|\bar{Y}_n(x)w_n(x)\,dx \to_P 0. \tag{4.34}$$

In view of (4.29), (4.32)–(4.34), we have

$$\begin{aligned}&\int_{\tau_0}^{\tau}\frac{\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*}^{3/2}(x)}\dot{\lambda}_{\theta_n^*}^2(x)\bar{Y}_n(x)w_n(x)\,dx\\&\quad+2\int_{\tau_0}^{\tau}\frac{\lambda_{\theta_n^*}^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n^*}^{1/2}(x)}\ddot{\lambda}_{\theta_n^*}(x)\bar{Y}_n(x)w_n(x)\,dx \to_P A(\theta_\lambda).\end{aligned} \tag{4.35}$$

From (2.6), (4.16) and the definition of w_n,

$$\sup_x\left\{w_n(x)\left|\lambda^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)\right|\right\}=o_p(n^{\varepsilon})\sup_x\left\{w_n(x)\left|\lambda(x)-\hat{\lambda}_n(x)\right|\right\}$$

for every $\varepsilon > 0$,

which, together with (4.35), (4.28) and Lemma 5, implies that $\hat{\theta}_n - \theta_n = o_p(n^{-1/4-\delta})$ for some $\delta > 0$. Similarly, it can also be shown that

$$\eta_n - \theta_n = o_p(n^{-1/4-\delta'}) \quad \text{for some } \delta' > 0. \tag{4.36}$$

Again by the mean value theorem, for some η_n^* between θ_n and η_n,

$$\begin{aligned}&\int_{\tau_0}^{\tau}\frac{\lambda^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda_{\theta_n}^{1/2}(x)}\dot{\lambda}_{\theta_n}(x)\bar{Y}_n(x)w_n(x)\,dx\\&\quad-\int_{\tau_0}^{\tau}\frac{\lambda^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda_{\eta_n}^{1/2}(x)}\dot{\lambda}_{\eta_n}(x)\bar{Y}_n(x)w_n(x)\,dx\\&=\int_{\tau_0}^{\tau}\left(\lambda^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)\right)\left(\frac{\ddot{\lambda}_{\eta_n^*}(x)}{\lambda_{\eta_n^*}^{1/2}(x)}-\frac{\dot{\lambda}_{\eta_n^*}^2(x)}{2\lambda_{\eta_n^*}^{3/2}(x)}\right)\bar{Y}_n(x)w_n(x)\,dx(\theta_n-\eta_n)\\&=o_p(n^{-1/2}),\end{aligned} \tag{4.37}$$

where the last equality follows from (4.16), (4.36) and Lemmas 4 and 5.

In view of (4.28), (4.35) and (4.37) we see that for (4.18) to hold, it suffices to show that

$$2\sqrt{n}\int_{\tau_0}^{\tau}\frac{\lambda^{1/2}(x)-\hat{\lambda}_n^{1/2}(x)}{\lambda_{\eta_n}^{1/2}(x)}\dot{\lambda}_{\eta_n}(x)\bar{Y}_n(x)w_n(x)\,dx \to_{\mathscr{D}} N(0, D(\theta_\lambda)). \tag{4.38}$$

Applying (2.6) we can express the left-hand side of (4.38) as

$$\begin{aligned}&\sqrt{n}\int_{\tau_0}^{\tau}\left[\lambda(x)-\hat{\lambda}_n(x)\right]\frac{\dot{\lambda}_{\eta_n}(x)\bar{Y}_n(x)}{\lambda_{\eta_n}^{1/2}(x)\lambda^{1/2}(x)}w_n(x)\,dx\\&\quad+\sqrt{n}\int_{\tau_0}^{\tau}\frac{\left[\lambda(x)-\hat{\lambda}_n(x)\right]^2}{\lambda_{\eta_n}^{1/2}(x)\lambda^{1/2}(x)\left[\lambda^{1/2}(x)+\hat{\lambda}_n^{1/2}(x)\right]^2}\dot{\lambda}_{\eta_n}(x)\bar{Y}_n(x)w_n(x)\,dx.\end{aligned} \tag{4.39}$$

The second term in (4.39) is $o_p(1)$ by Lemmas 4 and 5 and (4.16). Hence (4.38) is equivalent to

$$(4.40)\quad \sqrt{n}\int_{\tau_0}^{\tau}\left[\lambda(x)-\hat{\lambda}_n(x)\right]\frac{\dot{\lambda}_{\eta_n}(x)\overline{Y}_n(x)}{\lambda_{\eta_n}^{1/2}(x)\lambda^{1/2}(x)}w_n(x)\,dx \to_{\mathscr{D}} N(0,D(\theta_\lambda)).$$

Let $H_n(x)=\dot{\lambda}_{\eta_n}(x)/[\lambda_{\eta_n}^{1/2}(x)\lambda^{1/2}(x)]$. Then we can write the left-hand side of (4.40) as

$$\begin{aligned}
&\sqrt{n}\int_{\tau_0}^{\tau}\left[\lambda(x)-\hat{\lambda}_n(x)\right]H_n\overline{Y}_n(x)w_n(x)\,dx\\
&\quad=\sqrt{n}\int_{\tau_0}^{\tau}H_n(x)\overline{Y}_n(x)w_n(x)\left\{d\Lambda(x)-\frac{d\overline{N}_n(x)}{\overline{Y}_n(x)}\right\}\\
(4.41)\qquad&\quad\quad+\sqrt{n}\int_{\tau_0}^{\tau}H_n(x)\overline{Y}_n(x)w_n(x)\\
&\quad\quad\quad\times\left\{\frac{d\overline{N}_n(x)}{\overline{Y}_n(x)}-d\Lambda(x)-\hat{\lambda}_n(x)\,dx+\tilde{\lambda}_n(x)\,dx\right\}\\
&\quad\quad+\sqrt{n}\int_{\tau_0}^{\tau}H_n(x)\overline{Y}_n(x)w_n(x)\left[\lambda(x)-\tilde{\lambda}_n(x)\right]dx.
\end{aligned}$$

From (4.21) and (4.16), it follows that the last term on the right-hand side of the preceding equation is $o_p(1)$. Moreover, the second term can be rewritten as

$$\begin{aligned}
&\sqrt{n}\int_{\tau_0}^{\tau}H_n(x)\overline{Y}_n(x)w_n(x)\left\{d\left[\hat{\Lambda}_n(x)-\Lambda(x)\right]\right.\\
&\qquad\qquad\left.-d_n^{-1}\int_{-\infty}^{\infty}K\left(\frac{x-u}{d_n}\right)d\left[\hat{\Lambda}_n(u)-\Lambda(u)\right]dx\right\}\\
(4.42)\quad&=\sqrt{n}\int_{\tau_0}^{\tau}H_n(x)\overline{Y}_n(x)w_n(x)\left\{d\left[\hat{\Lambda}_n(x)-\Lambda(x)\right]\right.\\
&\qquad\qquad\left.-d_n^{-2}\int_{-\infty}^{\infty}\left[\hat{\Lambda}_n(u)-\Lambda(u)\right]K'\left(\frac{x-u}{d_n}\right)du\,dx\right\}\\
&=\sqrt{n}\int_{\tau_0}^{\tau}H_n(x)\overline{Y}_n(x)w_n(x)\,d\left\{\hat{\Lambda}_n(x)-\Lambda(x)\right.\\
&\qquad\qquad\left.-\int_{-\infty}^{\infty}\left[\hat{\Lambda}_n(u)-\Lambda(u)\right]d_n^{-1}K\left(\frac{x-u}{d_n}\right)du\right\}
\end{aligned}$$

recalling that $\hat{\Lambda}_n(x) = \int_{-\infty}^{x} \bar{Y}_n^{-1}(s)\, d\bar{N}_n(s)$. In view of Lemma 6,

$$\sup_{x:\, w_n(x)>0} \left| \hat{\Lambda}_n(x) - \Lambda(x) - \int_{-\infty}^{\infty} \left[\hat{\Lambda}_n(u) - \Lambda(u)\right] d_n^{-1} K\left(\frac{x-u}{d_n}\right) du \right|$$
$$(4.43) \qquad \sup_{\substack{w_n(x)>0,\\ |y|\le c d_n}} \left|\hat{\Lambda}_n(x) - \Lambda(x) - \hat{\Lambda}_n(x+y) + \Lambda(x+y)\right| = o_p(n^{-1/2-\delta})$$

for some $\delta > 0$. Thus, it follows from (4.42), (4.43) and the easily verifiable fact that $\int_{\tau_0}^{\tau} |d(H_n(x)\bar{Y}_n(x)w_n(x))| = o_p(n^{\varepsilon})$ for every $\varepsilon > 0$, that

$$(4.44) \qquad \int_{\tau_0}^{\tau} H_n(x)\bar{Y}_n(x)w_n(x)\left\{\frac{d\bar{N}_n(x)}{\bar{Y}_n(x)} - d\Lambda(x) - \hat{\lambda}_n(x)\, dx + \tilde{\lambda}_n(x)\, dx\right\} = o_p(n^{-1/2}).$$

Therefore it remains to show the weak convergence of the first term on the right-hand side of (4.41). We shall prove this by applying Rebolledo's martingale central limit theorem [cf. Gill (1980), page 18]. To do so, we need to replace w_n by the predictable weight function

$$(4.45) \qquad v_n(x) = \begin{cases} 1, & \text{if } \bar{N}_n(x - d_n c) \ge c_1(\log n)^{-\alpha_1} \\ & \quad \text{and } \bar{Y}_n(x) \ge c_2(\log n)^{-\alpha_2}, \\ 0, & \text{otherwise,} \end{cases}$$

where α_i $i = 1, 2$ are the same constants as those in (4.2) which defines w_n. Also define stopping times $T_n = \inf\{x\colon \bar{Y}_n(x) < c_2(\log n)^{-\alpha_2}\}$. Then by the Rebolledo martingale central limit theorem we can easily show that

$$(4.46) \qquad \sqrt{n}\int_{\tau_0}^{T_n \wedge \cdot} H_n(x)v_n(x)\left\{d\bar{N}_n(x) - \bar{Y}_n(x)\, d\Lambda(x)\right\} \to_{\mathscr{D}[\tau_0, \tau]} W(\cdot),$$

where W is a zero-mean Gaussian martingale with

$$EW^2(t) = \int_{\tau_0}^{t} \dot{\lambda}_\theta^2(x)\left(\lambda_\theta(x)\lambda(x)\right)^{-1}\Gamma(x)\, dF(x).$$

Since W is continuous,

$$\sqrt{n}\int_{\tau_0}^{\tau} H_n(x)w_n(x)\left\{d\bar{N}_n(x) - \bar{Y}_n(x)\, d\Lambda(x)\right\}$$
$$= \sqrt{n}\int_{\tau_0}^{T_n - cd_n} H_n(x)v_n(x)\left\{d\bar{N}_n(x) - \bar{Y}_n(x)\, d\Lambda(x)\right\} \to_{\mathscr{D}} W(\tau).$$

Hence (4.18) follows. A similar argument leads to (4.19). □

5. Concluding remarks. We have introduced in this paper a Hellinger-type distance for hazard rate functions. For parametric inference, the distance can be used to define minimum distance estimators of unknown parameters. Among its features, it accommodates censored data in a natural way and exhibits flexibility in stabilizing tail fluctuation. In particular, it does not require that censoring variables be identically distributed. Thus it is suitable for applications in areas such as clinical trials and industrial life tests.

When the distance $\int[\lambda_\theta^{1/2}(t) - \hat{\lambda}_n^{1/2}(t)]^2\bar{Y}_n(t)\,dt$ as a function of θ is convex, we can use the classical Newton–Raphson method to compute $\hat{\theta}_n$. However, computational complication will arise when local minima exist. In low-dimensional situations, we can first use a systematic grid search to locate the region where the global minimizer $\hat{\theta}_n$ belongs and then apply a more efficient algorithm such as the Newton–Raphson method. Recently, Beran and Millar (1987) proposed a general stochastic search method for finding solutions defined as minimizers of certain distances and have investigated its convergence properties. Their method can certainly be applied to our setting. The readers are referred to their paper for details.

As we have shown in the preceding developments, when the underlying density (hazard) function belongs to the parametric family, the estimator $\hat{\theta}_n$ is asymptotically as efficient as the usual maximum likelihood estimator. On the other hand, if the parametric assumption fails, then $\hat{\theta}_n \to \theta_0$, which provides a closest fit in the sense that λ_{θ_0} has the shortest Hellinger-type distance; that is,

$$\int\left[\lambda_{\theta_0}^{1/2}(t) - \lambda^{1/2}(t)\right]^2[1 - F(t)]\Gamma(t)\,dt$$
$$= \min_\theta \int\left[\lambda_\theta^{1/2}(t) - \lambda^{1/2}(t)\right]^2[1 - F(t)]\Gamma(t)\,dt.$$

This may be interpreted as a robustness property.

Beran (1977) as well as Yang (1991) proved convolution and asymptotic minimax properties for their estimators. We believe that analogous results should also hold for our estimator, though we have not been able to prove them yet. It is also worth mentioning that the estimator discussed in Section 4 generally estimates a different parameter than that estimated by the usual MHDE when the underlying density does not belong to the prescribed parametric family, even in the absence of censoring. Therefore, it is usually rather difficult to compare the current estimator with Beran's (1977) or Yang's (1991) estimators.

Finally, the hazard rate function approach can also handle left truncated data, or, more generally, both left truncated and right censored data. A Nelson-type estimator of cumulative hazard function can be found in Wang, Jewell and Tsai (1986) and Keiding and Gill (1990) for the truncated data, and in Lai and Ying (1991) for both left truncated and right censored data. Let $\hat{\Lambda}_n^{TC}$ denote the Nelson-type cumulative hazard function estimator for the left truncated and right censored model and let $\bar{Y}_n^{TC}$ denote the corresponding risk set size process. Then Ramlau-Hansen's (1983) smoothing method can be applied to obtain a hazard rate function estimator, say $\hat{\lambda}_n^{TC}$. Therefore, we can use the Hellinger-type distance

$$\int_a^b\left[\sqrt{\hat{\lambda}_n^{TC}(t)} - \sqrt{\lambda_\theta(t)}\right]^2\bar{Y}_n^{TC}(t)\,dt$$

to obtain a minimum distance estimator $\hat{\theta}_n^{TC}$. Since similar counting process and its associated martingale theory has already been developed for the left truncated and right censored model, it is straightforward that the results of Sections 3 and 4 can be extended.

Acknowledgment. The author wishes to express his gratitude to the referees for their thorough reading of the paper and many valuable comments.

REFERENCES

BERAN, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BERAN, R. J. and MILLAR, P. W. (1987). Stochastic estimation and testing. *Ann. Statist.* **15** 1131–1154.

BORGAN, O. (1984). Maximum likelihood estimation in parametric counting process models with applications to censored failure time data. *Scand. J. Statist.* **11** 1–16.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York.

EFRON, B. and JOHNSTONE, I. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18** 38–62.

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

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

KEIDING, N. and GILL, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18** 582–602.

LAI, T. L. and YING, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regions. *J. Multivariate Anal.* **27** 334–358.

LAI, T. L. and YING, Z. (1991). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19** 417–442.

POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27** 43–70.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

SELLKE, T. and SIEGMUND, D. (1983). Sequential analysis of proportional hazards model. *Biometrika* **70** 315–326.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Process with Applications to Statistics*. Wiley, New York.

SIMPSON, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *J. Amer. Statist. Assoc.* **82** 802–807.

TAMURA, R. and BOOS, D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *J. Amer. Statist. Assoc.* **81** 223–229.

WANG, M. C., JEWELL, N. P. and TSAI, W. Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14** 1597–1605.

YANG, S. (1991). Minimum Hellinger distance estimation of parameter in the random censorship model. *Ann. Statist.* **19** 579–602.

DEPARTMENT OF STATISTICS
101 ILLINI HALL
725 SOUTH WRIGHT STREET
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820