

## ASYMPTOTICALLY OPTIMAL ESTIMATION OF SMOOTH FUNCTIONALS FOR INTERVAL CENSORING, CASE 2

BY RONALD GESKUS AND PIET GROENEBOOM

*Municipal Health Service, Amsterdam and Delft University of Technology*

For a version of the interval censoring model, case 2, in which the observation intervals are allowed to be arbitrarily small, we consider estimation of functionals that are differentiable along Hellinger differentiable paths. The asymptotic information lower bound for such functionals can be represented as the squared  $L_2$ -norm of the canonical gradient in the observation space. This canonical gradient has an implicit expression as a solution of an integral equation that does not belong to one of the standard types. We study an extended version of the integral equation that can also be used for discrete distribution functions like the nonparametric maximum likelihood estimator (NPMLE), and derive the asymptotic normality and efficiency of the NPMLE from properties of the solutions of the integral equations.

**1. Introduction.** In the interval censoring problem, one wants to obtain information on some distribution  $F$ , often representing an event time distribution, based on a sample of random intervals  $J_1, \dots, J_n$  in which unobservable  $X_1, \dots, X_n \sim F$  are known to be contained. In case 1, we have a sample of observation times  $T_i$  and we know whether  $X_i$  is smaller or larger than the corresponding observation time  $T_i$ . More formally: we observe  $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ , with  $\Delta_i = 1_{\{X_i \leq T_i\}}$ . Case 2 is usually denoted as the situation with two observation times  $(U_i, V_i)$  and the information whether  $X_i$  is left of  $U_i$ , between  $U_i$  and  $V_i$  or right of  $V_i$ .

For case 1, also denoted by *current status data*, quite a lot is known. It is already shown in Ayer, Brunk, Ewing, Reid and Silverman (1955) and van Eeden (1956) that there exists a one-step procedure for calculation of the nonparametric maximum likelihood estimator (NPMLE)  $\hat{F}_n$  of the distribution function  $F$ , based on isotonic regression theory. The asymptotic distribution of  $\hat{F}_n(t_0)$ , for fixed  $t_0 \in \mathbb{R}$ , is derived in Part II, Chapter 5 of Groeneboom and Wellner (1992);  $n^{1/3}$  is the obtained convergence rate. The same chapter discusses convergence properties of the NPMLE  $\mu(\hat{F}_n)$  of the mean  $\mu(F)$  and it is shown that, under some extra conditions,

$$\sqrt{n} \left( \mu(\hat{F}_n) - \mu(F) \right) \rightarrow_{\mathcal{D}} N \left( 0, \int \frac{F(x)[1 - F(x)]}{g(x)} dx \right) \text{ as } n \rightarrow \infty,$$

---

Received March 1996; revised February 1999.

AMS 1991 subject classifications. 60F17, 62E20, 62G05, 62G20, 45A05.

Key words and phrases. Nonparametric maximum likelihood, empirical processes, asymptotic distributions, asymptotic efficiency, integral equations.

with  $g(x)$  the density of the distribution of the observation times, and  $\rightarrow_{\mathcal{D}}$  denoting convergence in distribution. Huang and Wellner (1995) prove a similar result for functionals that are linear in  $F$ :  $K(F) = \int c dF$ . Then an extra factor  $c'(x)$  appears in the integral formula for the limit variance. Groeneboom's proof for the mean uses the convergence rate of the supremum distance between the NPMLE and the underlying distribution function, which is replaced by an easier argument based on  $L_2$ -distance properties of  $\hat{F}_n$  and smoothness of  $c'(x)/g(x)$  in Huang and Wellner's proof.

In case 2 one may expect better estimation results than in case 1, since one has more information on the location of the  $X_i$ 's. However, both theoretical as well as practical aspects of the problem are more complicated. Only iterative procedures are available for computation of the NPMLE  $\hat{F}_n$  of  $F$ . The *iterative convex minorant algorithm*, as introduced by Groeneboom in part II of Groeneboom and Wellner (1992), converges quickly in computer experiments, and a slight modification of this algorithm is shown to converge in Jongbloed (1998).

When considering the asymptotic distribution of the NPMLE  $\hat{F}_n(t_0)$ , a distinction should be made. If the relative amount of mass of the  $(U, V)$ -distribution near the diagonal point  $(t_0, t_0)$ , compared to the amount of mass of  $F$  near  $t_0$  is very small we are more in a case 1-type situation and we still have a  $n^{1/3}$  convergence rate [Wellner (1995)]. The limit distribution of  $\hat{F}_n(t_0)$  has been established in Groeneboom (1996) for the situation that  $U$  and  $V$  remain bounded away. If the observation time distribution has sufficient mass along the diagonal, the convergence rate increases to  $(n \log n)^{1/3}$  [see Groeneboom and Wellner (1992)]. There is a conjecture on the asymptotic distribution of  $\hat{F}_n(t_0)$  in this situation, but the proof is still incomplete. Another "estimator" is  $F_n^{(1)}(t_0)$ , which is obtained by doing one step in the iterative convex minorant algorithm, with the true underlying distribution  $F$  as starting value. Of course, this procedure, which does not even lead to an estimator in the strict sense, has no practical value. However, it may be relevant for theoretical purposes, for the asymptotic distribution of  $F_n^{(1)}(t_0)$  is known, and it is conjectured to have the same asymptotic distribution as the NPMLE.

Contrary to  $F(t_0)$ , for case 1 as well as case 2, the mean  $\mu(F)$  is a smooth linear functional. This means that it is differentiable along Hellinger differentiable paths of distributions. For any functional having this property, one can derive a Hájek–LeCam convolution theorem type information lower bound, giving the best possible limit variance that can be attained under  $\sqrt{n}$  convergence rate. See for example, van der Vaart (1991), part I of Groeneboom and Wellner (1992) or Bickel, Klaassen, Ritov and Wellner (1993) for the general theory and the application to case 1. For case 1, an explicit expression of the information lower bound can be derived and is given by

$$\int c'(x)\phi(x) dx,$$

with  $\phi(x)$  being the function

$$\phi(x) = c'(x) \frac{F(x)[1 - F(x)]}{g(x)}.$$

So this lower bound is attained by the NPMLE.

The function  $\phi$  appearing in the information lower bound has an analogue in case 2. However, contrary to case 1, an explicit formula for  $\phi$  is unknown and unlikely to exist, except for some very special choices of the distributions of  $X$  and  $(U, V)$  [see Geskus (1997)]. Nevertheless it can be proved, without knowing  $\phi$  explicitly, that the NPMLE  $K(\hat{F}_n)$ , estimating the smooth functional  $K(F)$ , shows asymptotically optimal behavior in case 2 as well, using properties of the integral equation for  $\phi$ . Just as in the problem of estimation of  $F(t_0)$ , two situations can be distinguished, showing different behavior. In Geskus and Groeneboom (1996, 1997), together from now on denoted as GG, the simplest case is treated, with the  $(U, V)$ -distribution having no mass along the diagonal. Smoothness properties of  $\phi$  are derived from the structure of the integral equation and are sufficient to give the proof. The technical report Geskus and Groeneboom (1995) contains both papers. See also Geskus (1997), which contains a rather extensive treatment of the application of the general lower bound theory to case 1 and this simple version of case 2. In the present paper, we treat the situation where the observation times can be arbitrarily close. In the long run we will have observations for which we get very close to observing  $X$  itself, a situation that never occurs in case 1 and the case 2 treated in GG. Now the analysis is much more complicated, since we really have to deal with the singularity of the integrand of the integral equation on the diagonal.

Our general outline is as follows. In Section 2, after a specification of the problem, the relevant lower bound calculations are given. In Section 3 we show asymptotic efficiency of the NPMLE; here we benefit from some recent results on the Hellinger distance between the NPMLE and the underlying distribution function in van de Geer (1996). Section 3 also contains our main result. Theorem 3.2, showing that the NPMLE  $K(\hat{F}_n)$  of a smooth functional  $K(F_0)$  of the underlying distribution function  $F_0$  at  $\sqrt{n}$ -rate, is asymptotically normal and has an asymptotic variance that is equal to the information lower bound.

In Section 4 we present some simulation results showing that the variance of the NPMLE is already very close to the information lower bound for reasonable sample sizes. In this section we also show some pictures of the solution of the integral equation and its derivative for the case that the distributions are uniform [ $F$  uniform and the distribution of  $(U, V)$  uniform on the upper triangle of the unit square] and the smooth functional corresponds to the mean. These graphs are compared to graphs of a corresponding solution of the integral equation when  $F$  is replaced by the NPMLE (and the equation is transformed into another equation in an inverse scale). Finally, in the Appendix, some technical results are proved that are needed in Section 3.

Of course, situations with more than two observation times for each unobservable event time may occur as well. This is usually denoted as the case  $k$  situation. However, only the two observation times immediately around the event time give relevant information, so it very much resembles case 2. We will only consider case 2. See also GG, where case  $k$  is briefly discussed.

Unless otherwise stated, all norms in this paper should be read as  $L_2$ -norms. The dominating measure varies, but is denoted by a subscript, or should be clear from the context. The  $L_2^0$ -space denotes the subspace of functions that integrate to zero.

**2. The model. Lower bound considerations.** We will start with a brief formulation of the problem and a list of assumptions that are needed in order to perform the lower bound calculations. Moreover, a key equation on which our analysis will be based is given. A more elaborate introduction, including a derivation of this equation, can be found in GG and Geskus (1997).

2.1. *The model. General lower bound theory for smooth functionals.* Let  $F$  be a distribution function. We are interested in estimation of some functional  $K(F)$ . However, instead of a sample  $X_1, \dots, X_n \sim F$ , we are only able to observe the sample  $(U_1, V_1, \Delta_1, \Gamma_1), \dots, (U_n, V_n, \Delta_n, \Gamma_n)$  with  $\Delta_i = 1_{\{X_i \leq U_i\}}$  and  $\Gamma_i = 1_{\{U_i < X_i \leq V_i\}}$ . The following model assumptions are made:

(M1) Let  $K > 0$  and let  $S$  be a bounded interval  $\subset \mathbb{R}$ .  $F$  is contained in the class

$$\mathcal{F}_S := \left\{ F \mid \text{support}(F) \subset S, F \text{ absolutely continuous, } \sup_x |f(x)| \leq K \right\}.$$

Here  $F$  is the distribution on which we want to obtain information; however, we do not observe  $X_i \sim F$  directly.

(M2) Instead, we observe the pairs  $(U_i, V_i)$ , with distribution function  $H$ . Now  $H$  is contained in  $\mathcal{H}$ , the collection of all two-dimensional distributions on  $\{(u, v) \mid u < v\}$ , absolutely continuous with respect to two-dimensional Lebesgue measure and such that  $(U_i, V_i)$  is independent of  $X_i$ . Let  $h$  denote the density of  $(U_i, V_i)$ , with marginal densities and distribution functions  $h_1, H_1$  and  $h_2, H_2$  for  $U_i$  and  $V_i$ , respectively.

(M3) If both  $H_1$  and  $H_2$  put zero mass on some set  $A$ , then  $F \in \mathcal{F}_S$  has zero mass on  $A$  as well, so  $F \ll H_1 + H_2$ . This means that  $F$  does not have mass on sets in which no observations can occur.

Typically, under (M1),  $F$  has support  $[0, M]$ , and  $H$  has its mass on the triangle  $\{(u, v) \mid 0 \leq u < v \leq M\}$ . From now on we will restrict attention to this typical situation. More general choices of the support and the situation with  $H_1 + H_2$  having larger support than  $F$  have been treated in Geskus (1997), but are similar in essence. Without condition (M3), the functionals in which we are interested are not well defined. Moreover, if  $F$  has probability

mass on a region in which no observations can occur, one cannot estimate the structure of  $F$  on this region consistently. In many survival studies, events do occur outside the domain of observation, making estimation of functionals like the mean an impossible task.

The observable random vectors  $(U_i, V_i, \Delta_i, \Gamma_i)$  have density

$$q_{F,H}(u, v, \delta, \gamma) = h(u, v) F(u)^\delta (F(v) - F(u))^\gamma (1 - F(v))^{1-\delta-\gamma}$$

with respect to  $\lambda_2 \otimes \nu_2$ , where  $\nu_2$  denotes counting measure on the set  $\{(0, 1), (1, 0), (0, 0)\}$ . The letter “M” in conditions (M1) to (M3) stands for “model.”

With respect to the functional to be estimated we assume

(F1)  $K$  is differentiable along Hellinger differentiable paths of distributions from  $\mathcal{F}_S$ .

The canonical gradient for this functional is denoted by  $\tilde{\kappa}_F$ . What functionals satisfy this pathwise differentiability requirement? An important class of functionals are the functionals that are linear in  $F$ ,

$$K(F) = \int c(x) dF(x).$$

All moment functionals  $F \mapsto \int x^k dF(x)$  belong to this class. Estimation of the distribution function at a fixed point concerns a linear functional as well: for  $K(F) = F(t_0)$  we have  $c(x) = 1_{[0, t_0]}(x)$ . In Bickel, Klaassen, Ritov and Wellner (1993), Proposition A.5.2, it is shown that linear functionals on  $\mathcal{F}_S$  with

$$\sup_{F \in \mathcal{F}_S} E_F c(X)^2 < \infty$$

are pathwise differentiable at any  $F \in \mathcal{F}_S$ , with canonical gradient

$$\tilde{\kappa}_F(x) = c(x) - \int c(x) dF(x).$$

So if we were able to observe the  $X_i$ 's directly, we would obtain the information lower bound

$$\|c(X) - E_F(c(X))\|_F^2.$$

For nonlinear functionals, there is no general method that immediately establishes pathwise differentiability and supplies the formula for the canonical gradient. An example of a nonlinear functional to which our theory can be applied is

$$K(F) = \int F^2(x) w(x) dx.$$

If the function  $w$  is bounded, a slight extension of the proof in Bickel, Klaassen, Ritov and Wellner (1993), as given in Geskus (1997), shows that this functional has canonical gradient

$$\tilde{\kappa}_F(x) = 2 \int_{s=x}^M F(s) w(s) ds - \int_{x=0}^M \int_{s=x}^M 2F(s) w(s) ds dF(x).$$

In order to show that the NPMLE  $K(\hat{F}_n)$  asymptotically attains the lower bound, we have to make the following extra assumption:

$$(F2) \quad K(G) - K(F) = \int \tilde{\kappa}_F(x) d(G - F)(x) + \mathcal{O}(\|G - F\|_\lambda^2),$$

for all distribution functions  $G$  with support contained in  $[0, M]$ , with  $\lambda$  denoting Lebesgue measure on  $\mathbb{R}$ . For linear functionals (F2) holds without the  $\mathcal{O}$ -term. However, the functional

$$K(F) = \int F^2(x) w(x) dx$$

also satisfies (F2).

In the interval censoring model we do not observe  $X \sim F$  directly, and the estimated  $K(F)$  is only implicitly defined as a functional  $\Theta(Q_{F,H})$  on the class of probability measures on the observation space, with  $H$  acting as a nuisance parameter. What about differentiability and information lower bounds in this model? The score operator  $L_1$ , relating the censoring model to the unattainable model without censoring is, in our situation,

$$(2.1) \quad \begin{aligned} [L_1 a](u, v, \delta, \gamma) &= E[a(X) \mid U = u, V = v, \Delta = \delta, \Gamma = \gamma] \\ &= \frac{\delta \int_0^u a dF}{F(u)} \\ &\quad + \frac{\gamma \int_u^v a dF}{F(v) - F(u)} + \frac{(1 - \delta - \gamma) \int_v^M a dF}{1 - F(v)} \quad \text{a.e.} - [Q_{F,H}] \end{aligned}$$

This operator may be defined on  $L_2(F)$ , with range in  $L_2(Q_{F,H})$ . However, since it relates scores, our main interest lies in the domain  $L_2^0(F)$ . Then its range is contained in  $L_2^0(Q_{F,H})$ . The adjoint of  $L_1$  on  $L_2^0(Q_{F,H})$  can be written as  $[L_1^* b](x) = E[b(U, V, \Delta, \Gamma) \mid X = x]$  and we get

$$(2.2) \quad \begin{aligned} [L_1^* b](x) &= \int_{u=x}^M \int_{v=u}^M b(u, v, 1, 0) h(u, v) dv du \\ &\quad + \int_{u=0}^x \int_{v=x}^M b(u, v, 0, 1) h(u, v) dv du \\ &\quad + \int_{u=0}^x \int_{v=u}^x b(u, v, 0, 0) h(u, v) dv du \quad \text{a.e.} - [F]. \end{aligned}$$

Now we have pathwise differentiability of  $\Theta(Q_{F,H})$  if and only if

$$\tilde{\kappa}_F \in \mathcal{R}(L_1^*)$$

and if this holds, then the canonical gradient is the unique element  $\tilde{\theta}_{F,H}$  in  $\overline{\mathcal{R}(L_1)} \subset L_2^0(Q_{F,H})$  satisfying

$$(2.3) \quad L_1^* \tilde{\theta}_{F,H} = \tilde{\kappa}_F.$$

[See van der Vaart (1991).]

Many functionals that are pathwise differentiable in the model without censoring lose this property in the interval censoring model. Due to the

smoothness of the adjoint operator, any functional  $K$  with a canonical gradient that is not a.e. equal to a continuous function cannot be obtained under  $L_1^*$ . So not all linear functionals remain pathwise differentiable. For example,  $K(F) = F(t_0)$ , with canonical gradient  $1_{[0, t_0]}(\cdot) - F(t_0)$ , is discontinuous at  $t_0$ , and therefore does not belong to the range of  $L_1^*$ . This corresponds with  $F(t_0)$  not being estimable at  $\sqrt{n}$ -rate. However, functionals with a canonical gradient that is sufficiently smooth will be shown to remain differentiable under censoring. Hence for these functionals the information lower bound theory holds.

If the canonical gradients  $\tilde{\theta} = \tilde{\theta}_{F, H}$  and  $\tilde{\kappa}_F$  satisfy some extra conditions, the information lower bound  $\|\tilde{\theta}\|_{Q_{F, H}}^2$  has an alternative formulation.

**THEOREM 2.1.** *Let  $\tilde{\theta}$  be contained in  $\mathcal{R}(L)$ , say  $\tilde{\theta} = La_0$  for some  $a_0 \in L_2^0(F)$ . Assume that the function  $x \mapsto \tilde{\kappa}_F(x)$  is differentiable with bounded derivative. Then we have*

$$\begin{aligned} \|\tilde{\theta}\|_{Q_{F, H}}^2 &= \langle a_0, \tilde{\kappa}_F \rangle_F \\ &= \int_0^M \tilde{\kappa}'_F(x) \phi_0(x) dx \end{aligned}$$

with  $\phi_0(x) = \int_x^M a_0(t) dF(t)$ .

For the proof, see Theorem 3.3 in Geskus and Groeneboom (1995).

This theorem holds more generally. However, in the interval censoring model, both case 1 and case 2, we have the extra property that the function,

$$\phi(x) = \int_x^M a(t) dF(t) \quad \text{with } a \in L_2^0(F) \text{ as defined,}$$

also appears explicitly in the score operator  $L_1$ . Therefore it will play an important role. It will be called the *integrated score function*. From its definition we know that  $\phi$  satisfies  $\phi(0) = \phi(M) = 0$  and that  $\phi$  is continuous for  $F \in \mathcal{F}_S$ .

In Section 2.2 we will pay attention to the structure of the lower bound. Section 3 will be devoted to showing that the NPMLE  $\hat{\Theta}_n$  of  $\Theta(Q_{F, H})$  satisfies

$$\sqrt{n} (\hat{\Theta}_n - \Theta(Q_{F, H})) \rightarrow_{\mathcal{D}} N(0, \|\tilde{\theta}\|_{Q_{F, H}}^2).$$

**2.2. Lower bounds for interval censoring case 2.** We restrict ourselves to the case  $\tilde{\theta} \in \mathcal{R}(L_1)$ . So the case  $\tilde{\theta} \in \overline{\mathcal{R}(L_1)} \setminus \mathcal{R}(L_1)$  will not be considered. Solvability of the equation

$$(2.4) \quad \tilde{\kappa}_F(x) = [L_1^* L_1 a](x) \quad \text{a.e.-}F$$

in the variable  $a \in L_2^0(F)$  will be investigated. The support of  $F$  may consist of a finite number of disjoint intervals. However, (2.4) is not defined on intervals where  $F$  does not put mass, and these intervals do not play any further role. So without loss of generality, we may assume the support of  $F$  to

consist of one interval  $[0, M]$ . By the structure of the score operator  $L_1$  this can be reformulated as an equation in  $\phi$ . If we suppose (2.4) to hold for all  $x \in [0, M]$ , taking derivatives on both sides yields the following integral equation:

$$(2.5) \quad \phi(x) + d_F(x) \left[ \int_{t=0}^x \frac{\phi(x) - \phi(t)}{F(x) - F(t)} h(t, x) dt - \int_{t=x}^M \frac{\phi(t) - \phi(x)}{F(t) - F(x)} h(x, t) dt \right] = k(x) d_F(x),$$

with  $d_F(x)$  being the function

$$d_F(x) = \frac{F(x)[1 - F(x)]}{h_1(x)[1 - F(x)] + h_2(x)F(x)},$$

writing  $k(x)$  instead of  $\tilde{\kappa}'_F(x)$ . Although  $k$  may depend on  $F$ , we do not explicitly express this dependence. This is done, since in proving asymptotic efficiency of the NPMLE we have to consider (2.5) for convex combinations  $F = (1 - \alpha)F_0 + \alpha\hat{F}_n$ , where  $F_0 \in \mathcal{F}_S$  (the unknown distribution) is continuous and  $\hat{F}_n$  (the NPMLE of  $F_0$ ) is purely discrete. Solvability and structure of the solution to (2.5) will be investigated for such combinations, with  $k$  still determined by the underlying distribution  $F_0$  (so  $k = \tilde{\kappa}'_{F_0}$ ). Apart from the model conditions (M1) to (M3), some extra conditions will have to be introduced in order to make the proofs in this section possible. For the distributions we assume

- (D1)  $h_1(x) + h_2(x) > 0$  for all  $x \in [0, M]$ .
- (D2) The function  $h(u, v)$  is continuous. The partial derivatives  $\Delta_t^1(x) = (\partial/\partial x)h(x, t)$  and  $\Delta_t^2(x) = (\partial/\partial x)h(t, x)$  exist, except for at most a finite number of points  $x$ , where left and right derivatives with respect to  $x$  do exist for each  $t$ . The derivatives are bounded, uniformly in  $t$  and  $x$ .
- (D3)  $F$  is a nondegenerate distribution function with at most finitely many points of jump  $x_i \in (0, M)$ . Let  $D = \{x_0 = 0, x_1, \dots, x_m, x_{m+1} = M\}$  denote the ordered set of jump points of  $F$ , augmented with the endpoints of the interval  $[0, M]$ . We assume that  $F$  is differentiable between jumps, except for at most a finite number of points, where left and right derivatives exist. Everywhere outside  $D$ , the derivative is bounded and greater than or equal to  $c$  for some  $c > 0$  (so we assume  $f_0(x) \geq c$  for all  $x \in [0, M]$ ). The set of points of jump may be empty. Note that if  $F$  has jumps, we assume that  $F$  has derivative greater than or equal to  $c$  also on the (nonempty) intervals  $(0, x_1)$  and  $(x_m, M)$  (where we allow  $x_1 = x_m$ , though).

For the functional we should have

- (F3)  $k$  is differentiable, except for at most a finite number of points  $x$ , where left and right derivatives exist. The derivative is bounded, uniformly in  $x$ .



Note that the letter “D” in conditions (D1) to (D3) stands for “distribution” and the letter “F” in (F1) to (F3) is for “functional.”

Of course, (D2) implies continuity of  $h_1$  and  $h_2$ : (D1) is the equivalent of  $g > 0$  in case 1 and is needed. It implies that  $d_F$  is bounded. In case 1, the function  $\phi$  has an explicit representation of the form

$$\phi(x) = k \frac{F[1 - F]}{g},$$

whereas in case 2,  $\phi$  can only be expressed implicitly as a solution to (2.5). If (2.5) is solvable, its solution  $\phi$  can be shown to contain a factor  $F(1 - F)$ , just as in case 1. The structure of  $d_F$  already suggests this factor to be present. Validity of the factorization is shown by inserting

$$\phi = F(1 - F)\xi$$

in (2.5). Some reordering yields an integral equation in  $\xi$ , which will be shown to be solvable. This  $\xi$ -equation has the following form:

$$(2.6) \quad \xi(x) + c_F(x) \left[ \int_{t=0}^x \frac{\xi(x) - \xi(t)}{F(x) - F(t)} h^*(t, x) dt - \int_{t=x}^M \frac{\xi(t) - \xi(x)}{F(t) - F(x)} h^\circ(x, t) dt \right] = k(x)c_F(x),$$

with  $c_F(x)$  given by

$$(2.7) \quad c_F^{-1}(x) = \int_{t=0}^x [1 - F(t)]h(t, x) dt + \int_{t=x}^M F(t)h(x, t) dt \\ = h_2(x)E\{1 - F(U) \mid V = x\} + h_1(x)E\{F(V) \mid U = x\}$$

and

$$(2.8) \quad h^*(t, x) = F(t)[1 - F(t)]h(t, x) \quad \text{if } t \leq x, \\ h^\circ(x, t) = F(t)[1 - F(t)]h(x, t) \quad \text{if } x \leq t.$$

This equation is similar in structure to the  $\phi$ -equation. So the lemmas and theorems in the remainder of this section apply to both the  $\phi$ -equation (2.5) and the  $\xi$ -equation (2.6). Most of the proofs will only be given for the  $\phi$ -equation.

Unlike the situation treated in GG, we now assume that the observation density does have mass along the diagonal. This has the consequence that the integral equation may no longer be a Fredholm integral equation. However, we first consider a “desingularized” integral equation, to which the theory on Fredholm integral equations of the second kind can be applied [Kress (1989)].

If  $F$  has jumps, the solution of the integral equation will in general also have jumps. However, the key observation in analyzing the integral equation and in proving the efficiency of the NPMLE is that, even when  $F$  has discontinuities, we can make a change of scale in such a way that the solution of the integral equation can be extended to a Lipschitz function in the transformed scale.

We first introduce some notation. Let  $G(t) = F^{-1}(t)$ ,  $t \in [0, 1]$ , with a derivative  $g$  which exists except for at most a finite number of points, where, however,  $G$  has left and right derivatives. Furthermore, let  $\bar{k}(t) = k(G(t))$ ,  $\bar{H}(t, u) = H(G(t), G(u))$  and likewise  $\bar{h}(t, u) = h(G(t), G(u))$ , and let  $\bar{d}_F$  be defined by

$$(2.9) \quad \bar{d}_F(t) = \frac{t(1-t)}{(1-t)\bar{h}_1(t) + t\bar{h}_2(t)},$$

where  $\bar{h}_i = h_i \circ G$ ,  $i = 1, 2$ . Note that, if  $F$  has jumps,  $\bar{d}_F \neq d_F \circ G$ . Also note that  $\bar{k}$ ,  $\bar{d}$  and  $\bar{h}$  are continuous. In a similar way, we define

$$\bar{c}_F(t)^{-1} = \int_0^t (1-s)\bar{h}(s, t) dG(s) + \int_t^1 s\bar{h}(t, s) dG(s)$$

and

$$(2.10) \quad \begin{aligned} \bar{h}^*(t, u) &= t(1-t)\bar{h}(t, u) \quad \text{if } t \leq u, \\ \bar{h}^\circ(u, t) &= t(1-t)\bar{h}(u, t), \quad \text{if } u \leq t. \end{aligned}$$

We have the following lemma.

LEMMA 2.1. (i) *The integral equation*

$$(2.11) \quad \begin{aligned} \bar{\phi}_\varepsilon(t) = \bar{d}_F(t) &\left\{ \bar{k}(t) - \int_0^t \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')}{(t-t') \vee \varepsilon} \bar{h}(t', t) dG(t') \right. \\ &\left. + \int_t^1 \frac{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)}{(u-t) \vee \varepsilon} \bar{h}(t, u) dG(u) \right\} \end{aligned}$$

has a unique continuous solution  $\bar{\phi}_\varepsilon$ , satisfying

$$(2.12) \quad \inf_{x \in [0, M]} d_F(x)k(x) \leq \bar{\phi}_\varepsilon(t) \leq \sup_{x \in [0, M]} d_F(x)k(x),$$

for all  $t \in [0, 1]$  and  $\varepsilon > 0$ .

For points  $t$  in the range of  $F$ , say  $t = F(x)$ , we have  $\bar{\phi}_\varepsilon(t) = \phi_\varepsilon(x)$ .

(ii) *The integral equation*

$$(2.13) \quad \begin{aligned} \bar{\xi}_\varepsilon(t) = \bar{c}_F(t) &\left\{ \bar{k}(t) - \int_0^t \frac{\bar{\xi}_\varepsilon(t) - \bar{\xi}_\varepsilon(t')}{(t-t') \vee \varepsilon} \bar{h}^*(t', t) dG(t') \right. \\ &\left. + \int_t^1 \frac{\bar{\xi}_\varepsilon(u) - \bar{\xi}_\varepsilon(t)}{(u-t) \vee \varepsilon} \bar{h}^*(t, u) dG(u) \right\} \end{aligned}$$

has a unique continuous solution  $\bar{\xi}_\varepsilon$ , satisfying

$$(2.14) \quad \inf_{x \in [0, M]} c_F(x)k(x) \leq \bar{\xi}_\varepsilon(t) \leq \sup_{x \in [0, M]} c_F(x)k(x),$$

for all  $t \in [0, 1]$  and  $\varepsilon > 0$ .

For points  $t$  in the range of  $F$ , say  $t = F(x)$ , we have  $\bar{\xi}_\varepsilon(t) = \xi_\varepsilon(x)$ .

PROOF. (i) By the Fredholm theory, as used, for example, in Geskus and Groeneboom [(1996), Theorem 5, page 82], the  $\bar{\phi}_\varepsilon$ -equation (2.11) can be shown to have a unique continuous solution, for each  $\varepsilon > 0$ . Note that the integration in (2.11) is only with respect to  $dG(t')$  and  $dG(u)$  and therefore only involves values belonging to the range of  $F$ . So for points  $t$  in the range of  $F$  we have

$$\bar{\phi}_\varepsilon(t) = \phi_\varepsilon(G(t)).$$

Let  $m = \operatorname{arg\,min}[\bar{\phi}_\varepsilon]$  and  $s = \operatorname{arg\,max}[\bar{\phi}_\varepsilon]$ . We have

$$\bar{\phi}_\varepsilon(s) \leq \bar{d}_F(s)\bar{k}(s) \leq \sup_{x \in [0, M]} d_F(x)k(x),$$

since  $\bar{\phi}_\varepsilon(s) - \bar{\phi}_\varepsilon(t) \geq 0$ ,  $t \in [0, 1]$ , and similarly

$$\bar{\phi}_\varepsilon(m) \geq \bar{d}_F(m)\bar{k}(m) \geq \inf_{x \in [0, M]} d_F(x)k(x),$$

since  $\bar{\phi}_\varepsilon(m) - \bar{\phi}_\varepsilon(t) \leq 0$ ,  $t \in [0, 1]$ . Hence we have (2.12).

(ii) The argument is completely similar to the argument given for (i).  $\square$

The following lemma is the crux of the proof of the existence of the solution to the original integral equation.

LEMMA 2.2. *The functions  $\bar{\phi}_\varepsilon$  are Lipschitz on  $[0, 1]$ , uniformly in  $\varepsilon > 0$ .*

PROOF. We will use similar notation to that in Lemma 2.1. Let  $x_1, \dots, x_m$  be the points of jump of  $F$  and let  $x_0 = 0$ ,  $x_{m+1} = M$ . Furthermore, let  $\tau_i = F(x_i)$ ,  $i = 0, \dots, m + 1$ . For  $i = 0, \dots, m$ , the interval  $[\tau_i, \tau_{i+1}]$  can be divided into two parts.

(i) The interval  $[\tau_i, \tau'_i]$ , where  $\tau'_i = F(x_{i+1} -)$ . The interval  $[\tau_i, \tau'_i]$  corresponds to the interval  $[x_i, x_{i+1})$  in the original scale. The function  $G$  is strictly increasing and differentiable on the interval  $(\tau_i, \tau'_i)$ , and is right and left differentiable at  $\tau_i$  and  $\tau'_i$ , respectively.

(ii) The interval  $[\tau'_i, \tau_{i+1}]$ . This interval corresponds to the jump of  $F$  at  $x_{i+1}$ . Here the function  $G$  is constant, again having right and left derivatives at the respective endpoints.

If  $i = m$ , the second interval only consists of the point 1. Let

$$D' = \{\tau_0, \dots, \tau_{m+1}\} \cup \{\tau'_0, \dots, \tau'_m\} \\ \cup \left\{ \text{discontinuity points of } \bar{k}'(t), \bar{d}'_F(t), \right. \\ \left. \Delta_{t'}^1(t) = \frac{\partial}{\partial t} \bar{h}(t, t') \text{ for } t \leq t', \text{ and } \Delta_u^2(t) = \frac{\partial}{\partial t} \bar{h}(u, t) \text{ for } t \geq u \right\}.$$

Then  $\bar{\phi}_\varepsilon(t)$  is differentiable for  $t \notin D'$ , and has left and right derivatives for  $t \in D'$ . Using

$$\begin{aligned} \bar{k}(t) &= \int_0^t \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')}{(t - t') \vee \varepsilon} \bar{h}(t', t) dG(t') + \int_t^1 \frac{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)}{(u - t) \vee \varepsilon} \bar{h}(t, u) dG(u) \\ &= \frac{\bar{\phi}_\varepsilon(t)}{\bar{d}_F(t)} = \bar{\xi}_\varepsilon(t) [(1 - t)\bar{h}_1(t) + t\bar{h}_2(t)], \end{aligned}$$

and using left or right derivatives when  $t \in D'$ , we have

$$\begin{aligned} \bar{\phi}'_\varepsilon(t) &= \bar{d}'_F(t) \bar{\xi}_\varepsilon(t) [(1 - t)\bar{h}_1(t) + t\bar{h}_2(t)] \\ &\quad + \bar{d}_F(t) \left\{ \bar{k}'(t) - \int_0^t \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')}{(t - t') \vee \varepsilon} \frac{\partial}{\partial t} \bar{h}(t', t) dG(t') \right. \\ &\quad \left. + \int_t^1 \frac{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)}{(u - t) \vee \varepsilon} \frac{\partial}{\partial t} \bar{h}(t, u) dG(u) \right\} \\ (2.15) \quad &- \bar{d}_F(t) \left\{ \int_{t': t-t' > \varepsilon} \left\{ \frac{\bar{\phi}'_\varepsilon(t)}{t - t'} - \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')}{(t - t')^2} \right\} d\bar{H}(t', t) \right. \\ &\quad \left. + \int_{u, u-t > \varepsilon} \left\{ \frac{\bar{\phi}'_\varepsilon(t)}{u - t} - \frac{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)}{(u - t)^2} \right\} d\bar{H}(t, u) \right\} \\ &- \bar{d}_F(t) \bar{\phi}'_\varepsilon(t) \varepsilon^{-1} \left\{ \int_{t-\varepsilon}^t \bar{h}(t', t) g(t') dt' + \int_t^{t+\varepsilon} \bar{h}(t, u) g(u) du \right\}. \end{aligned}$$

Note that  $(\partial/\partial t)\bar{H}(t, u) = \bar{h}(t, u)g(t)$  and similarly for the other partial derivative of  $\bar{H}$ . Moving the terms containing  $\bar{\phi}'_\varepsilon$  to the left-hand side of (2.15) shows that  $\bar{\phi}'_\varepsilon(t)$  has a finite upper bound, using Lemma 2.1. Moreover,  $\bar{\phi}'_\varepsilon$  is piecewise continuous on the closed intervals from one point in  $D'$  to the subsequent one. So  $\bar{\phi}'_\varepsilon$  attains a maximum value, which may be a right or left derivative. The rest of the proof is devoted to showing that this maximum value is uniform in  $\varepsilon$ .

Let  $M_\varepsilon = \sup_{t \in [0, 1]} \bar{\phi}'_\varepsilon(t)$  and suppose that  $\bar{\phi}'_\varepsilon$  attains its supremum at a point  $s$ . Note that  $M_\varepsilon \geq 0$ , since  $\bar{\phi}_\varepsilon(0) = \bar{\phi}_\varepsilon(1) = 0$  and  $\bar{\phi}_\varepsilon$  is continuous. Then, if  $0 < t < s - \varepsilon$ ,

$$\frac{\bar{\phi}'_\varepsilon(s)}{s - t} - \frac{\bar{\phi}_\varepsilon(s) - \bar{\phi}_\varepsilon(t)}{(s - t)^2} \geq \frac{\int_t^s \{\bar{\phi}'_\varepsilon(s) - \bar{\phi}'_\varepsilon(u)\} du}{(s - t)^2} \geq 0.$$

Likewise, if  $1 > t > s + \varepsilon$ , we get

$$\frac{\bar{\phi}'_\varepsilon(s)}{t - s} - \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(s)}{(t - s)^2} \geq 0.$$

So these parts work in the opposite direction, and are harmless in (2.15). Now let  $K_\varepsilon(t)$  be defined by

$$K_\varepsilon(t) = \bar{d}_F(t)\bar{k}'(t) + \bar{d}'_F(t)\bar{\xi}_\varepsilon(t)[(1-t)\bar{h}_1(t) + t\bar{h}_2(t)] - \bar{d}_F(t) \left\{ \int_0^t \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')}{(t-t') \vee \varepsilon} \frac{\partial}{\partial t} \bar{h}(t',t) dG(t') - \int_t^1 \frac{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)}{(u-t) \vee \varepsilon} \frac{\partial}{\partial t} \bar{h}(t,u) dG(u) \right\}$$

and let  $C_\varepsilon(t)$  be defined by

$$(2.16) \quad C_\varepsilon(t) = 1 + \bar{d}_F(t)\varepsilon^{-1} \left\{ \int_{t-\varepsilon}^t \bar{h}(t',t)g(t') dt' + \int_t^{t+\varepsilon} \bar{h}(t,u)g(u) du \right\}, \quad t \in [0, 1].$$

Then we have

$$(2.17) \quad \bar{\phi}'_\varepsilon(s)C_\varepsilon(s) \leq K_\varepsilon(s),$$

implying

$$(2.18) \quad M_\varepsilon \leq \sup_{t \in [0, 1]} K_\varepsilon(t)/C_\varepsilon(t).$$

In a similar way, if  $m_\varepsilon = \inf_{t \in [0, 1]} \bar{\phi}'_\varepsilon(t)$ , we get

$$(2.19) \quad m_\varepsilon \geq \inf_{t \in [0, 1]} K_\varepsilon(t)/C_\varepsilon(t).$$

Let the function  $A_\delta$  be defined by

$$A_\delta(t) = \bar{d}_F(t) \left\{ \int_{t-\delta}^t \left| \frac{\partial}{\partial t} \bar{h}(t',t) \right| dG(t) + \int_t^{t+\delta} \left| \frac{\partial}{\partial t} \bar{h}(t,u) \right| dG(u) \right\}, \quad t \in [0, 1].$$

Fix  $\delta > 0$  such that, for all  $t \in [0, 1]$ ,

$$(2.20) \quad A_\delta(t)/C_\varepsilon(t) \leq \frac{1}{2}.$$

Note that  $\delta > 0$  can be chosen independently of  $\varepsilon > 0$ , since

$$\lim_{\varepsilon \downarrow 0} C_\varepsilon(t) = 1 + 2\bar{d}_F(t)\bar{h}(t,t)g(t), \quad t \in (0, 1).$$

Then we get from (2.20), for each  $t \in [0, 1]$ , by applying the mean value theorem on the ratios  $\{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')\}/(t-t')$  and  $\{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)\}/(u-t)$ ,

$$\begin{aligned} & \bar{d}_F(t) \left\{ \int_{t-\delta}^t \left| \frac{\bar{\phi}_\varepsilon(t) - \bar{\phi}_\varepsilon(t')}{(t-t') \vee \varepsilon} \frac{\partial}{\partial t} \bar{h}(t',t) \right| dG(t') + \int_t^{t+\delta} \left| \frac{\bar{\phi}_\varepsilon(u) - \bar{\phi}_\varepsilon(t)}{(u-t) \vee \varepsilon} \frac{\partial}{\partial t} \bar{h}(t,u) \right| dG(u) \right\} / C_\varepsilon(t) \\ & \leq A_\delta(t) \max\{M_\varepsilon, |m_\varepsilon|\} / C_\varepsilon(t) \leq \frac{1}{2} \max\{M_\varepsilon, |m_\varepsilon|\}. \end{aligned}$$

Defining  $B_\delta(t)$  by

$$B_\delta(t) = \bar{d}_F(t)|\bar{k}'(t)| + |\bar{d}'_F(t)| \left[ (1-t)\bar{h}_1(t) + t\bar{h}_2(t) \right] \sup_{t' \in [0,1]} \{ \bar{c}_F(t')|\bar{k}(t') \} \\ + \frac{2\bar{d}_F(t)}{\delta} \sup_{t' \in [0,1]} \{ \bar{d}_F(t')|\bar{k}(t') \} \\ \times \left\{ \sup_{t' \in [0,t]} \left| \frac{\partial}{\partial t} \bar{h}(t',t) \right| + \sup_{u \in [t,1]} \left| \frac{\partial}{\partial t} \bar{h}(t,u) \right| \right\},$$

we get, for  $t \in [0, 1]$ ,

$$\begin{aligned} & \bar{d}_F(t)|\bar{k}'(t)| + |\bar{d}'_F(t)|\bar{\xi}_\varepsilon(t) \left[ (1-t)\bar{h}_1(t) + t\bar{h}_2(t) \right] \\ & + \bar{d}_F(t) \left\{ \int_0^{t-\delta} \frac{|\bar{\phi}_\varepsilon(t)| + |\bar{\phi}_\varepsilon(t')|}{t-t'} \left| \frac{\partial}{\partial t} \bar{h}(t',t) \right| dG(t') \right. \\ & \quad \left. + \int_{t+\delta}^1 \frac{|\bar{\phi}_\varepsilon(t)| + |\bar{\phi}_\varepsilon(u)|}{u-t} \left| \frac{\partial}{\partial t} \bar{h}(t,u) \right| dG(u) \right\} \\ (2.21) \quad & \leq \bar{d}_F(t)|\bar{k}'(t)| + |\bar{d}'_F(t)|\bar{\xi}_\varepsilon(t) \left[ (1-t)\bar{h}_1(t) + t\bar{h}_2(t) \right] \\ & + \frac{2\bar{d}_F(t)}{\delta} \sup_{t' \in [0,1]} |\bar{\phi}_\varepsilon(t')| \\ & \quad \times \left\{ \int_0^{t-\delta} \left| \frac{\partial}{\partial t} \bar{h}(t',t) \right| dG(t') + \int_{t+\delta}^1 \left| \frac{\partial}{\partial t} \bar{h}(t,u) \right| dG(u) \right\} \\ & \leq B_\delta(t) \leq c, \end{aligned}$$

for some constant  $c$ , independent of  $\varepsilon$  and  $t$ . Hence, for each  $t \in [0, 1]$ ,

$$|\bar{\phi}'_\varepsilon(t)| \leq A_\delta(t)/C_\varepsilon(t) + B_\delta(t)/C_\varepsilon(t) \leq \frac{1}{2} \max\{M_\varepsilon, |m_\varepsilon|\} + B_\delta(t)/C_\varepsilon(t),$$

implying

$$(2.22) \quad \frac{1}{2} \max\{M_\varepsilon, |m_\varepsilon|\} \leq \sup_{t \in [0,1]} B_\delta(t)/C_\varepsilon(t) \leq \sup_{t \in [0,1]} c/C_\varepsilon(t) \leq c',$$

for some constant  $c'$  independent of  $\varepsilon$ .

Hence  $\bar{\phi}'_\varepsilon(t)$  is bounded on  $[0, 1]$ , uniformly in  $\varepsilon$  and  $t$ , implying that  $\bar{\phi}_\varepsilon$  is Lipschitz, uniformly in  $\varepsilon > 0$ .  $\square$

We now have the following theorem.

**THEOREM 2.2.** *Let  $G(t) = F^{-1}(t)$ ,  $t \in [0, 1]$ , with a derivative  $g$  which exists except for at most a finite number of points, where  $G$  has left and right derivatives. Furthermore, let  $\bar{k}(t) = k(G(t))$ ,  $\bar{H}(t, u) = H(G(t), G(u))$ ,  $\bar{h}(t, u) = h(G(t), G(u))$ , and let  $\bar{d}_F$  be defined by*

$$(2.23) \quad \bar{d}_F(t) = \frac{t(1-t)}{(1-t)\bar{h}_1(t) + t\bar{h}_2(t)},$$

where  $\bar{h}_i = h_i \circ G$ ,  $i = 1, 2$ . Then

(i) *The integral equation*

$$(2.24) \quad \bar{\phi}(t) = \bar{d}_F(t) \left\{ \bar{k}(t) - \int_0^t \frac{\bar{\phi}(t) - \bar{\phi}(t')}{t - t'} d\bar{H}(t', t) + \int_t^1 \frac{\bar{\phi}(u) - \bar{\phi}(t)}{u - t} d\bar{H}(t, u) \right\}, \quad t \in [0, 1],$$

has a unique solution which is Lipschitz on  $[0, 1]$ .

(ii) *The Lipschitz norm in (i) has the following upper bound. Let  $C(t)$  be defined by*

$$(2.25) \quad C(t) = 1 + 2\bar{d}_F(t)g(t)\bar{h}(t, t).$$

Moreover, let  $A_\delta(t)$  and  $B_\delta(t)$  be defined by

$$(2.26) \quad A_\delta(t) = \bar{d}_F(t) \left\{ \int_{t-\delta}^t \left| \frac{\partial}{\partial t} \bar{h}(t', t) \right| dG(t') + \int_t^{t+\delta} \left| \frac{\partial}{\partial t} \bar{h}(t, u) \right| dG(u) \right\},$$

and

$$(2.27) \quad B_\delta(t) = \bar{d}_F(t) |\bar{k}'(t)| + |\bar{d}_F(t)| [(1-t)\bar{h}_1(t) + t\bar{h}_2(t)] \sup_{t' \in [0, 1]} \{ \bar{c}_F(t') |\bar{k}(t')| \} + \frac{2\bar{d}_F(t)}{\delta} \sup_{t' \in [0, 1]} \{ \bar{d}_F(t') |\bar{k}(t')| \} \times \left\{ \sup_{t' \in [0, 1]} \left| \frac{\partial}{\partial t} \bar{h}(t', t) \right| + \sup_{u \in [t, 1]} \left| \frac{\partial}{\partial t} \bar{h}(t, u) \right| \right\}$$

At the points in

$D' = \{ \text{discontinuity points of } g(t), \text{ augmented with } 0 \text{ and } 1 \}$

$\cup \left\{ \text{discontinuity points of } \bar{k}'(t), \bar{d}_F'(t), \right.$

$\Delta_{t'}^1(t) = \frac{\partial}{\partial t} \bar{h}(t, t')$  for  $t \leq t'$ , and  $\Delta_u^2(t) = \frac{\partial}{\partial t} \bar{h}(u, t)$  for  $t \geq u \left. \right\}$ ,

$A_\delta$  and  $B_\delta$  have two versions, one corresponding to taking left derivatives and one corresponding to taking right derivatives.

Then there exists a  $\delta > 0$  such that

$$\sup_{t \in [0, 1]} A_\delta(t)/C(t) \leq 1/2$$

and we have

$$(2.28) \quad |\bar{\phi}(u) - \bar{\phi}(t)| \leq c(u - t), \quad 0 \leq t < u \leq 1,$$

where  $c$  is given by

$$(2.29) \quad c = 2 \sup_{t \in [0, 1]} B_\delta(t)/C(t).$$

(iii) *The integral equation (2.5) has a unique solution  $\phi$ .*

PROOF. (i) By the preceding two lemmas, the set  $\{\bar{\phi}_\varepsilon: \varepsilon \leq \varepsilon_0\}$  (for some  $\varepsilon_0 > 0$ ) is bounded and equicontinuous. Hence, by the Arzelà–Ascoli theorem, each sequence  $\bar{\phi}_{\varepsilon_n}$ ,  $\varepsilon_n \downarrow 0$ , has a subsequence  $(\bar{\phi}_{\varepsilon_m})$ , converging in the supremum metric to a continuous function  $\bar{\phi}$  on  $[0, 1]$ . By Lebesgue’s dominated convergence theorem we get, for such a subsequence  $(\bar{\phi}_{\varepsilon_m})$ ,

$$(2.30) \quad \begin{aligned} \bar{\phi}(x) &= \lim_{m \rightarrow \infty} \bar{\phi}_{\varepsilon_m}(x) \\ &= \bar{d}_F(x) \left\{ \bar{k}(x) - \int_0^x \frac{\bar{\phi}(x) - \bar{\phi}(t)}{x - t} \bar{h}(t, x) dG(t) \right. \\ &\quad \left. + \int_x^1 \frac{\bar{\phi}(t) - \bar{\phi}(x)}{t - x} \bar{h}(x, t) dG(t) \right\}. \end{aligned}$$

Uniqueness of the solution follows in the same way as in Lemma 2.1.

(ii) It was shown in (2.22) in the proof of Lemma 2.2 that

$$\sup_{t \in [0, 1]} |\phi'_\varepsilon(t)| \leq 2 \sup_{t \in [0, 1]} B_\delta(t)/C_\varepsilon(t),$$

where  $C_\varepsilon$  is defined by (2.16). But since

$$\lim_{\varepsilon \downarrow 0} C_\varepsilon(t) = 1 + 2\bar{d}_F(t)\bar{h}(t, t)g(t),$$

for  $t \in [0, 1]$ , (2.28) now follows.

(iii) We define  $\phi$  by  $\phi(x) = \bar{\phi}(F(x))$ . If  $t = F(x)$ , we get, by a change of variables,

$$\begin{aligned} \phi(x) &= \bar{\phi}(t) \\ &= \bar{d}_F(t) \left\{ \bar{k}(t) - \int_0^t \frac{\bar{\phi}(t) - \bar{\phi}(t')}{t - t'} d\bar{H}(t', t) \right. \\ &\quad \left. + \int_t^1 \frac{\bar{\phi}(u) - \bar{\phi}(t)}{u - t} d\bar{H}(t, u) \right\} \\ &= d_F(x) \left\{ k(x) - \int_0^x \frac{\phi(x) - \phi(x')}{F(x) - F(x')} dH(x', x) \right. \\ &\quad \left. + \int_x^M \frac{\phi(y) - \phi(x)}{F(y) - F(x)} dH(x, y) \right\}, \end{aligned}$$



and hence  $\phi$  satisfies the original integral equation. Uniqueness of  $\phi$  follows from uniqueness of  $\bar{\phi}$  (since a solution  $\phi$  conversely defines a solution  $\bar{\phi}$  on the inverse scale).  $\square$

REMARK. The same arguments can be applied to prove existence of a solution to the  $\xi$ -equation. Hence  $\phi$  can be written as

$$\phi = F(1 - F)\xi.$$

Solvability of  $\tilde{\kappa}_F = L_1^* L_1 a$  can now immediately be seen.

COROLLARY 2.1. *The equation  $\tilde{\kappa}_F = L_1^* L_1 a$  is solvable.*

PROOF. By the Lipschitz property of  $\bar{\phi}$  we have, for any  $0 \leq x < y \leq M$ ,

$$\frac{|\phi(y) - \phi(x)|}{|F(y) - F(x)|} = \frac{|\bar{\phi}(F(y)) - \bar{\phi}(F(x))|}{|F(y) - F(x)|} \leq K,$$

for some constant  $K$ . Thus the Radon–Nikodym derivative  $d\phi/dF$  is a.e.-[ $F$ ] bounded by  $K$ .  $\square$

For the canonical gradient we get, if  $t < u$ ,

$$\begin{aligned} \tilde{\theta}_F(t, u, \delta, \gamma) &= [L_1 a](t, u, \delta, \gamma) \\ &= -\delta \frac{\phi(t)}{F(t)} - \gamma \frac{\phi(u) - \phi(t)}{F(u) - F(t)} \\ &\quad + (1 - \delta - \gamma) \frac{\phi(u)}{1 - F(u)}. \end{aligned} \tag{2.31}$$

**3. Asymptotic efficiency of the NPMLE.** In this section we will denote the unknown distribution function of the unobservable random variables  $X_i$  by  $F_0$ . As in Section 2, we will assume that  $F_0$  is continuous. Let  $\hat{F}_n$  be the NPMLE of  $F_0$ , based on the sample of observations  $(U_1, V_1, \Delta_1, \Gamma_1), \dots, (U_n, V_n, \Delta_n, \Gamma_n)$ . It is obtained by maximizing the likelihood

$$(3.1) \quad \prod_{i=1}^n F(U_i)^{\Delta_i} (F(V_i) - F(U_i))^{\Gamma_i} (1 - F(V_i))^{1 - \Delta_i - \Gamma_i} h(U_i, V_i)$$

over the class of piecewise constant right-continuous (sub-)distribution functions on  $[0, M]$ , having jumps only at a subset of the points  $U_i$  and  $V_i$ ,  $i = 1, \dots, n$ . The properties of the function thus obtained are discussed in Groeneboom and Wellner (1992) and GG.

A rather important property of the NPMLE is that it does not depend on  $\prod h(U_i, V_i)$ , so we do not have to perform any preliminary density estimation or bandwidth choice. The fact that we do not have to solve a bandwidth problem is one of the great advantages of the nonparametric maximum likelihood approach in the present problem.

By the restriction that  $\hat{F}_n$  only has mass at the observation times, also made in Groeneboom and Wellner (1992), we get a piecewise constant func-

tion. Let  $x_0 = 0, x_{m+1} = M$  and let  $x_1 < \dots < x_m$  be the points of jump of  $F$  in the interval  $(0, M)$ . Then  $\hat{F}_n$  satisfies the following properties.

PROPOSITION 3.1. *Any function  $\sigma$  that is constant on the same intervals  $J_i = [x_{i-1}, x_i)$  as  $\hat{F}_n$  satisfies*

$$\int_{t \in J_i} \sigma(t) \left\{ \frac{\delta}{\hat{F}_n(t)} - \frac{\gamma}{\hat{F}_n(u) - \hat{F}_n(t)} \right\} dQ_n(t, u, \delta, \gamma) + \int_{u \in J_i} \sigma(u) \left\{ \frac{\gamma}{\hat{F}_n(u) - \hat{F}_n(t)} - \frac{1 - \delta - \gamma}{1 - \hat{F}_n(u)} \right\} dQ_n(t, u, \delta, \gamma) = 0$$

for  $i = 2, \dots, m$ .

PROOF. See Groeneboom and Wellner [(1992), part II, Proposition 1.3] and Geskus and Groeneboom [(1997), Corollary 1, page 207].

PROPOSITION 3.2.

$$\text{Prob} \left\{ \lim_{n \rightarrow \infty} \|\hat{F}_n - F_0\|_\infty = 0 \right\} = 1.$$

For the proof, see Groeneboom and Wellner [(1992), part II, Sections 4.1 (case 1) and 4.3 (case 2)].

PROPOSITION 3.3.

$$\|\hat{F}_n - F_0\|_{H_i} = \mathcal{O}_p(n^{-1/3}(\log n)^{1/6}) \quad \text{as } n \rightarrow \infty, \text{ for } i = 1, 2.$$

PROOF. See Geskus and Groeneboom [(1997), Corollary 2, page 209] and van de Geer [(1996), Example 3.2].

The following result is needed in the proof of Lemma 3.1.

PROPOSITION 3.4.

$$\lim_{n \rightarrow \infty} \text{Pr} \{ \hat{F}_n \text{ is defective} \} = 0.$$

See Geskus and Groeneboom [(1997), Proposition 1, page 206]. Although the conditions on  $H$  are different there, the proof is the same, since the difference in conditions has no bearing on this particular property.

In addition to the smoothness conditions (D1) to (D3), given in Section 2, we assume

$$(D4) \quad h(t, t) = \lim_{u \downarrow t} h(t, u) \geq c > 0,$$

for all  $t \in (0, M)$  and some  $c > 0$ .

As in GG, our definition of the canonical gradient  $\tilde{\theta}$  will be extended to piecewise constant distribution functions  $F$  with finitely many discontinuities, based on the solution  $\phi_F$  of a discrete version of the integral equation

(2.5). (In order to stress dependence on  $F$ , we will write  $\phi_F$  instead of  $\phi$ .) However, since  $F(v) - F(u)$  no longer remains bounded away from zero on the region where  $H$  puts mass, we have to use an approach different from the one in GG. On one hand, the quotient,

$$\frac{\phi_F(v) - \phi_F(u)}{F(v) - F(u)},$$

for  $u$  and  $v$  in the same interval of constancy of  $F$ , can only be defined correctly if  $\phi_F$  is constant on the same interval. On the other hand,  $d_F$ ,  $h$  and  $\tilde{\kappa}'_{F_0}$  in general are not constant on these intervals, making a completely discrete version of the integral equation impossible. Therefore, instead of one function  $\phi_F$  we now need a pair of functions  $(\phi_F, \psi_F)$ , satisfying

$$(3.2) \quad \phi_F(x) = d_F(x) \left\{ k(x) - \int_0^x r_F(t, x) h(t, x) dt + \int_x^M r_F(x, t) h(x, t) dt \right\},$$

where  $r_F(t, u)$  is defined by

$$(3.3) \quad r_F(t, u) = \begin{cases} \frac{\phi_F(u) - \phi_F(t)}{F(u) - F(t)}, & \text{if } F(t) < F(u), \\ \frac{\psi_F(u) - \psi_F(t)}{F_0(u) - F_0(t)}, & \text{if } F(t) = F(u), t < u, \end{cases}$$

where  $\phi_F$  is constant on the same intervals as  $F$ .

Since  $\phi_F$  is constant, the only real integral part is the  $\psi_F$ -part; the remaining part of the integral can be written as a summation. The key to the proof of the existence of a solution pair  $(\phi_F, \psi_F)$  and also to the other proofs in this section are a representation of the equation for  $\phi_F$  on an inverse scale and the construction of a continuous extension of the equation for  $\phi_F$  on this inverse scale (similar techniques were used in Section 2). Using a similar notation to Section 2, we denote by  $G$  the inverse of  $F$ , where, for purely discrete distribution functions  $F$ , we take the right-continuous version of the inverse, defined by

$$G(t) = \inf\{x \in [0, M]: F(x) > t\}, \quad t \geq 0.$$

Furthermore, we define

$$\bar{k}_F = k \circ G, \quad \bar{h}_{1,F} = h_1 \circ G, \quad \bar{h}_{2,F} = h_2 \circ G \quad \text{and} \\ \bar{d}_F(t) = \frac{t(1-t)}{(1-t)\bar{h}_{1,F}(t) + t\bar{h}_{2,F}(t)}.$$

and likewise  $\bar{H}(t, u) = H(G(t), G(u))$ ,  $0 \leq t \leq u \leq 1$ .

For part (iii) of Theorem 3.1, we will also need the following notation:

$$(3.4) \quad \Delta_i(g) = \int_{x_i}^{x_{i+1}} g(t) dt$$

$$(3.5) \quad \Delta_{ij}(h) = \int_{u=x_i}^{x_{i+1}} \int_{v=x_j}^{x_{j+1}} h(u, v) dv du$$

$$(3.6) \quad \tilde{d}_i = \frac{z_i(1 - z_i)}{\Delta_i(h_1)(1 - z_i) + \Delta_i(h_2)z_i}.$$

The following theorem shows the existence of the solution pair. Moreover, it gives a uniform Lipschitz condition for the functions  $\bar{\phi}_F$  and  $\psi_F$ , which will be a crucial tool in showing the Donsker property for  $\tilde{\theta}_F$ .

**THEOREM 3.1.** *Let the following conditions on  $F_0, H$  and  $\tilde{\kappa}_{F_0}$  be satisfied: (M1) to (M3); (D1) to (D4); (F1) to (F3). Furthermore, let  $\mathcal{F}_{[0, M]}$  be the set of discrete nondefective distribution functions on  $[0, M]$  with finitely many points of jump, contained in  $(0, M)$ . Then there exists an  $\varepsilon > 0$  such that, for  $F \in \mathcal{F}$ , where  $\mathcal{F}$  is defined by  $\{F \in \mathcal{F}_{[0, M]} : \sup_{x \in [0, M]} |F(x) - F_0(x)| \leq \varepsilon\}$ ,*

(i) *There exists a unique Lipschitz function  $\bar{\phi}_F : [0, 1] \rightarrow \mathbb{R}$  such that, for  $t \in [0, 1] \setminus D$ ,*

$$(3.7) \quad \begin{aligned} \bar{\phi}_F(t) = \bar{d}_F(t) & \left\{ \bar{k}_F(t) - \int_{t' \in [0, t)} \frac{\bar{\phi}_F(t) - \bar{\phi}_F(t')}{t - t'} d\bar{H}(t', t) \right. \\ & \left. + \int_{u \in (t, 1]} \frac{\bar{\phi}_F(u) - \bar{\phi}_F(t)}{u - t} d\bar{H}(t, u) \right\}, \end{aligned}$$

where  $D$  is the (finite) set of discontinuities of the right-continuous inverse  $G = F^{-1}$  in  $(0, 1)$ , augmented with 0 and 1. The function  $\bar{\phi}_F$  is Lipschitz, uniformly for  $F \in \mathcal{F}$ .

(ii) *There exists a pair  $(\phi_F, \psi_F)$ , solving the integral equation (3.2), where  $\phi_F$  is absolutely continuous with respect to  $F$  and the function  $\psi_F$  is Lipschitz on each interval between jumps of  $F$ , uniformly for  $F \in \mathcal{F}$ , with a Lipschitz norm not depending on the interval.*

(iii) *Let  $z_i = F(x_i)$  and  $y_i = \phi_F(x_i)$ ,  $i = 1, \dots, m$ . Then, using the definitions (3.4) to (3.6), we have that the vector  $y = (y_1, \dots, y_m)'$  is the unique solution of the set of linear equations,*

$$(3.8) \quad \begin{aligned} y_i & \left\{ \tilde{d}_i^{-1} + \sum_{j < i} \frac{\Delta_{ji}(h)}{z_i - z_j} + \sum_{j > i} \frac{\Delta_{ij}(h)}{z_j - z_i} \right\} \\ & = \Delta_i(k) + \sum_{j < i} \frac{\Delta_{ji}(h)}{z_i - z_j} y_j + \sum_{j > i} \frac{\Delta_{ij}(h)}{z_j - z_i} y_j, \quad i = 1, \dots, m. \end{aligned}$$

Theorem 3.1 will be proved by approximating the purely discrete distribution function  $F$  by the function  $F_\alpha = (1 - \alpha)F_0 + \alpha F$  and by studying the behavior of the corresponding function  $\phi_{F_\alpha}$ , as  $\alpha \uparrow 1$ . The rather technical proof is given in the Appendix. By Theorem 3.1, the definition of the function  $\theta_F$  can be extended to piecewise constant distribution functions  $F \in \mathcal{F}$  by defining

$$(3.9) \quad \tilde{\theta}_F(t, u, \delta, \gamma) = -\frac{\delta\phi_F(t)}{F(t)} - \gamma r_F(t, u) + \frac{(1 - \delta - \gamma)\phi_F(u)}{1 - F(u)},$$

where  $\phi_F$  and  $\psi_F$  solve (3.2), and where  $\phi_F(t)/F(t)$  and  $\phi_F(u)/[1 - F(u)]$  are defined to be zero if  $F(t) = 0$  or if  $1 - F(u) = 0$ , respectively. Note that  $\theta_F$  no longer has an interpretation as canonical gradient.

In the sequel we will write  $Q_F$  instead of  $Q_{F,H}$ . We are now ready to formulate our main result.

**THEOREM 3.2.** *Let the conditions of Theorem 3.1 be satisfied. Then*

$$(3.10) \quad \sqrt{n} \left( K(\hat{F}_n) - K(F_0) \right) \rightarrow_{\mathcal{D}} N\left(0, \|\tilde{\theta}_{F_0}\|_{Q_{F_0}}^2\right) \quad \text{as } n \rightarrow \infty.$$

**PROOF.** Note that it is sufficient to show the following:

$$(3.11) \quad \sqrt{n} \left( K(\hat{F}_n) - K(F_0) \right) = \sqrt{n} \int \tilde{\theta}_{F_0} d(Q_n - Q_{F_0}) + o_p(1).$$

Moreover, using the uniform consistency of  $\hat{F}_n$  (see Proposition 3.2), we may assume that  $\hat{F}_n \in \mathcal{F}$ , for all large  $n$ , where  $\mathcal{F}$  is defined as in Theorem 3.1. The proof consists of the following steps.

(i) By conditions (D1) and (F2), and Proposition 3.3 we have

$$\sqrt{n} \left( K(\hat{F}_n) - K(F_0) \right) = \sqrt{n} \int \tilde{\kappa}_{F_0} d(\hat{F}_n - F_0) + o_p(1).$$

(ii) In Lemma 3.1 the following will be shown:

$$\int \tilde{\kappa}_{F_0} d(F - F_0) = - \int \tilde{\theta}_F dQ_{F_0},$$

if  $F \in \mathcal{F}$ .

(iii) Unlike the situation in GG,  $\phi_{\hat{F}_n}$  is constant on the same intervals as  $\hat{F}_n$ . Since  $\gamma = 0$ , if  $\hat{F}_n(u) = \hat{F}_n(t)$ , Proposition 3.1 can be used to obtain

$$\int \tilde{\theta}_{\hat{F}_n} dQ_n = 0,$$

yielding

$$-\sqrt{n} \int \tilde{\theta}_{\hat{F}_n} dQ_{F_0} = \sqrt{n} \int \tilde{\theta}_{\hat{F}_n} d(Q_n - Q_{F_0}).$$

(iv) This is further split into

$$\begin{aligned} \sqrt{n} \int \tilde{\theta}_{\hat{F}_n} d(Q_n - Q_{F_0}) &= \sqrt{n} \int \tilde{\theta}_{F_0} d(Q_n - Q_{F_0}) \\ &\quad + \sqrt{n} \int (\tilde{\theta}_{\hat{F}_n} - \tilde{\theta}_{F_0}) d(Q_n - Q_{F_0}). \end{aligned}$$

The last term will be shown to be  $o_p(1)$  (Theorem 3.3).  $\square$

LEMMA 3.1. *Let  $\mathcal{F}$  be defined as in Theorem 3.1. Then we have under the conditions of Theorem 3.1, for all  $F \in \mathcal{F}$ ,*

$$\int \tilde{\kappa}_{F_0} d(F - F_0) = - \int \tilde{\theta}_F dQ_{F_0}.$$

PROOF. Let, for any distribution function  $F$ ,  $L_F: L_2(F) \rightarrow L_2(Q_F)$  denote the conditional expectation operator

$$\begin{aligned} [L_F a](u, v, \delta, \gamma) &= \frac{\delta \int_0^u a dF}{F(u)} + \frac{\gamma \int_u^v a dF}{F(v) - F(u)} \\ &\quad + \frac{(1 - \delta - \gamma) \int_v^M a dF}{1 - F(v)} \quad \text{a.e.}-[Q_F], \end{aligned}$$

with adjoint  $L^*$ , given by the conditional expectation

$$[L^* b](x) = E[b(U, V, \Delta, \Gamma) | X = x] \quad \text{a.e.}-F.$$

Since the adjoint is an expectation, conditionally on the value of the random variable  $X \sim F$ , its structure does not depend on  $F$ ;  $F$  only determines where it has to be defined (the a.e.- $F$  part). Still  $a \in L_2^0(F)$  implies  $L_F(a) \in L_2^0(Q_F)$ . The ratios occurring in  $\tilde{\theta}_F$ ,  $F \in \mathcal{F}$ , are bounded, since, by Theorem 3.1,  $\bar{\phi}_F$  and  $\psi_F$  are Lipschitz functions, if  $F \in \mathcal{F}$ . Hence  $\tilde{\theta}_F \in L_2(Q_{F_0})$ , for  $F \in \mathcal{F}$ . Let  $1 \in L_2(F)$  denote the constant function  $1(x) \equiv 1$ ,  $x \in \mathbb{R}$ . Under  $L_F$  this transforms into the constant function  $1^\circ(t, u, \delta, \gamma) \equiv 1$  on  $L_2(Q_F)$ . Now we have

$$\begin{aligned} \int \tilde{\theta}_F dQ_{F_0} &= \langle \tilde{\theta}_F, 1^\circ \rangle_{Q_{F_0}} \\ &= \langle \tilde{\theta}_F, L_{F_0}(1) \rangle_{Q_{F_0}} \\ &= \langle L^*(\tilde{\theta}_F), 1 \rangle_{F_0} \\ &= \int L^*(\tilde{\theta}_F) dF_0. \end{aligned}$$

If we can prove

$$L^*(\tilde{\theta}_F) = \tilde{\kappa}_{F_0} - \int \tilde{\kappa}_{F_0} dF \quad \text{a.e.}-F_0,$$

we are done. This is shown as follows: recall that the integral equation was obtained by taking derivatives on both sides of the equation  $\tilde{\kappa}_{F_0}(x) = [L^*\tilde{\theta}_{F_0}](x)$  for all  $x \in [0, M]$ . Now we will go the other way, integrate, but replace  $\tilde{\theta}_{F_0}$  by  $\tilde{\theta}_F$ , obtaining

$$[L^*\tilde{\theta}_F](x) = [\tilde{\kappa}_{F_0}](x) + C \quad \text{for all } x \in [0, M].$$

For the constant  $C$  we have, using that  $F$  is nondefective,

$$\begin{aligned} C &= \int C dF \\ &= \int L^*(\tilde{\theta}_F) dF - \int \tilde{\kappa}_{F_0} dF. \end{aligned}$$

It is easily shown that  $\tilde{\theta}_F$  is contained in  $L^0_2(Q_F)$ . [However, it is *not* contained in  $\mathcal{R}(L_F)$ , because of the part  $\psi_F$  of the solution pair  $(\phi_F, \psi_F)$ .] Now we have

$$\begin{aligned} \langle L^*(\tilde{\theta}_F), 1 \rangle_F &= \langle \tilde{\theta}_F, L_F(1) \rangle_{Q_F} \\ &= \langle \tilde{\theta}_F, 1^\circ \rangle_{Q_F} \\ &= 0. \end{aligned} \quad \square$$

The hard part of the proof of Theorem 3.2 is to show that

$$(3.12) \quad \sqrt{n} \int (\tilde{\theta}_{\hat{F}_n} - \tilde{\theta}_{F_0}) d(Q_n - Q_{F_0}) = o_p(1).$$

We will prove the Donsker property for the “middle part”  $\gamma r_F$  of  $\tilde{\theta}_F$  and only indicate the very similar (simpler) proofs for the two other parts of  $\tilde{\theta}_F$  at the end of the proof of Theorem 3.3. Since the proof is rather involved, we first sketch the general ideas.

We start by defining a neighborhood, shrinking with  $n$ , such that the probability that  $\hat{F}_n$  belongs to  $\mathcal{F}_n$  tends to 1, as  $n \rightarrow \infty$ . Let, for  $F \in \mathcal{F}$ ,

$$(3.13) \quad \begin{aligned} q_F(t, u, \delta, \gamma) &= \delta F(t) + \gamma\{F(u) - F(t)\} \\ &\quad + (1 - \gamma - \delta)\{1 - F(u)\} \end{aligned}$$

and

$$(3.14) \quad \begin{aligned} q_{F_0}(t, u, \delta, \gamma) &= \delta F_0(t) + \gamma\{F_0(u) - F_0(t)\} \\ &\quad + (1 - \delta - \gamma)\{1 - F_0(u)\}. \end{aligned}$$

It is proved in van de Geer (1996), that

$$(3.15) \quad h^2(q_{\hat{F}_n}, q_{F_0}) = \mathcal{O}_p(n^{-2/3}(\log n)^{1/3}),$$

where  $h(q_F, q_{F_0})$  is the Hellinger distance between the densities  $q_F$  and  $q_{F_0}$  w.r.t. the product of the measure induced by  $H$  and counting measure on

$\{0, 1\}^2 \setminus \{(1, 1)\}$ . The result (3.15) has already been used above, since Proposition 3.3 is based on it. Now, if  $\mathcal{F}_n$  is the set of distribution functions  $F \in \mathcal{F}$ , satisfying

$$(3.16) \quad h^2(q_F, q_{F_0}) \leq n^{-2/3} \log n,$$

we have

$$\Pr\{\hat{F}_n \in \mathcal{F}_n\} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

In fact, the upper bound  $n^{-2/3} \log n$ , defining the class  $\mathcal{F}_n$ , could be replaced by

$$c_n n^{-2/3} (\log n)^{1/3},$$

where we only need  $c_n \rightarrow \infty$ , as  $n \rightarrow \infty$ , but we are a little bit wasteful with our powers of  $\log n$  in an attempt to avoid an accumulation of constants in the upper bounds.

We need to study properties of the empirical integrals

$$Q_n(\tilde{\theta}_F - \tilde{\theta}_{F_0})^2 \quad \text{and} \quad Q_n(\tilde{\theta}_G - \tilde{\theta}_G)^2,$$

for  $F, G \in \mathcal{F}_n$ . The denominators in  $\tilde{\theta}_{F_0}$ ,  $\tilde{\theta}_F$  and  $\tilde{\theta}_G$  can be arbitrarily close to zero. If  $F \in \mathcal{F}$ , then  $F(u) - F(t)$  will be zero on a region of positive Lebesgue measure, in which case we get for the “middle part”  $\gamma r_F$  of  $\tilde{\theta}_F$ :

$$\gamma r_F(t, u) = \gamma\{\psi_F(u) - \psi_F(t)\} / \{(F_0(u) - F_0(t))\}.$$

We will face these difficulties by considering three regions of integration:

$$(3.17) \quad C_{n,\eta}(F) = \{w: q_F(w) > \eta q_{F_0}(w), q_{F_0}(w) > n^{-1/3}\},$$

$$(3.18) \quad D_\eta(F) = \{w: q_F(w) \leq \eta q_{F_0}(w)\}$$

and

$$(3.19) \quad C_n(F_0) = \{w: q_{F_0}(w) \leq n^{-1/3}\},$$

for some  $\eta \in (0, 1)$ , where the elements  $w$  of the sets, defined above, are of the form  $w = (t, u, \delta, \gamma)$ . On the region  $C_{n,\eta}(F)$ ,  $\tilde{\theta}_F$  has a behavior which is comparable to the behavior of  $\tilde{\theta}_{F_0}$ ; on the other regions we just use the uniform boundedness of  $\tilde{\theta}_F$  and the fact that the integrals over these regions become sufficiently small.

For the entropy calculations we shall use ratios  $r_{F_k, G_k, \bar{\phi}_k}$  of the form

$$\frac{\bar{\phi}_k(G_k(u)) - \bar{\phi}_k(F_k(t))}{G_k(u) - F_k(t)},$$

where  $F_k$  and  $G_k$  are distribution functions such that  $F_k \leq F \leq G_k$  [ $(F_k, G_k)$  is a “bracket” for  $F$ ] and where  $\bar{\phi}_k$  is a Lipschitz function approximating  $\phi_F$ . In this way, the good behavior of the ratios  $r_F$  on the region  $C_{n,\eta}(F)$  is preserved on the same region by the approximating ratio  $r_{F_k, G_k, \bar{\phi}_k}$ . Next we



will apply the chaining lemma, using a chain with a kind of “funnel” structure, preserving this good behavior on the region  $C_{n,\eta}(F)$ . Note that the approximating ratios are outside the original class of ratios  $r_F$ .

In dealing with the region  $D_\eta(F)$ , defined by (3.18), we will need the functions

$$(3.20) \quad g_F = \frac{q_F - q_{F_0}}{q_F + q_{F_0}},$$

for which Lemma A.1 in van de Geer (1996) holds. This lemma, specialized to our situation, is given below for easy reference.

LEMMA 3.2. *Let, for  $F \in \mathcal{F}_n$ ,  $S_n(F)$  and  $\mathcal{E}_n$  be defined by*

$$S_n(F) = \left\{ \int_{q_{F_0} > \sigma_n} g_F^2 dQ_n \right\}^{1/2} \quad \text{and} \quad \mathcal{E}_n = \{g_F \mathbf{1}_{\{q_{F_0} > \sigma_n\}} : F \in \mathcal{F}_n\},$$

where (we take)  $\sigma_n = n^{-1/3}$ . Let  $(\rho_n)_{n \geq 1}$  be a nondecreasing sequence of real numbers greater than or equal to 1. Then, given  $0 < \nu < 2$  and  $0 < C < \infty$ , there exists an  $0 < L < \infty$  depending on  $\nu$  and  $C$ , such that for

$$\tau_n \geq n^{-1/(2+\nu)} \rho_n^{\nu/(2+\nu)} \quad \text{for all } n,$$

we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{S_n(F)}{h(q_F, q_{F_0}) \vee (L\tau_n)} \geq 8 \text{ for some } F \in \mathcal{F}_n \right\} \\ \leq \limsup_{n \rightarrow \infty} 4 \Pr \left\{ \sup_{\delta > 0} \left( \frac{\delta}{\rho_n} \right)^\nu H(\delta, \mathcal{E}_n, Q_n) > C \right\}, \end{aligned}$$

where  $H(\delta, \mathcal{E}_n, Q_n)$  denotes the  $\delta$ -entropy of  $\mathcal{E}_n$  for the  $L_2$ -distance w.r.t.  $Q_n$ .

We will also need the following two lemmas.

LEMMA 3.3. (i) *Let the function  $a_n$  be defined by*

$$a_n = \mathbf{1}_{\{q_{F_0} > n^{-1/3}\}} / q_{F_0}^2.$$

Then

$$(3.21) \quad Q_n a_n = \mathcal{O}_p(\log n)$$

(ii) *Let the function  $b_n$  be defined by*

$$b_n = \mathbf{1}_{\{q_{F_0} \leq n^{-1/3}\}}.$$

Then

$$(3.22) \quad Q_n b_n = \mathcal{O}_p(n^{-2/3}).$$

PROOF. Both statements are simple consequences of the Markov inequality,

$$\begin{aligned} \Pr\{Q_n a_n > c \log n\} &\leq (c \log n)^{-1} Q_{F_0} a_n \\ &= (c \log n)^{-1} \int_{q_{F_0} > n^{-1/3}} q_{F_0}^{-2} dQ_{F_0} = c^{-1} \mathcal{O}(1), \end{aligned}$$

and, likewise,

$$\Pr\{Q_n b_n > cn^{-2/3}\} \leq c^{-1} n^{2/3} Q_{F_0} b_n = c^{-1} n^{2/3} \int_{q_{F_0} \leq n^{-1/3}} dQ_{F_0} = c^{-1} \mathcal{O}(1). \quad \square$$

LEMMA 3.4. *Let, for  $\eta \in (0, 1)$ , the set  $D_\eta(F)$  be defined by (3.18). Then*

$$(3.23) \quad \sup_{F \in \mathcal{F}_n} Q_n D_\eta(F) = \mathcal{O}_p(n^{-2/3} \log n).$$

PROOF. We have

$$\begin{aligned} (3.24) \quad Q_n D_\eta(F) &\leq (1 - \eta)^{-2} \int_{D_\eta(F)} \left\{ \frac{q_F - q_{F_0}}{q_{F_0}} \right\}^2 dQ_n \\ &\leq 4(1 - \eta)^{-2} Q_n \left\{ \frac{q_F - q_{F_0}}{q_F + q_{F_0}} \right\}^2, \end{aligned}$$

where we use in the first step that

$$|q_F - q_{F_0}| = q_{F_0} \left| \frac{q_F}{q_{F_0}} - 1 \right| \geq (1 - \eta) q_{F_0},$$

if  $q_F \leq \eta q_{F_0}$ .

Furthermore, by Lemmas 3.2 and 3.3,

$$\begin{aligned} (3.25) \quad \sup_{F \in \mathcal{F}_n} \int_{q_{F_0} > n^{-1/3}} \left\{ \frac{q_F - q_{F_0}}{q_F + q_{F_0}} \right\}^2 dQ_n &= \sup_{F \in \mathcal{F}_n} \int_{q_{F_0} > n^{-1/3}} g_F^2 dQ_n \\ &= \mathcal{O}_p(n^{-2/3} \log n). \end{aligned}$$

Relation (3.25) follows from Lemma 3.2, by taking  $\nu = 1$ ,  $\tau_n = n^{-1/3}(\log n)^{1/2}$  and  $\rho_n = \log n$ . The metric entropy  $H(\delta, \mathcal{S}_n, \mathbf{Q}_n)$  of the class  $\mathcal{S}_n$ , with the  $L_2$ -distance with respect to  $\mathbf{Q}_n$ , is then  $\mathcal{O}_p(\delta^{-1}(\log n)^{1/2})$  uniformly in  $\delta > 0$ . This is seen by first noting that

$$\frac{q_F - q_{F_0}}{q_F + q_{F_0}} = \frac{2q_F}{q_F + q_{F_0}} - 1,$$

and next that, for two distribution functions  $F_1$  and  $F_2$ ,

$$\left| \frac{q_{F_1}}{q_{F_1} + q_{F_0}} - \frac{q_{F_2}}{q_{F_2} + q_{F_0}} \right| \leq \frac{|q_{F_1} - q_{F_2}|}{q_{F_0}}.$$

By the results of Birman and Solomjak (1967) or Ball and Pajor (1990), applied on classes of uniformly bounded monotone functions, the  $\delta$ -entropy of the class of functions  $q_F$  for the  $L_2$ -distance w.r.t. the probability measure  $\tilde{Q}_n$ , defined by

$$\tilde{Q}_n(B) = \int_{B \cap \{q_{F_0} > n^{-1/3}\}} q_{F_0}^{-2} dQ_n \Big/ \int_{q_{F_0} > n^{-1/3}} q_{F_0}^{-2} dQ_n,$$

is  $\mathcal{O}(\delta^{-1})$ , implying  $H(\delta, \mathcal{S}_n, \mathbf{Q}_n) = \mathcal{O}_p(\delta^{-1}(\log n)^{1/2})$ , using part (i) of Lemma 3.3. The entropy results needed here can also be found in van der Vaart and Wellner [(1996), Theorem 2.7.5, page 159; Theorem 2.6.9, page 142 and Example 2.6.21, page 149] where, in fact,  $\log N(\varepsilon, \mathcal{F}, L_2(\mathbf{Q})) \leq K/\varepsilon$  follows from Theorem 2.6.9 by taking  $\mathcal{F}$  in the latter to be the class of indicators  $\{1_{[0,t]}; t \in \mathbb{R}\}$ , with  $V = 2$ .

Since, by  $|g_F| \leq 1$  and part (ii) of Lemma 3.3,

$$\sup_{F \in \mathcal{S}_n} \int_{q_{F_0} \leq n^{-1/3}} \left\{ \frac{q_F - q_{F_0}}{q_F + q_{F_0}} \right\}^2 dQ_n \leq \int_{q_{F_0} \leq n^{-1/3}} dQ_n \leq \mathcal{O}_p(n^{-2/3}),$$

we get, in combination with (3.25),

$$(3.26) \quad \sup_{F \in \mathcal{S}_n} \int \left\{ \frac{q_F - q_{F_0}}{q_F + q_{F_0}} \right\}^2 dQ_n = \mathcal{O}_p(n^{-2/3} \log n).$$

The result now follows from (3.24) and (3.26).  $\square$

From now on we will concentrate on the behavior of the “middle part” of  $\tilde{\theta}_F$ . Consider triples  $(F_k, G_k, \bar{\phi}_k)$ , where  $F_k$  and  $G_k$  are distribution functions belonging to  $\mathcal{F}$  and  $\bar{\phi}_k$  belongs to a uniform class of Lipschitz functions on  $[0, 1]$ , with the same uniform Lipschitz norm  $c_{\text{Lip}}$  and upper bound as the functions  $\bar{\phi}_F$ ,  $F \in \mathcal{F}$ , of Theorem 3.1. For these triples we define

$$(3.27) \quad r_{F_k, G_k, \bar{\phi}_k}(t, u) = \frac{\bar{\phi}_k(G_k(u)) - \bar{\phi}_k(F_k(t))}{G_k(u) - F_k(t)} \quad \text{if } G_k(u) > F_k(t),$$

and, for pairs  $(t, u)$  such that  $F_k(t) = G_k(u)$ , we define  $r_{F_k, G_k, \bar{\phi}_k}(t, u) = 0$ . Moreover, we define the semimetric

$$\begin{aligned}
 & d_n((F_k, G_k, \bar{\phi}_k), (F_l, G_l, \bar{\phi}_l)) \\
 &= 2 \max_{t \in [0, 1]} |\bar{\phi}_k(t) - \bar{\phi}_l(t)| \\
 &\quad \times \left\{ \int_{F_0(u) - F_0(t) > n^{-1/3}} \frac{1}{\{F_0(u) - F_0(t)\}^2} \gamma dQ_n \right\}^{1/2} \\
 (3.28) \quad &+ \left\{ \int_{F_0(u) - F_0(t) > n^{-1/3}} \left\{ \frac{F_k(t) - F_l(t)}{F_0(u) - F_0(t)} \right\}^2 \gamma dQ_n \right\}^{1/2} \\
 &+ \left\{ \int_{F_0(u) - F_0(t) > n^{-1/3}} \left\{ \frac{G_k(u) - G_l(u)}{F_0(u) - F_0(t)} \right\}^2 \gamma dQ_n \right\}^{1/2}.
 \end{aligned}$$

We now have the following result.

LEMMA 3.5. *Let, for distribution functions  $F$  and  $G$  such that  $F \leq G$ , the set  $C_{n, \eta}(F, G)$  be defined by*

$$\begin{aligned}
 (3.29) \quad C_{n, \eta}(F, G) &= \{(t, u) : F_0(u) - F_0(t) > n^{-1/3}, \\
 &\quad G(u) - F(t) \geq \eta\{F_0(u) - F_0(t)\}\}.
 \end{aligned}$$

Then we have for all pairs of distribution functions  $(F_k, G_k)$  and  $(F_l, G_l)$  such that  $F_k \leq G_k$  and  $F_l \leq G_l$ ,

$$\begin{aligned}
 & Q_n(r_{F_k, G_k, \bar{\phi}_k} - r_{F_l, G_l, \bar{\phi}_l})^2 \gamma 1_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} \\
 & \leq C^2 d_n((F_k, G_k, \bar{\phi}_k), (F_l, G_l, \bar{\phi}_l))^2,
 \end{aligned}$$

where  $r_{F, G, \bar{\phi}}$  is defined by (3.27), and where  $C > 0$  is a constant, only depending on  $\eta \in (0, 1)$  and the Lipschitz norm  $c_{\text{Lip}}$ , corresponding to the uniform Lipschitz class of functions  $\bar{\phi}_F, F \in \mathcal{F}$ .

PROOF. If  $G_k(u) - F_k(t) > 0$  and  $G_l(u) - F_l(t) > 0$ , we have the decomposition

$$\begin{aligned}
 & \{r_{F_k, G_k, \bar{\phi}_k}(t, u) - r_{F_l, G_l, \bar{\phi}_l}(t, u)\} \gamma \\
 &= \left\{ \frac{\bar{\phi}_k(G_k(u)) - \bar{\phi}_k(F_k(t))}{G_k(u) - F_k(t)} - \frac{\bar{\phi}_l(G_l(u)) - \bar{\phi}_l(G_l(t))}{G_l(u) - F_l(t)} \right\} \gamma \\
 &= \frac{\bar{\phi}_k(G_k(u)) - \bar{\phi}_k(F_k(t))}{G_k(u) - F_k(t)} \{G_l(u) - F_l(t) - [G_k(u) - F_k(t)]\}
 \end{aligned}$$

$$\begin{aligned} & \times \frac{\gamma}{G_l(u) - F_l(t)} \\ & + \left\{ \bar{\phi}_k(G_k(u)) - \bar{\phi}_k(F_k(t)) - [\bar{\phi}_l(G_l(u)) - \bar{\phi}_l(F_l(t))] \right\} \\ & \times \frac{\gamma}{G_l(u) - F_l(t)}. \end{aligned}$$

Hence

$$\begin{aligned} & Q_n(r_{F_k, G_k, \bar{\phi}_k} - r_{F_l, G_l, \bar{\phi}_l})^2 \gamma \mathbf{1}_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} \\ & \leq c_{\text{Lip}}^2 \eta^{-2} \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \times \{G_l(u) - F_l(t) - [G_k(u) - F_k(t)]\}^2 \gamma dQ_n \\ & \quad + \eta^{-2} \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \times \left\{ \bar{\phi}_k(G_k(u)) - \bar{\phi}_k(F_k(t)) - [\bar{\phi}_l(G_l(u)) - \bar{\phi}_l(F_l(t))] \right\}^2 \gamma dQ_n \\ & \leq c_{\text{Lip}}^2 \eta^{-2} \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \times \{G_l(u) - F_l(t) - [G_k(u) - F_k(t)]\}^2 \gamma dQ_n \\ (3.30) \quad & + 2\eta^{-2} \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \times \left\{ \bar{\phi}_k(G_l(u)) - \bar{\phi}_k(F_l(t)) - [\bar{\phi}_l(G_l(u)) - \bar{\phi}_l(F_l(t))] \right\}^2 \gamma dQ_n \\ & \quad + 4c_{\text{Lip}}^2 \eta^{-2} \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \quad \times \left[ (G_k(u) - G_l(u))^2 + (F_k(t) - F_l(t))^2 \right] \gamma dQ_n \\ & \leq C^2 \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \quad \times \left[ (G_k(u) - G_l(u))^2 + (F_k(t) - F_l(t))^2 \right] \gamma dQ_n \\ & \quad + C^2 \int_{C_{n, \eta}(F_k, G_k) \cap C_{n, \eta}(F_l, G_l)} (F_0(u) - F_0(t))^{-2} \\ & \quad \quad \times \left[ (\bar{\phi}_k(G_l(u)) - \bar{\phi}_l(G_l(u)))^2 + (\bar{\phi}_k(F_l(t)) - \bar{\phi}_l(F_l(t)))^2 \right] \gamma dQ_n \\ & \leq C^2 d_n((F_k, G_k, \bar{\phi}_k), (F_l, G_l, \bar{\phi}_l))^2, \end{aligned}$$

where  $C > 0$  is a constant, only depending on  $\eta$  and the Lipschitz norm  $c_{\text{Lip}}$ , corresponding to the uniform Lipschitz class of functions  $\bar{\phi}_F, F \in \mathcal{F}$ .  $\square$

Note that we can also apply Lemma 3.5 in the approximation of a ratio  $r_F$ , by making the identification  $r_F = r_{F, F, \bar{\phi}_F}$  and by noting that  $C_{n, \eta}(F, F) = C_{n, \eta}(F)$ . For the set of functions  $\mathcal{F}_n$  we now have the following theorem, which finishes the proof of Theorem 3.2.

**THEOREM 3.3.** *Let  $\mathcal{F}_n$  be the set of distribution functions  $F \in \mathcal{F}$ , defined by (3.16). Then we have, under the conditions of Theorem 3.2, for each  $\varepsilon > 0$ ,*

$$(3.31) \quad \Pr\left\{ \sup_{F \in \mathcal{F}_n} \left| \sqrt{n} (\mathbf{Q}_n - \mathbf{Q}_{F_0}) (\tilde{\theta}_F - \tilde{\theta}_{F_0}) \right| > \varepsilon \right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

**PROOF.** We will [using the notation of Pollard (1984), page 150] denote the empirical process  $\sqrt{n}(\mathbf{Q}_n - \mathbf{Q}_{F_0})$  by  $E_n$  and the symmetrized empirical process  $E_n$  by  $E_n^0$ . Fix an (arbitrary)  $\varepsilon > 0$ . By the symmetrization lemma we have

$$\begin{aligned} & \Pr\{|E_n(r_F - r_{F_0})\gamma| > \varepsilon \text{ for some } F \in \mathcal{F}_n\} \\ & \leq 4 \Pr\{|E_n^0(r_F - r_{F_0})\gamma| > \frac{1}{4}\varepsilon \text{ for some } F \in \mathcal{F}_n\}. \end{aligned}$$

Let  $\varepsilon > 0$  and  $\eta \in (0, 1)$  be fixed in the following. We are going to show that

$$(3.32) \quad \Pr\{|E_n^0(r_F - r_{F_0})\gamma| > \frac{1}{4}\varepsilon \text{ for some } F \in \mathcal{F}_n \mid \xi_n\} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

for all

$$\xi_n = ((T_1, U_1, \delta_1, \gamma_1), \dots, (T_n, U_n, \delta_n, \gamma_n)),$$

such that

$$(3.33) \quad \int_{q_{F_0} \leq n^{-1/3}} d\mathbf{Q}_n \leq n^{-2/3} \log n,$$

$$(3.34) \quad \int_{q_{F_0} > n^{-1/3}} q_{F_0}^{-2} d\mathbf{Q}_n \leq (\log n)^2$$

and

$$(3.35) \quad \sup_{F \in \mathcal{F}_n} \mathbf{Q}_n D_\eta(F) = \sup_{F \in \mathcal{F}_n} \int_{q_F \leq \eta q_{F_0}} d\mathbf{Q}_n \leq n^{-2/3} (\log n)^2,$$

are satisfied for the empirical measure  $\mathbf{Q}_n$ , corresponding to  $\xi_n$ . By the preceding lemmas, the probability that these conditions are *not* satisfied for the sample  $\xi_n$  tends to zero, as  $n \rightarrow \infty$ . In (3.33) to (3.35) we use again our method of absorbing constants into extra powers of  $\log n$ .

Let  $\varepsilon_n = \frac{1}{16}\varepsilon/\sqrt{n}$  and let, for each  $\delta > 0$ ,  $\mathcal{S}_\delta$  be a (minimal) net of triples  $(F_k, G_k, \bar{\phi}_k)$  such that for any  $F \in \mathcal{F}_n$  there exists a triple  $(F_k, G_k, \bar{\phi}_k) \in \mathcal{S}_\delta$  satisfying  $F_k \leq F \leq G_k, F_k, G_k \in \mathcal{F}$  and

$$d_n((F_k, G_k, \bar{\phi}_k), (F, F, \bar{\phi}_F)) < \delta/C,$$

where the constant  $C > 0$  is as in Lemma 3.5, and where the Lipschitz norm of  $\bar{\phi}_k$  is bounded above by the Lipschitz norm of the class  $\{\bar{\phi}_F: F \in \mathcal{F}\}$ . Then (3.32) will hold if we can show that, for some  $\varepsilon'_n \leq \varepsilon_n$ ,

$$(3.36) \quad \Pr\left(|E_n^0(r_{F_k, G_k, \bar{\phi}_k} - r_{F_0})\gamma| > \frac{1}{8}\varepsilon \text{ for some } (F_k, G_k, \bar{\phi}_k) \in \mathcal{S}'_{\varepsilon'_n} \mid \xi_n\right) \rightarrow 0$$

as  $n \rightarrow \infty$ ,

since, conditionally on  $\xi_n$ ,

$$\begin{aligned} & \left|E_n^0(r_F - r_{F_0})\gamma\right| \\ & \leq \left|E_n^0(r_F - r_{F_k, G_k, \bar{\phi}_k})\gamma\right| + \left|E_n^0(r_{F_k, G_k, \bar{\phi}_k} - r_{F_0})\gamma\right| \\ & \leq n^{1/2}Cd_n\left((F_k, G_k, \bar{\phi}_k), (F, F, \bar{\phi}_F)\right) + \left|E_n^0(r_{F_k, G_k, \bar{\phi}_k} - r_{F_0})\gamma\right| \\ & \quad + \mathcal{O}\left(n^{-1/6}(\log n)^2\right) \\ & \leq \left|E_n^0(r_{F_k, G_k, \bar{\phi}_k} - r_{F_0})\gamma\right| + n^{1/2}\varepsilon_n + o(1) \\ & \leq \left|E_n^0(r_{F_k, G_k, \bar{\phi}_k} - r_{F_0})\gamma\right| + \frac{1}{16}\varepsilon + \frac{1}{16}\varepsilon = \left|E_n^0(r_{F_k, G_k, \bar{\phi}_k} - r_{F_0})\gamma\right| + \frac{1}{8}\varepsilon, \end{aligned}$$

for all large  $n$ .

We now construct a chain in the following way. Let

$$\delta_i = 3^{-i}n^{-1/12}, \quad i = 0, 1, \dots,$$

and let  $k$  be the smallest integer such that  $3^{-k}n^{-1/12} \leq \varepsilon_n$ . Define  $\mathcal{S}_{\delta_k} = \mathcal{S}'_{\delta_k}$  and let, recursively,  $\mathcal{S}_{\delta_{i-1}}$  be a minimal  $\delta_{i-1}$ -net for the semidistance  $d_n$ , such that for each triple  $(F_{\delta_i}, G_{\delta_i}, \bar{\phi}_{\delta_i}) \in \mathcal{S}_{\delta_i}$  there exists a triple  $(F_{\delta_{i-1}}, G_{\delta_{i-1}}, \bar{\phi}_{\delta_{i-1}}) \in \mathcal{S}_{\delta_{i-1}}$  satisfying  $F_{\delta_{i-1}}, G_{\delta_{i-1}} \in \mathcal{F}$ ,  $F_{\delta_{i-1}} \leq F_{\delta_i} \leq G_{\delta_i} \leq G_{\delta_{i-1}}$  and

$$d_n\left((F_{\delta_{i-1}}, G_{\delta_{i-1}}, \bar{\phi}_{\delta_{i-1}}), (F_{\delta_i}, G_{\delta_i}, \bar{\phi}_{\delta_i})\right) < \delta_{i-1}/C,$$

where the constant  $C > 0$  is as in Lemma 3.5. The cardinality  $N_n(\delta_i)$  of  $\mathcal{S}_{\delta_i}$  satisfies

$$\log N_n(\delta_i) \leq c\delta_i^{-1} \log n, \quad i = 0, \dots, k,$$

for some constant  $c > 0$ , using (3.34) and the entropy results in Birman and Solomjak (1967). Hence,

$$(3.37) \quad \sum_{i=0}^{k-1} \delta_i \sqrt{\log N_n(\delta_{i+1})} \leq c(\log n)^{1/2} n^{-1/24},$$

for some constant  $c > 0$ .

Furthermore, let, recursively, starting with an element in  $\mathcal{S}_{\delta_k} = \mathcal{S}'_{\delta_k}$   $(F_{\delta_{i-1}}, G_{\delta_{i-1}}, \bar{\phi}_{\delta_{i-1}})$  be the closest point to  $(F_{\delta_i}, G_{\delta_i}, \bar{\phi}_{\delta_i})$  in  $\mathcal{S}_{\delta_{i-1}}$  for the semidistance  $d_n$ , such that  $F_{\delta_{i-1}} \leq F_{\delta_i} \leq G_{\delta_i} \leq G_{\delta_{i-1}}$ . This is what we meant by the ‘‘funnel structure’’; the pair  $(F_{\delta_i}, G_{\delta_i})$  is ‘‘narrowing’’ the pair  $(F_{\delta_{i-1}}, G_{\delta_{i-1}})$ .

Defining  $H(\delta_i) = \{2 \log(N_n(\delta_i)^2/\delta_i)\}^{1/2}$  and  $\eta_i = \delta_i H_n(\delta_{i+1})$ , we get, for all large  $n$ , using Hoeffding’s inequality [see, e.g., Pollard (1984) page 161, for a similar construction of the  $\eta_i$ ’s] and using the fact that, by (3.37),  $\sum_{i=0}^{k-1} \eta_i =$

$o(1)$ ,

$$\begin{aligned}
 & \Pr \left\{ \max_{(F_{\delta_k}, G_{\delta_k}, \phi_{\delta_k}) \in \mathcal{F}_{\delta_k}} \left| E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}}) \gamma \right| > \varepsilon/32 \mid \xi_n \right\} \\
 & \leq \Pr \left\{ \max_{(F_{\delta_k}, G_{\delta_k}, \phi_{\delta_k}) \in \mathcal{F}_{\delta_k}} \left| E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}}) \right. \right. \\
 & \qquad \qquad \qquad \left. \left. \times \gamma \mathbf{1}_{C_{n, \eta}(F_{\delta_k}, G_{\delta_k})} \right| > \varepsilon/64 \mid \xi_n \right\} \\
 & \leq \Pr \left\{ \max_{(F_{\delta_k}, G_{\delta_k}, \phi_{\delta_k}) \in \mathcal{F}_{\delta_k}} \left| E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}}) \right. \right. \\
 & \qquad \qquad \qquad \left. \left. \times \gamma \mathbf{1}_{C_{n, \eta}(F_{\delta_k}, G_{\delta_k})} \right| > \sum_{i=0}^{k-1} \eta_i \mid \xi_n \right\} \\
 & \leq \sum_{i=0}^{k-1} \Pr \left\{ \max_{(F_{\delta_{i+1}}, G_{\delta_{i+1}}, \bar{\phi}_{\delta_{i+1}}) \in \mathcal{F}_{\delta_{i+1}}} \left| E_n^0(r_{F_{\delta_{i+1}}, G_{\delta_{i+1}}, \bar{\phi}_{\delta_{i+1}}} - r_{F_{\delta_i}, G_{\delta_i}, \bar{\phi}_{\delta_i}}) \gamma \right| \right. \\
 & \qquad \qquad \qquad \left. \times \mathbf{1}_{C_{n, \eta}(F_{\delta_{i+1}}, G_{\delta_{i+1}})} > \eta_i \mid \xi_n \right\} \\
 & \leq 2 \sum_{i=0}^{k-1} N_n(\delta_{i+1}) \exp\left\{-\frac{1}{2} \eta_i^2 / \delta_i^2\right\} \\
 & \leq 2 \sum_{i=0}^{k-1} N_n(\delta_{i+1}) \exp\left\{-\log(N_n(\delta_{i+1})^2 / \delta_{i+1})\right\} \\
 & \leq 2 \sum_{i=0}^k \delta_i = \mathcal{O}(n^{-1/12}).
 \end{aligned}$$

We also have, if  $A_0$  is the set of triples that can occur at the (coarse) end of a chain as constructed above, and if  $(F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}) \in A_0$ ,

$$\begin{aligned}
 |E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_0}) \gamma| & \leq |E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}}) \gamma| \\
 & \quad + |E_n^0(r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}} - r_{F_0}) \gamma|,
 \end{aligned}$$

for some  $(F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k})$  at the beginning (fine end) of the chain. Moreover,

$$\begin{aligned}
 & \left| E_n^0(r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}} - r_{F_0}) \gamma \right| \\
 & \leq \left| E_n^0(r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}} - r_F) \gamma \right| + \left| E_n^0(r_F - r_{F_0}) \gamma \right| \\
 & \leq n^{1/2} \varepsilon_n + \left| E_n^0(r_F - r_{F_0}) \gamma \right| = \varepsilon/32 + \left| E_n^0(r_F - r_{F_0}) \gamma \right|,
 \end{aligned}$$

for some  $F \in \mathcal{F}_n$ , by the properties of  $\mathcal{F}_{\delta_k}$  and the construction of the chain.



But for an  $F \in \mathcal{F}_n$  we have, by Hoeffding’s inequality, for all large  $n$ ,

$$\begin{aligned} \Pr\left\{ \left| E_n^0(r_F - r_{F_0})\gamma \right| > \varepsilon/32 \mid \xi_n \right\} &\leq 2 \exp\left\{ -\frac{1}{2}(\varepsilon/32)^2 / Q_n(r_F - r_{F_0})^2 \right\} \\ &\leq 2 \exp\left\{ -cn^{1/12}\varepsilon^2 / (\log n)^2 \right\}, \end{aligned}$$

for a constant  $c > 0$ , since by Lemmas 3.5, 5.1 and 5.2, and (3.33) to (3.35),

$$\sup_{F \in \mathcal{F}_n} Q_n(r_F - r_{F_0})^2 \gamma \leq kn^{-1/6}(\log n)^2,$$

for some  $k > 0$  and all large  $n$ . Hence, using the correspondence between  $(F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0})$  and  $(F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k})$  along the chain, but this time doing the counting at the “coarse end” of the chain, we get

$$\begin{aligned} &\Pr\left\{ \sup_{(F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}) \in A_0} \left| E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_0})\gamma \right| > \frac{3}{32}\varepsilon \mid \xi_n \right\} \\ &\leq \Pr\left\{ \sup_{(F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}) \in A_0} \left| E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}})\gamma \right| > \frac{\varepsilon}{32} \mid \xi_n \right\} \\ &\quad + \Pr\left\{ \sup_{(F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}) \in A_0} \left| E_n^0(r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}} - r_{F_0})\gamma \right| > \frac{\varepsilon}{16} \mid \xi_n \right\} \\ &\leq \exp\left\{ c_1 n^{1/12} \log n - c_2 \varepsilon^2 n^{1/6} / (\log n)^2 \right\} + o(1) \rightarrow 0 \quad \text{as } n \rightarrow \infty, \end{aligned}$$

for constants  $c_1, c_2 > 0$ , since the number of triples in the  $\delta_0$ -net for the semidistance  $d_n$  is  $\exp(\mathcal{O}(n^{1/12} \log n))$ . Thus we get

$$\begin{aligned} &\Pr\left\{ \max_{(F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}) \in \mathcal{F}_{\delta_k}} \left| E_n^0(r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}} - r_{F_0})\gamma \right| > \frac{\varepsilon}{8} \mid \xi_n \right\} \\ &\leq \Pr\left\{ \max_{(F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}) \in \mathcal{F}_{\delta_k}} \left| E_n^0(r_{F_{\delta_k}, G_{\delta_k}, \bar{\phi}_{\delta_k}} - r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}})\gamma \right| > \frac{\varepsilon}{32} \mid \xi_n \right\} \\ &\quad + \Pr\left\{ \sup_{(F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}) \in A_0} \left| E_n^0(r_{F_{\delta_0}, G_{\delta_0}, \bar{\phi}_{\delta_0}} - r_{F_0})\gamma \right| > \frac{3}{32}\varepsilon \mid \xi_n \right\} \\ &\rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

This proves (3.36).

In a very similar way, it is shown that

$$(3.38) \quad \Pr\left\{ \left| E_n \left( \frac{\phi_{\hat{F}_n}(t)}{\hat{F}_n(t)} - \frac{\phi_{F_0}(t)}{F_0(t)} \right) \delta \right| > \varepsilon \right\} \rightarrow 0,$$

and

$$(3.39) \quad \Pr \left\{ \left| E_n \left( \frac{\phi_{\hat{F}_n}(u)}{1 - \hat{F}_n(u)} - \frac{\phi_{F_0}(u)}{1 - F_0(t)} \right) (1 - \gamma - \delta) \right| > \varepsilon \right\} \rightarrow 0,$$

as  $n \rightarrow \infty$ . For example, to prove (3.38), we condition on a sample

$$\xi_n = ((T_1, U_1, \delta_1, \gamma_1), \dots, (T_n, U_n, \delta_n, \gamma_n)),$$

such that conditions (3.33) to (3.35) are satisfied for the empirical measure  $Q_n$ , corresponding to  $\xi_n$ . We then approximate pairs  $(F, \bar{\phi}_F)$  by pairs  $(F_k, \bar{\phi}_k)$ , where  $F_k \geq F$ . The remaining part of the argument is the same (and in fact easier). The reason for treating the ratios  $r_F$  separately was mainly notational.

The result now follows from (3.32), (3.38) and (3.39).  $\square$

#### 4. Simulations.

4.1. *Computation of  $\phi_{\hat{F}_n}$  and  $\bar{\phi}_{\hat{F}_n}$ .* If  $F$  is a purely discrete distribution function, we know from Theorem 3.1 that  $\phi_F$ , as given by (3.2), is a piecewise constant function as well. In this equation, we do not need the  $\psi_F$ -part in order to obtain the  $\phi$ -solution. We know from part (iii) of Theorem 3.1 that the values of  $\phi_F$  can be found from a finite set of linear equations  $Ay = b$ , where [as noted in the proof of part (iii) of Theorem 3.1] the matrix  $A$  is a symmetric, strictly diagonally dominant  $M$ -matrix. Such a matrix can be shown to be positive definite [see Berman and Plemmons (1979)]. So Cholesky decomposition can be used, which is a fast algorithm and numerically stable.

The solution of the integral equation in the transformed scale is easily obtained from this, since the integral parts are with respect to a measure that has mass restricted to the values  $\tau_i = \hat{F}_n(x_i)$ . In Figure 1 we give a picture of the NPMLE and in Figures 2 to 4 we give the solutions  $\bar{\phi}_{\hat{F}_n}$ ,  $\bar{\xi}_{\hat{F}_n}$  and  $\bar{\phi}'_{\hat{F}_n}$ , respectively, based on a random sample of size  $n = 300$  from a uniform distribution on  $[0, 1]$ , censored by two uniformly distributed observation times (so  $H$  is the uniform distribution on the upper triangle of the unit square), where  $k \equiv 1$ , which is the derivative of the canonical gradient for the mean functional.

These solutions are compared with the solution  $\bar{\phi}_{F_0}$ , with  $F_0 = U(0, 1)$ , implying  $\bar{\phi}_{F_0} \equiv \phi_{F_0}$ . This solution is obtained in Geskus (1992), based on a power series expansion using Legendre polynomials. The function  $\bar{\xi}_{\hat{F}_n}$  is defined by  $\bar{\xi}_{\hat{F}_n}(t) = \bar{\phi}_{\hat{F}_n}(t)/(t(1 - t))$ ,  $t \in (0, 1)$  and likewise  $\bar{\xi}_{F_0}(t) = \bar{\phi}_{F_0}(t)/(t(1 - t))$ ,  $t \in (0, 1)$ .

The number of jumps of the NPMLE was 15 and the locations of the jumps are indicated by small vertical bars (slightly smaller than the tickmarks at 0.25, etc.) on the  $x$ -axis in Figure 1. On the other hand, in Figures 2 to 4, the

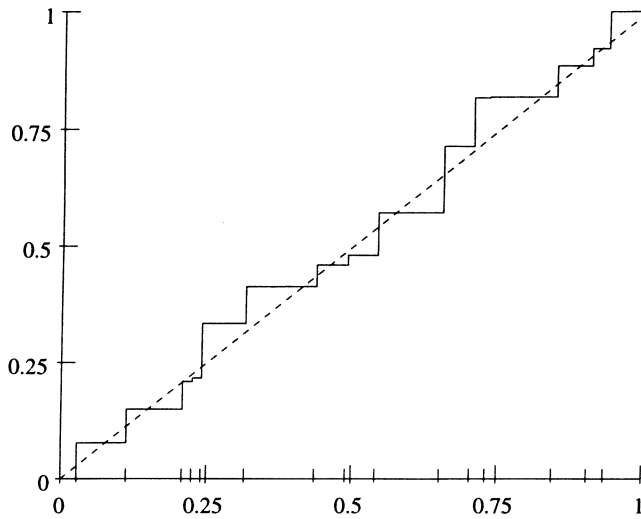


FIG. 1.  $\hat{F}_n$  based on sample size 300 and  $F_0$  (dashed).

small vertical bars on the  $x$ -axis denote the values of  $\hat{F}_n$  at these points of jump. The derivative  $\bar{\phi}'_{\hat{F}_n}$  is actually continuous in this case (this will generally not be the case), and has cusps at the points  $\hat{F}_n(x_i)$ . In Geskus (1997) it is shown that the cusps of the derivative  $\bar{\phi}'_{\hat{F}_n}$  are located on the curve  $t \mapsto \frac{1}{2}(1 - 2t)\bar{\xi}_{\hat{F}_n}(t)$ ,  $t \in (0, 1)$ .

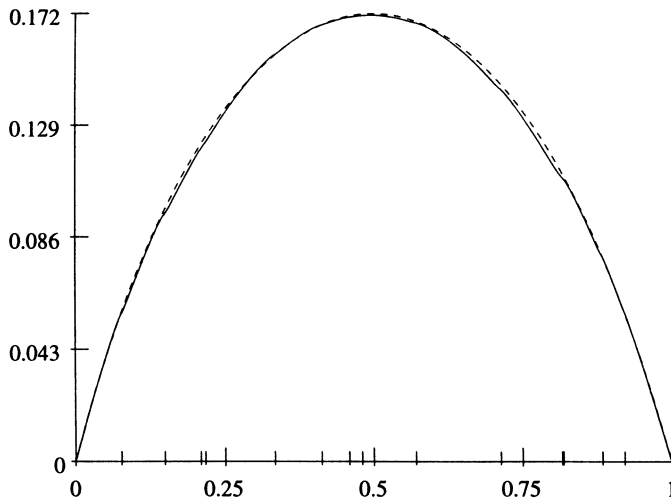


FIG. 2.  $\bar{\phi}'_{\hat{F}_n}$  and  $\bar{\phi}'_{F_0}$  (dashed).

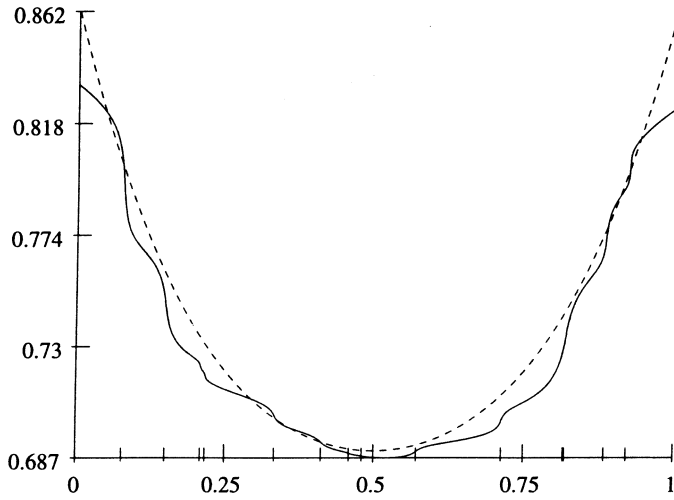


FIG. 3.  $\bar{\xi}_{\hat{F}_n}$  and  $\bar{\xi}_{F_0}$  (dashed).

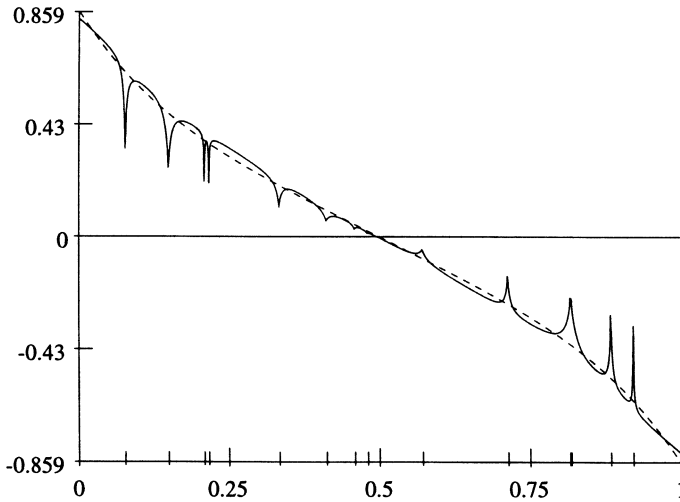


FIG. 4.  $\bar{\phi}'_{\hat{F}_n}$  and  $\bar{\phi}'_{F_0}$  (dashed).

4.2. *A simulation of  $K(\hat{F}_n)$ .* For the same uniform case as above, we did a computer experiment of 10,000 samples of magnitude 1000, and estimated the mean  $\mu(F_0)$  by the NPMLE  $\mu(\hat{F}_{1000})$ . Estimating the variance of  $\sqrt{1000}(\mu(\hat{F}_{1000}) - \mu(F_0))$  by the unbiased estimator  $S_{10,000}^2$  yielded the number 0.11917, while analytic computations as in Geskus (1992) yield 0.1198987 for the information lower bound. So the estimate is very close to the information lower bound.

APPENDIX

PROOF OF THEOREM 3.1. (i) Let  $F_\alpha = (1 - \alpha)F_0 + \alpha F$ . For  $\alpha \in (0, 1)$ , the function  $F_\alpha$  is strictly increasing and continuous between jumps and hence the solution  $\phi_{F_\alpha}$  to the integral equation exists by Theorem 2.2 in Section 2. For simplicity of notation, we will denote  $\phi_{F_\alpha}$  by  $\phi_\alpha$ . Furthermore, we will use the same notation as in the proof of Lemma 2.1:  $x_1, \dots, x_m$  are the points of jump of  $F$ , contained in  $(0, M)$  and  $x_0 = 0, x_{m+1} = M$ ; moreover,  $\tau_i = F(x_i), i = 0, \dots, m + 1$ .

Let  $G_\alpha = F_\alpha^{-1}$ , with derivative  $g_\alpha$ . Furthermore, let  $\bar{k}_\alpha = k \circ G_\alpha, \bar{h}_{i,\alpha} = h_i \circ G_\alpha, \bar{H}_\alpha(t, u) = H(G_\alpha(t), G_\alpha(u))$  and  $\bar{h}_\alpha(t, u) = h(G_\alpha(t), G_\alpha(u)), 0 < t < u < 1$ . Likewise, we define  $\bar{d}_\alpha = \bar{d}_{F_\alpha}$  and  $\bar{c}_\alpha = \bar{c}_{F_\alpha}$ . Then, by Theorem 2.2,  $\bar{\phi}_\alpha$  is the unique solution of the integral equation

$$\begin{aligned} \bar{\phi}_\alpha(t) = \bar{d}_\alpha(t) & \left\{ \bar{k}_\alpha(t) - \int_{t' \in [0, t)} \frac{\bar{\phi}_\alpha(t) - \bar{\phi}_\alpha(t')}{t - t'} d\bar{H}_\alpha(t', t) \right. \\ & \left. + \int_{u \in (t, 1]} \frac{\bar{\phi}_\alpha(u) - \bar{\phi}_\alpha(t)}{u - t} d\bar{H}_\alpha(t, u) \right\}, \quad t \in [0, 1]. \end{aligned}$$

Let the set  $D_\alpha$  be defined by

$$\begin{aligned} D_\alpha = \{ & \text{discontinuity points of } g_\alpha(t), \text{ augmented with } 0 \text{ and } 1\} \\ & \cup \left\{ \text{discontinuity points of } \bar{k}'_\alpha(t), \bar{d}'_\alpha(t), \right. \\ & \left. \Delta_{t'}^1(y) = \frac{\partial}{\partial t} \bar{h}_\alpha(t, t') \text{ for } t \leq t', \text{ and } \Delta_u^2(t) = \frac{\partial}{\partial t} \bar{h}_\alpha(u, t) \text{ for } t \geq u \right\} \end{aligned}$$

and let  $A_{\alpha, \delta}(t)$  and  $B_{\alpha, \delta}(t)$  be defined by

$$\begin{aligned} (A.1) \quad A_{\alpha, \delta}(t) = \bar{d}_\alpha(t) & \left\{ \int_{t-\delta}^t \left| \frac{\partial}{\partial t} \bar{h}_\alpha(t', t) \right| dG_\alpha(t') \right. \\ & \left. + \int_t^{t+\delta} \left| \frac{\partial}{\partial t} \bar{h}_\alpha(t, u) \right| dG_\alpha(u) \right\}, \end{aligned}$$

and

$$\begin{aligned} (A.2) \quad B_{\alpha, \delta}(t) = \bar{d}_\alpha(t) & |\bar{k}'_\alpha(t)| \\ & + |\bar{d}'_\alpha(t)| \left[ (1 - t) \bar{h}_{1,\alpha}(t) + t \bar{h}_{2,\alpha}(t) \right] \sup_{t' \in [0, 1]} \{ \bar{c}_\alpha(t') |\bar{k}_\alpha(t') \} \\ & + \frac{2\bar{d}_\alpha(t)}{\delta} \sup_{t' \in [0, 1]} \{ \bar{d}_\alpha(t') |\bar{k}_\alpha(t') \} \\ & \times \left\{ \sup_{t' \in [0, t]} \left| \frac{\partial}{\partial t} \bar{h}_\alpha(t', t) \right| + \sup_{u \in [t, 1]} \left| \frac{\partial}{\partial t} \bar{h}_\alpha(t, u) \right| \right\}. \end{aligned}$$

Moreover, let

$$(A.3) \quad C_\alpha(t) = 1 + 2\bar{d}_\alpha(t)g_\alpha(t)\bar{h}_\alpha(t, t).$$

As in Theorem 2.2, we have that at points of  $D_\alpha$  the functions  $A_{\alpha, \delta}$  and  $B_{\alpha, \delta}$  have two versions, one corresponding to taking left derivatives and one corresponding to taking right derivatives. Then there exists a  $\delta > 0$  such that

$$\sup_{t \in [0, 1]} A_{\alpha, \delta}(t)/C_\alpha(t) \leq 1/2.$$

By Theorem 2.2 we have

$$|\bar{\phi}_\alpha(u) - \bar{\phi}_\alpha(t)| \leq K_\alpha(u - t), \quad 0 \leq t < u \leq 1,$$

where  $K_\alpha$  is given by

$$(A.4) \quad K_\alpha = 2 \sup_{t \in [0, 1]} B_{\alpha, \delta}(t)/C_\alpha(t),$$

for  $\delta > 0$  such that  $\sup_{t \in [0, 1]} A_{\alpha, \delta}(t)/C_\alpha(t) \leq 1/2$ . We need to show that we can choose  $\delta$  and  $K_\alpha$  independently of  $\alpha$  and  $F$  in a small (supremum distance) neighborhood of  $F_0$ .

If (using the same notation as in the proof of Lemma 2.2)  $t$  belongs to an interval  $(\tau_i, \tau'_i)$ , on which  $G_\alpha$  increases, then, going back to the original scale, we get

$$\begin{aligned} \frac{A_{\alpha, \delta}(t)}{C_\alpha(t)} &\leq \sup_{x \in (0, M)} [2h(x, x)]^{-1} \left[ \int_{x': F_\alpha(x) - \delta < F_\alpha(x') < F_\alpha(x)} \left| \frac{\partial}{\partial x} h(x', x) \right| dx' \right. \\ &\quad \left. + \int_{y: F_\alpha(x) < F_\alpha(y) < F_\alpha(x) + \delta} \left| \frac{\partial}{\partial x} h(x, y) \right| dy \right]. \end{aligned}$$

The essential observation here is that, although  $A_{\alpha, \delta}(t)$  tends to  $\infty$ , as  $\alpha \uparrow 1$ , for points  $t$  in the range of  $F_\alpha$ , the ratio  $A_{\alpha, \delta}(t)/C_\alpha(t)$  stays bounded, since the factor  $g_\alpha(t)$ , causing the steep increase of  $A_{\alpha, \delta}(t)$  via  $(\partial/\partial t)\bar{h}_\alpha$ , also occurs in the denominator  $C_\alpha(t)$ .

If, on the other hand,  $t$  belongs to an interval  $(\tau'_i, \tau_{i+1})$ , on which  $G_\alpha$  is constant, then  $A_{\alpha, \delta}(t) = 0$ , since  $g_\alpha(t) = 0$  on such an interval. Hence we can choose  $\delta > 0$  such that

$$\sup_{t \in [0, 1]} A_{\alpha, \delta}(t)/C_\alpha(t) \leq 1/2$$

for all  $\alpha \in (0, 1)$  and all  $F$  such that  $\sup_{x \in [0, M]} |F(x) - F_0(x)| \leq \varepsilon$ , for a fixed suitably chosen  $\varepsilon > 0$ .

In a similar way we get, using (2.27) in Section 2,

$$\begin{aligned}
 \frac{B_{\alpha, \delta}(t)}{C_{\alpha}(t)} &\leq \sup_{x \in (0, M)} \frac{|k'(x)|}{2h(x, x)} \\
 &+ \sup_{x \in (0, M)} \frac{1 + |h_1(x) - h_2(x)| + |h'_1(x)| + |h'_2(x)|}{h(x, x)} \\
 (A.5) \quad &\times \sup_{x \in (0, M)} \frac{k(x)}{\varepsilon_1 h(x, x)} \\
 &+ \frac{1}{\delta \inf_{x \in (0, M)} h(x, x)} \sup_{x \in (0, M)} \frac{2k(x)}{h_1(x) + h_2(x)} \\
 &\times \left\{ \sup_{x, y \in (0, M)} \left| \frac{\partial}{\partial x} h(x, y) \right| + \sup_{x, y \in (0, M)} \left| \frac{\partial}{\partial y} h(x, y) \right| \right\},
 \end{aligned}$$

for some  $\varepsilon_1 > 0$ , uniform over  $\alpha$  and  $F$ , implying that  $\sup_{t \in [0, 1]} B_{\alpha, \delta}(t)/C_{\alpha}(t)$  and hence also  $K_{\alpha}$  in (5.4) has a finite upper bound [given by the right-hand side of (A.5)] which is independent of  $\alpha$  and  $F$ .

It follows that the sequence  $(\bar{\phi}_{\alpha})$  is equicontinuous and hence has a subsequence, converging to a function  $\bar{\phi}_F$  which is Lipschitz on  $[0, 1]$ . Let  $(\alpha_n)_{n=1, 2, \dots}$  be a sequence of numbers such that  $\alpha_n \uparrow 1$  and  $\bar{\phi}_{\alpha_n} \rightarrow \bar{\phi}_F$  in the supremum distance. Define

$$(A.6) \quad \phi_F(x) := \bar{\phi}_F(F(x)).$$

Then, by the equicontinuity of the sequence  $(\bar{\phi}_{\alpha_n})$ , we obtain, for each  $x \in [0, M]$ ,

$$(A.7) \quad \phi_F(x) := \bar{\phi}_F(F(x)) = \lim_{n \rightarrow \infty} \bar{\phi}_{\alpha_n}(F_{\alpha_n}(x)) = \lim_{n \rightarrow \infty} \phi_{\alpha_n}(x),$$

$x \in [0, M].$

Now let  $t \in [0, 1] \setminus D$ . Then  $t$  is a point of continuity of  $G$  and does not belong to the range of  $F$ . We have

$$\begin{aligned}
 (A.8) \quad \bar{\phi}_F(t) &= \lim_{n \rightarrow \infty} \bar{\phi}_{\alpha_n}(t) \\
 &= \lim_{n \rightarrow \infty} \bar{d}_{\alpha_n}(t) \left\{ \bar{k}_{\alpha_n}(t) - \int_{t' \in [0, t)} \frac{\bar{\phi}_{\alpha_n}(t) - \bar{\phi}_{\alpha_n}(t')}{t - t'} d\bar{H}_{\alpha_n}(t', t) \right. \\
 &\quad \left. + \int_{u \in (t, 1]} \frac{\bar{\phi}_{\alpha_n}(u) - \bar{\phi}_{\alpha_n}(t)}{u - t} d\bar{H}_{\alpha_n}(t, u) \right\} \\
 &= \bar{d}_F(t) \left\{ \bar{k}_F(t) - \lim_{n \rightarrow \infty} \int_{t' \in [0, t)} \frac{\bar{\phi}_F(t) - \bar{\phi}_{\alpha_n}(t')}{t - t'} d\bar{H}_{\alpha_n}(t', t) \right. \\
 &\quad \left. + \lim_{n \rightarrow \infty} \int_{u \in (t, 1]} \frac{\bar{\phi}_{\alpha_n}(u) - \bar{\phi}_F(t)}{u - t} d\bar{H}_{\alpha_n}(t, u) \right\}.
 \end{aligned}$$

Suppose  $F(x_i) < y < F(x_{i+1})$ . Hence  $G(y) = x_{i+1}$ . Then (5.8) can be written

$$\begin{aligned}
 \bar{\phi}_F(t) &= \frac{t(1-t)}{h_1(x_{i+1})(1-t) + h_2(x_{i+1})t} \\
 \text{(A.9)} \quad &\times \left\{ k(x_{i+1}) - \lim_{n \rightarrow \infty} \int_{x' \in [0, G_{\alpha_n}(t)]} \frac{\bar{\phi}_F(t) - \phi_{\alpha_n}(x')}{t - F_{\alpha_n}(x')} dH(x', G_{\alpha_n}(t)) \right. \\
 &\quad \left. + \lim_{n \rightarrow \infty} \int_{y \in (G_{\alpha_n}(t), M]} \frac{\phi_{\alpha_n}(y) - \bar{\phi}_F(t)}{F_{\alpha_n}(y) - t} dH(G_{\alpha_n}(t), y) \right\}
 \end{aligned}$$

and by the dominated convergence theorem and (5.7), we get

$$\begin{aligned}
 \bar{\phi}_F(t) &= \frac{t(1-t)}{h_1(x_{i+1})(1-t) + h_2(x_{i+1})t} \\
 &\times \left\{ k(x_{i+1}) - \int_{x': F(x') < t} \frac{\bar{\phi}_F(t) - \phi_F(x')}{t - F(x')} dH(x', x_{i+1}) \right. \\
 &\quad \left. + \int_{y: F(y) > t} \frac{\phi_F(y) - \bar{\phi}_F(t)}{F(y) - t} dH(x_{i+1}, y) \right\} \\
 \text{(A.10)} \quad &= \frac{t(1-t)}{\bar{h}_1(t)(1-t) + \bar{h}_2(t)t} \\
 &\times \left\{ \bar{k}_F(t) - \int_{t' \in [0, t]} \frac{\bar{\phi}_F(t) - \bar{\phi}_F(t')}{t - t'} d\bar{H}(t', t) \right. \\
 &\quad \left. + \int_{u \in (t, 1]} \frac{\bar{\phi}_F(u) - \bar{\phi}_F(t)}{u - t} d\bar{H}(t, u) \right\}.
 \end{aligned}$$

Letting  $t \downarrow F(x_i)$ , we find

$$\begin{aligned}
 \phi_F(x_i) &= \bar{\phi}_F(F(x_i)) \\
 &= \frac{F(x_i)(1 - F(x_i))}{h_1(x_{i+1})(1 - F(x_i)) + h_2(x_{i+1})F(x_i)} \\
 \text{(A.11)} \quad &\times \left\{ k(x_{i+1}) - \int_{x': F(x') < F(x_i)} \frac{\phi_F(x_i) - \phi_F(x')}{F(x_i) - F(x')} dH(x', x_{i+1}) \right. \\
 &\quad - \bar{\phi}'_F(F(x_i)) \int_{x': F(x') = F(x_i)} dH(x', x_{i+1}) \\
 &\quad \left. + \int_{y: F(y) > F(x_i)} \frac{\phi_F(y) - \phi_F(x_i)}{F(y) - F(x_i)} dH(x_{i+1}, y) \right\},
 \end{aligned}$$



and letting  $t \uparrow F(x_{i+1})$ , we get

$$\begin{aligned}
 \phi_F(x_{i+1}) &= \bar{\phi}_F(F(x_{i+1})) \\
 &= \frac{F(x_{i+1})(1 - F(x_{i+1}))}{h_1(x_{i+1})(1 - F(x_{i+1})) + h_2(x_{i+1})F(x_{i+1})} \\
 (A.12) \quad &\times \left\{ k(x_{i+1}) - \int_{x': F(x') < F(x_{i+1})} \frac{\phi_F(x_{i+1}) - \phi_F(x')}{F(x_{i+1}) - F(x')} dH(x', x_{i+1}) \right. \\
 &\quad + \bar{\phi}'_F(F(x_{i+1}) -) \int_{y: F(y) = F(x_{i+1})} dH(x_{i+1}, y) \\
 &\quad \left. + \int_{y: F(y) > F(x_{i+1})} \frac{\phi_F(y) - \phi_F(x_{i+1})}{F(y) - F(x_{i+1})} dH(x_{i+1}, y) \right\}.
 \end{aligned}$$

Note that the right and left derivatives

$$\bar{\phi}'_F(F(x_i) +) := \lim_{t \downarrow F(x_i)} \{ \bar{\phi}_F(t) - \bar{\phi}_F(F(x_i)) \} / (t - F(x_i))$$

and

$$\bar{\phi}'_F(F(x_{i+1}) -) := \lim_{t \uparrow F(x_{i+1})} \{ \bar{\phi}_F(F(x_{i+1})) - \bar{\phi}_F(t) \} / (F(x_{i+1}) - t)$$

exist, since the other functions appearing in (A.10) are continuous and have finite limits as  $t \downarrow F(x_i)$  or  $t \uparrow F(x_{i+1})$ , respectively. These one-sided derivatives, so to speak, “catch” the discontinuities in the functions  $\bar{k}$ ,  $\bar{h}_1$ ,  $\bar{h}_2$  and  $\bar{H}$ , if one crosses a point  $F(x_i)$  in the range; the function  $\bar{\phi}_F$  is continuous at such a point and can be defined there by either taking the left-hand limit [involving  $h_i(x_i), k(x_i)$ , etc. at  $F(x_i)$ ], or the right-hand limit [involving  $h_i(x_{i+1}), k(x_{i+1})$ , etc. at  $F(x_i)$ ]. Uniqueness of  $\bar{\phi}_F$  will be proved below in part (iii).

(ii) We define, for  $\alpha \in (0, 1)$ , the functions  $\psi_\alpha: [0, M] \rightarrow \mathbb{R}$  by

$$(A.13) \quad \psi_\alpha(x) = \frac{\phi_\alpha(x) - \phi_\alpha(x_i)}{1 - \alpha}, \quad x \in [x_i, x_{i+1}), \quad i = 0, \dots, m,$$

and  $\psi_\alpha(M) = \psi_\alpha(M -)$ . Using (A.5), we get, for  $x, y$  in the same interval  $[x_i, x_{i+1})$ ,

$$\begin{aligned}
 (A.14) \quad |\psi_\alpha(y) - \psi_\alpha(x)| &= \frac{|\phi_\alpha(y) - \phi_\alpha(x)|}{1 - \alpha} = \frac{|\bar{\phi}_\alpha(F_\alpha(y)) - \bar{\phi}_\alpha(F_\alpha(x))|}{1 - \alpha} \\
 &\leq c \frac{|F_\alpha(y) - F_\alpha(x)|}{1 - \alpha} = c |F_0(y) - F_0(x)|,
 \end{aligned}$$

where  $c > 0$  is independent of  $\alpha$  and  $F \in \mathcal{F}$ , since we use the Lipschitz constant for  $\bar{\phi}_\alpha$  in (A.14). By the continuity of  $\bar{\phi}_\alpha$ , we can extend the function

$\psi_\alpha$ , restricted to an interval  $[x_i, x_{i+1})$  to a continuous function  $\psi_{\alpha,i}$ , defined on the closed interval  $[x_i, x_{i+1}]$ . The functions  $\psi_{\alpha,i}$  are equicontinuous in  $\alpha$  on the intervals  $[x_i, x_{i+1}]$  and hence have a convergent subsequence, converging (in the supremum metric for functions defined on  $[x_i, x_{i+1}]$ ) to a continuous function  $\tilde{\psi}_i$ , defined on  $[x_i, x_{i+1}]$ . Let  $\psi_F: [0, M] \rightarrow \mathbb{R}$  be the function, such that

$$\psi_F(x) = \tilde{\psi}_i(x), \quad x \in [x_i, x_{i+1}), \quad i = 0, \dots, m, \quad \psi_F(M) = \tilde{\psi}_m(M),$$

and let  $(\psi_{\alpha_n})_{n=1,2,\dots}$  be a sequence such that the restriction of  $\psi_{\alpha_n}$  to an interval  $[x_i, x_{i+1})$ ,  $i = 0, \dots, m - 1$ , or the interval  $[x_m, M]$  converges to  $\psi_F$  in the supremum metric for continuous functions on such an interval. Since the sequence  $(\bar{\phi}_{\alpha_n})$  is also equicontinuous, we can also assume (by switching to a further subsequence) that  $\bar{\phi}_{\alpha_n}$  converges in the supremum metric to a Lipschitz function  $\bar{\phi}_F$ , as in part (i). Then we have

$$\begin{aligned} \frac{\psi_F(y) - \psi_F(x)}{F_0(y) - F_0(x)} &= \lim_{n \rightarrow \infty} \frac{\phi_{\alpha_n}(y) - \phi_{\alpha_n}(x)}{(1 - \alpha_n)\{F_0(y) - F_0(x)\}} \\ (A.15) \qquad \qquad \qquad &= \lim_{n \rightarrow \infty} \frac{\phi_{\alpha_n}(y) - \phi_{\alpha_n}(x)}{F_{\alpha_n}(y) - F_{\alpha_n}(x)}, \end{aligned}$$

for  $x_i \leq x < y < x_{i+1}$ ,  $i = 0, \dots, m$ . Since, by Theorem 2.2, part (iii),  $\phi_\alpha$  satisfies the integral equation

$$\begin{aligned} \phi_\alpha(x) = d_\alpha(x) &\left\{ k(x) - \int_{x' \in [0, x)} \frac{\phi_\alpha(x) - \phi_\alpha(x')}{F_\alpha(x) - F_\alpha(x')} dH(x', x) \right. \\ &\left. + \int_{y \in (x, M]} \frac{\phi_\alpha(y) - \phi_\alpha(x)}{F_\alpha(y) - F_\alpha(x)} dH(x, y) \right\}, \end{aligned}$$

we now get, by (A.7), (A.15) and the dominated convergence theorem,

$$\begin{aligned} \phi_F(x) = d_F(x) &\left\{ k(x) - \int_{x' \in [0, x)} r_F(x', x) dH(x', x) \right. \\ (A.16) \qquad \qquad \qquad &\left. + \int_{y \in (x, M]} r_F(x, y) dH(x, y) \right\}, \end{aligned}$$

where  $r_F$  is defined by (3.3).

The function  $\phi_F$  is absolutely continuous with respect to  $F$ , since, by the Lipschitz property of  $\bar{\phi}_F$ ,

$$\begin{aligned} |\phi_F(y) - \phi_F(x)| &= |\bar{\phi}_F(F(y)) - \bar{\phi}_F(F(x))| \\ &\leq c|F(y) - F(x)|, \quad x, y \in [0, M]. \end{aligned}$$

This shows in particular that constancy of  $F$  on an interval implies constancy of  $\phi_F$  on that same interval. Moreover, by (A.14) and the bounded differentiability of  $F_0$ , we have that  $\psi_\alpha$  is Lipschitz on each interval  $[x_i, x_{i+1})$ , and hence  $\psi_F$  is also Lipschitz on such an interval.

(iii) Multiplying both sides of (A.16) by  $d_F(x)^{-1}$  and integrating from  $x_i$  to  $x_{i+1}$ , the  $\psi_F$ -part cancels and we get a finite set of linear equations  $Ay = b$  for  $y$ , given by (3.8). The matrix  $A$  is a symmetric, strictly diagonally dominant  $M$ -matrix (also called a Stieltjes matrix). This implies that the matrix equation has a unique solution and hence that the piecewise constant (right-continuous) function  $\phi_F$  is uniquely determined [see Berman and Plemmons (1979)]. The unicity of  $\bar{\phi}_F$  is easily obtained from this, since the integral parts of the equation for  $\bar{\phi}_F$  are with respect to a measure that has mass restricted to the values  $t = F(x_i)$ .  $\square$

The following two lemmas are used at the end of the chaining argument in the proof of Theorem 3.3.

LEMMA A.1. *We have*

$$(A.17) \quad \sup_{F \in \mathcal{F}_n} \sup_{x \in [0, M]} |F(x) - F_0(x)| = \mathcal{O}(n^{-1/6}).$$

PROOF. If  $F \in \mathcal{F}_n$  then, for each  $x \in [0, M]$ , using the Cauchy-Schwarz inequality,

$$\begin{aligned} \left| \int_{\{t \in [0, M]: |t-x| \leq n^{-1/6}\}} \{F(t) - F_0(t)\} dt \right| &\leq \sqrt{2} n^{-1/12} \|F - F_0\|_2 \\ &\leq cn^{-5/12} \{\log n\}^{1/2}, \end{aligned}$$

for some  $c > 0$  independent of  $x$ , using  $\|F - F_0\|_2 = \mathcal{O}(n^{-1/3} \{\log n\}^{1/2})$  and Proposition 3.3. By the monotonicity of  $F$  we have, if  $x + n^{-1/6} < M$ ,

$$(A.18) \quad \begin{aligned} n^{1/6} \{F(x) - F_0(x)\} &\leq n^{1/3} \int_x^{x+n^{-1/6}} F(t) dt - n^{1/6} F_0(x) \\ &\leq n^{1/3} \left| \int_x^{x+n^{-1/6}} \{F(t) - F_0(t)\} dt \right| \\ &\quad + n^{1/3} \int_x^{x+n^{-1/6}} \{F_0(t) - F_0(x)\} dt \\ &\leq k_1, \end{aligned}$$

for some constant  $k_1 > 0$ , independent of  $x$ . Likewise, if  $x > n^{-1/6}$ ,

$$(A.19) \quad \begin{aligned} n^{1/6} \{F(x) - F_0(x)\} &\geq n^{1/3} \int_{x-n^{-1/6}}^x F(t) dt - n^{1/6} F_0(x) \\ &\geq -n^{1/3} \left| \int_{x-n^{-1/6}}^x \{F(t) - F_0(t)\} dt \right| \\ &\quad - n^{1/3} \int_{x-n^{-1/6}}^x \{F_0(x) - F_0(t)\} dt \\ &\geq -k_2, \end{aligned}$$

for some constant  $k_2 > 0$ , independent of  $x$ . For  $x \in [0, n^{-1/6}]$  we get, from (A.18),

$$n^{1/6}\{F(x) - F_0(x)\} \leq k_1,$$

but since we also have

$$n^{1/6}\{F(x) - F_0(x)\} \geq -n^{1/6}F_0(n^{-1/6}) \geq -c,$$

for some  $c > 0$ , if  $x \in [0, n^{-1/6}]$ , the result also holds if  $x \in [0, n^{-1/6}]$ . A similar argument is used if  $x \in [M - n^{-1/6}, M]$ .  $\square$

LEMMA 5.2. *Let the conditions of Theorem 3.1 be satisfied. Then*

$$(A.20) \quad \sup_{F \in \mathcal{F}_n} \sup_{t \in [0, 1]} |\bar{\phi}_F(t) - \bar{\phi}_{F_0}(t)| = \mathcal{O}(n^{-1/12}).$$

PROOF. For  $\alpha \in (0, 1)$ , let  $F_{n,\alpha} = (1 - \alpha)F_0 + \alpha F_n$  and let  $\bar{\phi}_{n,\alpha} = \bar{\phi}_{F_{n,\alpha}}$ , using the same notation as in the proof of Theorem 3.1, but with an index  $n$  added to denote that we consider distribution functions  $F \in \mathcal{F}_n$  instead of distribution functions  $F \in \mathcal{F}$ . For  $\alpha = 1$ , we define  $\bar{\phi}_{n,\alpha} = \bar{\phi}_{F_n}$ , where  $\bar{\phi}_{F_n}$  is defined as in Theorem 3.1 and for  $\alpha = 0$  we define  $\bar{\phi}_0 = \bar{\phi}_{F_0}$ . Since

$$\bar{\phi}_{F_n}(t) = \lim_{\alpha \uparrow 1} \bar{\phi}_{n,\alpha}(t),$$

it is sufficient to prove

$$(A.21) \quad \sup_{t \in [0, 1], \alpha \in (0, 1)} |\bar{\phi}_{n,\alpha}(t) - \bar{\phi}_0(t)| \leq cn^{-1/12},$$

for some  $c > 0$ .

Let  $G_0$  be the inverse of  $F_0$  and  $G_{n,\alpha}$  be the inverse of  $F_{n,\alpha}$ . Furthermore, let

$$\bar{d}_{n,\alpha}(t) = t(1 - t) / \{(1 - t)h_1(G_{n,\alpha}(t)) + th_2(G_{n,\alpha}(t))\},$$

and, similarly

$$\bar{d}_0(t) = t(1 - t) / \{(1 - t)h_1(G_0(t)) + th_2(G_0(t))\}.$$

Finally, let  $\bar{H}_{n,\alpha}(t, u) = H(G_{n,\alpha}(t), G_{n,\alpha}(u))$ ,  $\bar{H}_0(t, u) = H(G_0(t), G_0(u))$ ,  $\bar{k}_{n,\alpha} = k \circ G_{n,\alpha}$  and  $\bar{k}_0 = k \circ G_0$ . Note that

$$(A.22) \quad \sup_{t \in [0, 1], \alpha \in (0, 1)} |G_{n,\alpha}(t) - G_0(t)| \leq kn^{-1/6},$$

for some  $k > 0$ , which follows from Lemma 5.1 and from

$$G_{n,\alpha}(t) = \inf\{x: F_{n,\alpha}(x) > t\} \leq \inf\{x: F_0(x - cn^{-1/6}) > t\},$$

and

$$G_{n,\alpha}(t) = \inf\{x: F_{n,\alpha}(x) > t\} \geq \inf\{x: F_0(x + cn^{-1/6}) > t\},$$

for some  $c > 0$  and all  $\alpha \in [0, 1]$ . We now write

$$\begin{aligned}
 & \bar{\phi}_{n,\alpha}(t) - \bar{\phi}_0(t) \\
 &= -\bar{d}_{n,\alpha}(t) \\
 \text{(A.23)} \quad & \times \left\{ \int_{t' \in [0,t]} \frac{\bar{\phi}_{n,\alpha}(t) - \bar{\phi}_{n,\alpha}(t') - (\bar{\phi}_0(t) - \bar{\phi}_0(t'))}{t - t'} d\bar{H}_{n,\alpha}(t', t) \right. \\
 & \quad \left. - \int_{u \in (t,1]} \frac{\bar{\phi}_{n,\alpha}(u) - \bar{\phi}_{n,\alpha}(t) - (\bar{\phi}_0(u) - \bar{\phi}_0(t))}{u - t} d\bar{H}_{n,\alpha}(t, u) \right\} \\
 & + B_{n,\alpha}(t) + C_{n,\alpha}(t) + D_{n,\alpha}(t) + E_{n,\alpha}(t),
 \end{aligned}$$

where

$$\text{(A.24)} \quad B_{n,\alpha}(t) = -\bar{d}_{n,\alpha}(t) \int_{t' \in [0,t]} \frac{\bar{\phi}_0(t) - \bar{\phi}_0(t')}{t - t'} d(\bar{H}_{n,\alpha} - \bar{H}_0)(t', t),$$

$$\text{(A.25)} \quad C_{n,\alpha}(t) = -\bar{d}_{n,\alpha}(t) \int_{u \in (t,1]} \frac{\bar{\phi}_0(u) - \bar{\phi}_0(t)}{u - t} d(\bar{H}_{n,\alpha} - \bar{H}_0)(t, u)$$

$$\begin{aligned}
 \text{(A.26)} \quad & D_{n,\alpha}(t) = -\{\bar{d}_{n,\alpha}(t) - \bar{d}_0(t)\} \\
 & \times \left\{ \int_{t' \in [0,t]} \frac{\bar{\phi}_0(t) - \bar{\phi}_0(t')}{t - t'} d\bar{H}_0(t', t) \right. \\
 & \quad \left. - \int_{u \in (t,1]} \frac{\bar{\phi}_0(u) - \bar{\phi}_0(t)}{u - t} d\bar{H}_0(t, u) \right\}
 \end{aligned}$$

and

$$\text{(A.27)} \quad E_{n,\alpha}(t) = \bar{d}_{n,\alpha}(t) \bar{k}_{n,\alpha}(t) - \bar{d}_0(t) \bar{k}_0(t).$$

We first consider  $B_{n,\alpha}(t)$ . Let  $t \in [0, 1]$  be a point such that  $t > n^{-1/12}$ . Moreover, let  $t_n = t - n^{-1/12}$ . On  $[0, t_n]$  we take a grid of points  $u_j, j = 0, \dots, m$ , such that

$$\frac{1}{2}n^{-1/6} \leq u_j - u_{j-1} \leq n^{-1/6}, \quad j \geq 1$$

and  $u_0 = 0, u_m = t_n$ . Furthermore, we define the right-continuous piecewise constant function  $\chi_n$  by

$$\chi_n(t') = \frac{\bar{\phi}_0(t) - \bar{\phi}_0(u_j)}{t - u_j}, \quad t' \in [u_j, u_{j+1}), j < m$$

and by  $\chi_n(u) = 0$  on the remaining part of the interval. Then we get, by the Lipschitz property of  $\bar{\phi}_0$  and the fact that  $t - t' > n^{-1/12}$  if  $t' < t_n$ ,

$$\begin{aligned}
 \text{(A.28)} \quad & \int_{t' \in [0,t_n]} \left| \frac{\bar{\phi}_0(t) - \bar{\phi}_0(t')}{t - t'} - \chi_n(t') \right| d(\bar{H}_{n,\alpha} + \bar{H}_0)(t', t) \\
 & = \mathcal{O}(n^{-1/12}).
 \end{aligned}$$

Moreover, using  $\sup_t |G_{n,\alpha}(t) - G_0(t)| = \mathcal{O}(n^{-1/6})$ , the Lipschitz property of  $\bar{\phi}_0$  and the differentiability properties of  $H$ , we get

$$\begin{aligned}
 & \left| \int_{t' \in [0, t_n)} \chi_n(t') d(\bar{H}_{n,\alpha} - \bar{H}_0)(t', t) \right| \\
 (A.29) \quad & \leq \left| \chi_n(t_n) \{ \bar{H}_{n,\alpha}(t_n, t) - \bar{H}_0(t_n, t) \} \right| \\
 & \quad + \left| \int_{t' \in [0, t_n)} | \bar{H}_{n,\alpha}(t', t) - \bar{H}_0(t', t) | d\chi_n(t') \right| \\
 & = \mathcal{O}(n^{-1/6}) + n^{1/12} \mathcal{O}(n^{-1/6}) = \mathcal{O}(n^{-1/12}).
 \end{aligned}$$

We also have

$$(A.30) \quad \int_{t' \in [t_n, t)} \left| \frac{\bar{\phi}_0(t) - \bar{\phi}_0(t')}{t - t'} \right| d(\bar{H}_{n,\alpha} + \bar{H}_0)(t', t) = \mathcal{O}(n^{-1/12}),$$

by the uniform boundedness of  $\{\bar{\phi}_0(t) - \bar{\phi}_0(t')\}/(t - t')$ ,  $t > t'$ , the differentiability properties of  $H$ , and by  $\sup_t |G_{n,\alpha}(t) - G_0(t)| = \mathcal{O}(n^{-1/6})$ .

Since  $\bar{d}_{n,\alpha}$  is uniformly bounded, we now get from (A.28) to (A.30):

$$|B_{n,\alpha}(t)| = \mathcal{O}(n^{-1/12}).$$

Moreover, since  $\sup_t |G_{n,\alpha}(t) - G_0(t)| = \mathcal{O}(n^{-1/6})$ , uniformly in  $\alpha$ , the  $\mathcal{O}(n^{-1/12})$ -term is uniform in  $\alpha$  and  $t \in [0, 1]$ . Note that for  $t \in [0, n^{-1/12}]$ , we only have to use the uniform boundedness of  $\{\bar{\phi}_0(t) - \bar{\phi}_0(t')\}/(t - t')$ ,  $t > t'$ , and the property  $\bar{H}_{n,\alpha}(t', t) + \bar{H}_0(t', t) = \mathcal{O}(n^{-1/12})$ .

We similarly get

$$\sup_{t \in [0, 1], \alpha \in (0, 1)} |C_{n,\alpha}(t)| = \mathcal{O}(n^{-1/12}).$$

Since the integrals on the right-hand side of (A.26) are uniformly bounded, we have, using (A.22) and the differentiability properties of the marginal densities  $h_1$  and  $h_2$ ,

$$\sup_{t \in [0, 1], \alpha \in (0, 1)} |D_{n,\alpha}(t)| = \mathcal{O}(n^{-1/6}).$$

Using the differentiability condition (F3) for  $k$ , we similarly get

$$\sup_{t \in [0, 1], \alpha \in (0, 1)} |E_{n,\alpha}(t)| = \mathcal{O}(n^{-1/6}).$$

Thus,

$$\begin{aligned}
 (A.31) \quad & \sup_{t \in [0, 1], \alpha \in (0, 1)} |B_{n,\alpha}(t) + C_{n,\alpha}(t) + D_{n,\alpha}(t) + E_{n,\alpha}(t)| \\
 & = \mathcal{O}(n^{-1/12}).
 \end{aligned}$$

To complete the proof, we use the same argument as has been used several times before, notably in Section 2: let  $s$  be the point where the continuous function

$$(A.32) \quad t \mapsto \bar{\phi}_{n,\alpha}(t) - \bar{\phi}_0(t), \quad t \in [0, 1],$$

attains its maximum. Then, by (A.23),

$$\bar{\phi}_{n,\alpha}(s) - \bar{\phi}_0(s) \leq B_{n,\alpha}(s) + C_{n,\alpha}(s) + D_{n,\alpha}(s) + E_{n,\alpha}(s),$$

since the integrands on the right-hand side of (A.23) are nonpositive if  $s$  is the (respectively right or left) boundary point of the integration domain. Similarly, if  $m$  is the point where the function (A.32) attains its minimum, we get

$$\bar{\phi}_{n,\alpha}(m) - \bar{\phi}_0(m) \geq B_{n,\alpha}(m) + C_{n,\alpha}(m) + D_{n,\alpha}(m) + E_{n,\alpha}(m).$$

It follows that

$$(A.33) \quad \sup_{t \in [0, 1], \alpha \in (0, 1)} |\bar{\phi}_{n,\alpha}(t) - \bar{\phi}_0(t)| \\ \leq \sup_{t \in [0, 1], \alpha \in (0, 1)} |B_{n,\alpha}(t) + C_{n,\alpha}(t) + D_{n,\alpha}(t) + E_{n,\alpha}(t)|.$$

The result (A.21) now follows from (A.31) and (A.33), and the theorem then follows from (A.21).  $\square$

**Acknowledgment.** We thank Jon Wellner for many helpful discussions.

## REFERENCES

- AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.
- BALL, K. and PAJOR, A. (1990). The entropy of convex bodies with “few” extreme points. In *Geometry of Banach Spaces* 25–32. Cambridge Univ. Press.
- BERMAN, A. and PLEMMONS, R. J. (1979). *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York.
- BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.
- BIRMAN, M. Š. and SOLOMJAK, A. (1967). Piecewise-polynomial approximations of functions of the classes  $W_p^\alpha$ . *Mat. Sb.* **73** 295–317.
- GESKUS, R. B. (1992). Efficient estimation of the mean for interval censoring case II. Technical Report 92-83, Delft Univ. Technology, <ftp://ftp.twi.tudelft.nl/pub/publications/tech-reports/1992>.
- GESKUS, R. B. (1997). Estimation of smooth functionals with interval censored data, and something completely different. Ph.D. dissertation, Delft Univ. Technology.
- GESKUS, R. B. and GROENEBOOM, P. (1995). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2 and beyond. Technical Report 95-78, Delft Univ. Technology, <ftp://ftp.twi.tudelft.nl/pub/publications/tech-reports/1995>.
- GESKUS, R. B. and GROENEBOOM, P. (1996). Asymptotically optimal estimation of smooth functionals for interval censoring, part 1. *Statist. Neerlandica* **50** 69–88.
- GESKUS, R. B. and GROENEBOOM, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2. *Statist. Neerlandica* **51** 201–219.

- GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser, Boston.
- GROENEBOOM, P. (1996). Lectures on inverse problems. *Lectures on Probability Theory. Ecole d'Été de Probabilités de Saint-Flour XXIV*. Springer, Berlin.
- HUANG, J. and WELLNER, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Statist. Neerlandica* **49** 153–163. <http://www.stat.washington.edu:80/jaw/jaw.research.available.html>.
- JONGBLOED, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7** 310–321. [http://www.amstat.org/publications/jcgs/toc\\_98.html](http://www.amstat.org/publications/jcgs/toc_98.html).
- KRESS, R. (1989). *Linear Integral Equations*. Springer, New York.
- POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- VAN DE GEER, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. *J. Nonparametr. Statist.* **6** 293–310.
- VAN DER VAART, A. W. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204.
- VAN DER VAART, A. W. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. *Indag. Math.* **18** 444–455.
- WELLNER, J. (1995). Interval censoring case 2: alternative hypotheses. In *Analysis of Censored Data* (H. L. Koul and J. V. Deshpande, eds.) 271–291. IMS, Hayward, CA. <http://www.stat.washington.edu:80/jaw/jaw.research.available.html>.

MUNICIPAL HEALTH SERVICE  
DEPARTMENT OF PUBLIC HEALTH  
AND ENVIRONMENT  
NIEUWE ACHTERGRACHT 100  
1018 WT AMSTERDAM  
THE NETHERLANDS  
E-MAIL: rgeskus@gggd.amsterdam.nl

FACULTY OF INFORMATION TECHNOLOGY  
AND SYSTEMS  
DELFT UNIVERSITY OF TECHNOLOGY  
MEKELWEG 4  
2628 CD DELFT  
THE NETHERLANDS  
E-MAIL: p.groeneboom@twi.tudelft.nl