Statistica Neerlandica (1995) Vol. 49, nr. 2, pp. 153–163

Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1

J. Huang and J. A. Wellner[1]

Department of Statistics GN-22, University of Washington, Seattle, Washington 98195

We give a new proof of the asymptotic normality of a class of linear functionals of the nonparametric maximum likelihood estimator (NPMLE) of a distribution function with "case 1" interval censored data. In particular our proof simplifies the proof of asymptotic normality of the mean given in Groeneboom and Wellner (1992). The proof relies strongly on a rate of convergence result due to van de Geer (1993), and methods from empirical process theory.

Key Words & Phrases: asymptotic distribution, empirical processes, linear functionals, mean moments, nonparametric maximum likelihood.

1 Introduction

Suppose that $T \sim F_0$, and $Y \sim G$, T and Y are independent. Here T is the variable of main interest, Y is a censoring variable. Suppose that the only observable variables are

$$X = (Y, \delta)$$

where $\delta = 1_{[T \leqslant Y]}$. Let $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ be i.i.d. random variables with the same distribution as (Y, δ). The goal is to estimate F_0 or functionals of F_0, such as the mean.

It is shown in GROENEBOOM and WELLNER (1992) (hereafter referred to as "GW") that, although the Nonparametric Maximum Likelihood Estimator (NPMLE) $\hat{F}_n$ of F_0 only has $n^{1/3}$-convergence rate, the mean of F_0 can be estimated at $\sqrt{n}$ rate. Moreover, they show that the plug-in maximum likelihood estimator achieves the asymptotic efficiency bound. Groeneboom's proof given there (see pages 114–120) depends on the uniform convergence rate of the NPMLE $\hat{F}_n$, which is derived via a delicate exponential martingale argument. The main purpose here is to give a different proof, which only uses the L_2 or the Hellinger convergence rate of $\hat{F}_n$ established by VAN DE GEER (1983).

[1]Research supported in part by National Science Foundation grant DMS-9108409, NIAID grant 2R01 AI291968–04, NATO grant NWO Grant B61–238, and by the Stieljes Institute.

We begin with a discussion of the NPMLE $\hat{F}_n$ of F for the interval censoring model, including characterizations and computation thereof, and derive a key identity needed for our asymptotic normality proof. Section 3 contains a brief review of information bound theory for smooth functionals such as those which concern us here. This section also contains a discussion of the relationship of our approach with recent work of VAN DER LAAN (1993a), (1993b), (1994). Section 4 contains a restatement of the useful rate of convergence result obtained by VAN DE GEER (1993). The main theorem is stated and proved in section 5. Our hope is that the methods developed here may also be of some help in treating the estimation of smooth functionals in the much more difficult problem of interval centored data, case 2; see e.g. GESKUS (1992) and GESKUS and GROENEBOOM (1994).

2 Characterization, self-consistency and computation of the NPMLE

The NPMLE $\hat{F}_n$ of F_0 is the distribution function that maximizes

$$l_n(F) = \sum_{i=1}^{n} \{\delta_i \log F(Y_i) + (1-\delta_i) \log (1 - F(Y_i))\};$$

thus the maximization is carried out under a monotonicity constraint. $\hat{F}_n$ can be explicitly expressed via a min-max formula, or can be characterized via Fenchel's duality theorem and hence as the slope of the convex minorant of a particular sum diagram. In the following, we first establish an identity (3) based on the characterization of $\hat{F}_n$. This identity is crucial to our proof of the asymptotic normality of the NPMLE of the mean. Then we show that the self-consistency equation follows from the characterization of $\hat{F}_n$. For a thorough treatment of the characterization of $\hat{F}_n$ and its asymptotic properties, see GW (1992), part II, chapter 1, pages 35–52.

Let us relabel the data $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ in terms of the ordered values of $Y_1, \ldots, Y_n$ as $(Y_{(1)}, \delta_{(1)}), \ldots, (Y_{(n)}, \delta_{(n)})$, where $Y_{(1)} \leqslant \cdots \leqslant Y_{(n)}$. Write

$$l_n(F) = \sum_{i=1}^{n} \{\delta_{(i)} \log F(Y_{(i)}) + (1-\delta_{(i)}) \log (1 - F(Y_{(i)}))\}. \tag{1}$$

It is clear that $\hat{F}_n$ is determined only up to its values at the censoring times $Y_{(1)}, \ldots, Y_{(n)}$. We will take the right continuous step function with possible jumps at Y_i's as the NPMLE. First suppose that $\delta_{(1)} = 1$ and $\delta_{(n)} = 0$. Then to maximize (1), we must have $0 < \hat{F}_n(Y_{(1)}) \leqslant \hat{F}_n(Y_{(n)}) < 1$. Let $Y_{(1)} = \tau_0 < \tau_1 < \cdots < \tau_m < \tau_{m+1} = Y_{(n)}$ be the jump points of $\hat{F}_n$, that is,

$$\hat{F}_n(s) = \begin{cases} 0 & \text{if } s < \tau_0 \\ \hat{F}_{(n)}(\tau_j) & \text{if } \tau_j \leqslant s < \tau_{j+1}, \quad j = 0, 1, \ldots, m. \end{cases}$$

Notice that in this case, $\hat{F}_n$ is actually a subdistribution function. The likelihood does not tell us how to locate the remaining mass beyond $Y_{(n)}$. We will leave it unspecified.

From Propositions 1.1 and 1.2 of GW (1992), for $j = 0, 1, \ldots, m$, Fenchel duality implies that $x = \hat{F}_n(\tau_j)$ solves the "score" equation

$$\sum_{i:\tau_j \leqslant Y_{(i)} < \tau_{j+1}} \left\{ \frac{\delta_{(i)}}{x} - \frac{1-\delta_{(i)}}{1-x} \right\} = 0; \tag{2}$$

see GW (1992), pages 41–43 and especially (1.19), page 43. This implies

$$\sum_{i:\tau_j \leqslant Y_{(i)} < \tau_{j+1}} \{\delta_{(i)} - \hat{F}_n(Y_{(i)})\} = 0.$$

This in turn implies for any function γ,

$$\sum_{i:\tau_j \leqslant Y_{(i)} < \tau_{j+1}} \{\delta_{(i)} - \hat{F}_n(Y_{(i)})\}\gamma(\hat{F}_n(Y_{(i)})) = \gamma(\hat{F}_n(\tau_j)) \sum_{i:\tau_j \leqslant Y_{(i)} < \tau_{j+1}} \{\delta_{(i)} - \hat{F}_n(Y_{(i)})\} = 0.$$

Summing up the above equations over $j = 0, 1, \ldots, m$, we obtain

$$\sum_{i=1}^{n} \{(\delta_{(i)} - \hat{F}_n(Y_{(i)}))\gamma(\hat{F}_n(Y_{(i)}))\} = 0, \tag{3}$$

or, equivalently,

$$\int \{\delta - \hat{F}_n(y)\}\gamma(\hat{F}_n(y))\, \mathrm{d}\mathbb{P}_n(y, \delta) = 0 \tag{4}$$

where $\mathbb{P}_n$ is the empirical measure of the observed (Y_i, δ_i), $i = 1, \ldots, n$.

If for some $1 \leqslant k < n$, $\delta_{(1)} = \cdots = \delta_{(k)} = 0$ and $\delta_{(k+1)} = 1$, then to maximize (1) without violating the monotonicity constraint, we must have $\hat{F}_n(Y_{(1)}) = \cdots = \hat{F}_n(Y_{(k)}) = 0$ and $\hat{F}_n(Y_{(k+1)}) > 0$. In this case, define $\tau_0 = Y_{(k+1)}$. If $\delta_{(j)} = 0$ and $\delta_{(j+1)} = \cdots = \delta_{(n)} = 1$, we must have $\hat{F}_n(Y_{(j+1)}) = \cdots = \hat{F}_n(Y_{(n)}) = 1$ and $\hat{F}_n(Y_{(j)}) < 1$. In this case, define $\tau_m = Y_{(j)}$ and $\tau_{(m+1)} = Y_{(j+1)}$. It is clear that in either or both of these two cases, (3) and (4) still hold.

We remark here that equation (2) is useless in computing $\hat{F}_n$, because we do not know the jump points in the first place. $\hat{F}_n$ can actually be explicitly expressed by the "max-min" formula, i.e.,

$$\hat{F}_n(Y_{(i)}) = \max_{s \leqslant i} \min_{t \geqslant i} \frac{\sum_{s \leqslant j \leqslant t} \delta_{(j)}}{t - s + 1}, \quad i = 1, \ldots, n.$$

Our experience is that this formula is quite fast in computing $\hat{F}_n$ with large sample sizes.

We now show that (4) also implies the self-consistency equations, which are

$$\hat{F}_n(s) = E_{\hat{F}_n}\{\mathbb{F}_n(s) \mid Y_1, \ldots, Y_n, \delta_1, \ldots, \delta_n\}, \quad 0 \leqslant s \leqslant M,$$

or

$$\hat{F}_n(s) = \frac{1}{n}\sum_{i=1}^{n} \left\{ \delta_i \frac{\hat{F}_n(Y_i \wedge s)}{\hat{F}_n(Y_i)} + (1-\delta_i)\frac{\hat{F}_n(s) - \hat{F}_n(Y_i \wedge s)}{1 - \hat{F}_n(Y_i)} \right\}, \quad 0 \leqslant s \leqslant M \tag{5}$$

where $\mathbb{F}_n$ is the empirical distribution function of the unobservable random variables $T_1, \ldots, T_n$. See GW (1992), equation (1.6), page 38. Thus $\hat{F}_n$ is the conditional expectation of the empirical distribution function $\mathbb{F}_n$ given the observable data (Y_i, δ_i), $i = 1, \ldots, n$ under the self-induced probability measure $P_{\hat{F}_n}$. Starting from an initial value for $\hat{F}_n$ and updating it via (2.5) yields the familiar EM algorithm. However, when the sample size is moderate or large, the EM algorithm is painfully slow for computing $\hat{F}_n$, and may converge to solutions of the self-consistency equation other than the NPMLE.

We conclude this section by showing that the self-consistency equation (5) is satisfied by the NPMLE $\hat{F}_n$ as a consequence of its characterization via Fenchel duality and the resulting identity (3). Equation (5) can be rewritten as

$$\sum_{i=1}^{n}\left\{\left[\frac{\delta_i}{\hat{F}_n(Y_i)} - \frac{1-\delta_i}{1-\hat{F}_n(Y_i)}\right](\hat{F}_n(Y_i \wedge s) - \hat{F}_n(Y_i)\hat{F}_n(s))\right\} = 0,$$

or, equivalently, since $\hat{F}_n(Y_i \wedge s) = \hat{F}_n(Y_i) \wedge \hat{F}_n(s)$, as

$$\sum_{i=1}^{n}\left\{(\delta_i - \hat{F}_n(Y_i))\frac{(\hat{F}_n(Y_i) \wedge \hat{F}_n(s) - \hat{F}_n(Y_i)\hat{F}_n(s))}{\hat{F}_n(Y_i)(1-\hat{F}_n(Y_i))}\right\} = 0.$$

But this follows from (3) by choosing the function γ to be

$$\gamma(u) = \frac{u \wedge \hat{F}_n(s) - u\hat{F}_n(s)}{u(1-u)}.$$

3 Review of information bound calculations

Now let ψ be a fixed measurable function, and, for distribution functions F on $I \equiv [0, M]$, consider functionals of the form

$$\nu(F) = \int_I (1 - F(x))\psi(x)\,\mathrm{d}x. \tag{6}$$

By Fubini's theorem we have

$$\nu(F) = \int_I \Psi(x)\,\mathrm{d}F(x) \tag{7}$$

where $\Psi(x) \equiv \int_0^x \psi(t)\,\mathrm{d}t$, and hence ν is a linear functional. Such a functional $\nu(F)$ defined in terms of the unknown distribution function F is an "implicitly defined" functional since the observations $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ are from $P_{F,G} = \mathscr{L}(Y, \delta)$ rather than directly from F. If G has density g, then it is easily seen that $P_{F,G}$ has density

$$p_{F,G}(y, \delta) = F(y)^{\delta}(1 - F(y))^{1-\delta}g(y), \quad 0 \leqslant y \leqslant M, \delta \in \{0, 1\}.$$

Information bounds for implicitly defined functionals such as ν were nicely treated by VAN DER VAART (1991); for further discussions of van der Vaart's theorem, see GW

(1992), part I, section 3, pages 23–32, and BICKEL, KLAASSEN, RITOV, and WELLNER (1993), section 5.4, pages 201–210. In fact, the mean functional $\nu(F)$ corresponding to the choice $\psi \equiv 1$ was treated in all three of these references: see VAN DER VAART (1991), pages 196–197; GW (1992), pages 28–30; and BKRW (1993), pages 207–210. Extending those calculations to the case of a general function ψ is straightforward and the result is as follows: if I_κ^{-1} defined in (9) is finite, then $\kappa(P_{F,G}) \equiv \nu(F)$ is pathwise differentiable and the efficient influence function is

$$\tilde{l}_\kappa(y, \delta) \equiv \tilde{l}_\kappa(y, \delta; F_0, G) = -\{\delta - F_0(y)\} \frac{\psi(y)}{g(y)} 1_{[g(y) > 0]}, \tag{8}$$

and hence the information bound for estimation of $\kappa(P_{F_0,G}) = \nu(F_0)$ is

$$I_\kappa^{-1} = E\tilde{l}_\kappa^2(Y, \delta) = \int_0^M \frac{F_0(y)(1 - F_0(y))}{g(y)} \psi^2(y)\, dy. \tag{9}$$

Furthermore, general theory suggests that (under modest regularity conditions) an efficient estimator will be asymptotically linear with efficient influence function $\tilde{l}_\kappa$; see e.g. BKRW (1993), theorem 3.3.2(B), pages 63–64 and theorem 5.2.3, page 183. Thus we expect to be able to show that

$$\sqrt{n}(\nu(\hat{F}_n) - \nu(F_0)) = \sqrt{n}(\mathbb{P}_n - P)(\tilde{l}_\kappa(\cdot; F_0, G)) + o_p(1). \tag{10}$$

with $\tilde{l}_\kappa$ as given in (8). This is one of the principal motivations in the proof of our main theorem in section 5.

Our proofs in section 5 are also somewhat related to the recent work of VAN DER LAAN (1993a), (1993b), (1994). Suppose that:

(i) Θ is convex.
(ii) $\theta \longrightarrow P_\theta$ is linear.
(iii) $\theta \longrightarrow \nu(\theta)$ is linear.
(iv) $\hat{\theta}$ satisfies $\mathbb{P}_n \tilde{l}_\kappa(\cdot; \hat{\theta}_n) = 0$ where $\tilde{l}_\kappa$ is the efficient influence function for estimation of $\kappa(P_\theta) = \nu(\theta)$.

Then VAN DER LAAN (1993a, b) shows that

$$\nu(\hat{\theta}_n) - \nu(\theta) = (\mathbb{P}_n - P_\theta)(\tilde{l}_\kappa(\cdot; \hat{\theta}_n)). \tag{11}$$

Of course (11) implies that

$$\sqrt{n}(\nu(\hat{\theta}_n) - \nu(\theta)) = \sqrt{n}(\mathbb{P}_n - P_\theta)(\tilde{l}_\kappa(\cdot; \theta)) + \sqrt{n}(\mathbb{P}_n - P_\theta)(\tilde{l}_\kappa(\cdot; \hat{\theta}_n) - \tilde{l}_\kappa(\cdot; \theta))$$

where we expect the second term to be $o_p(1)$ under some modest continuity assumptions on the map $\theta \rightarrow \tilde{l}_\kappa(\cdot; \theta)$. Extensions of the identity (11) to problems involving a nuisance parameter are given in VAN DER LAAN (1994). In the present interval censoring model, the distribution function G is a nuisance parameter for estimation of F, and the efficient influence function $\tilde{l}_\kappa$ for estimation of $\kappa(P_{F,G}) = \nu(F)$

depends on the density $g = G'$ of G. It turns out in our case that one ingredient of van der Laan's identity does hold; namely

$$\nu(\hat{F}_n) - \nu(F_0) = -P_{F_0,G}\tilde{l}_\kappa(\cdot\,; \hat{F}_n, G)$$

does hold, as may be verified directly. However, the other key ingredient, the analogue of (iv) above, *does not hold*: that is,

$$\mathbb{P}_n\tilde{l}_\kappa(\cdot\,; \hat{F}_n, G) = \int \frac{\delta - \hat{F}_n(y)}{g(y)}\psi(y)\,\mathrm{d}\mathbb{P}_n(y, \delta) \neq 0. \tag{12}$$

One part of our proof in section 5 can be viewed as using the key identity (4) together with empirical process theory to show that

$$\mathbb{P}_n\tilde{l}_\kappa(\cdot\,; \hat{F}_n, G) = o_p(n^{-1/2}).$$

In fact, this is the hardest part of our proof, as it is the hard part of a corresponding step in the proof in HUANG (1994b), where a similar argument is used.

4 A convergence rate result

As demonstrated in WONG and SEVERINI (1991), BIRGÉ and MASSART (1993), VAN DE GEER (1993), and VAN DER VAART and WELLNER (1994), the convergence rate of a NPMLE in the L_2 or Hellinger distance is determined by the smoothness of the model and the entropy of the collection of likelihood functions. In many nonparametric estimation problems, methods are available for determining the convergence rate of the NPMLE (assuming that it exists and is unique). The interval censoring problem considered here is one of the examples considered by VAN DE GEER (1993); in example 4.8(a) she proved the following $n^{1/3}$-convergence rate of $\hat{F}_n$ in terms of Hellinger distance.

LEMMA 4.1. (VAN DE GEER, *1993). If F_0 is a distribution function on $I = [0, M]$ with $M > 0$, then*

$$\|\sqrt{\hat{F}_n} - \sqrt{F_0}\|_{L_2(G)} = O_p(n^{-1/3}).$$

Since

$$\int (\hat{F}_n - F_0)^2\,\mathrm{d}G = \int (\sqrt{\hat{F}_n} - \sqrt{F_0})^2(\sqrt{\hat{F}_n} + \sqrt{F_0})^2\,\mathrm{d}G < 4\int (\sqrt{\hat{F}_n} - \sqrt{F_0})^2\,\mathrm{d}G,$$

it follows that

$$\|\hat{F}_n - F_0\|_{L_2(G)} = O_p(n^{-1/3}). \tag{13}$$

This holds without any assumptions on the relationship of the support of F to the support of the observation distribution G. If we assume that $F_0 \ll G$, then (13) implies pointwise convergence of $\hat{F}_n$ to F_0 on the support of G, and hence on the support of F_0. Since the functionals $\nu(F)$ of interest here involve $F_0(x)$ for all x in the support of F_0, the domination condition $F_0 \ll G$, will be needed in our asymptotic normality theorem in the next section.

5 Asymptotic normality of the NPMLE's of the functionals ν

Consider estimation of the linear functional $\nu(F)$ as given in (6) and (7). A natural estimator of $\nu(F_0)$ is just $\nu(\hat{F}_n)$. Here is our main asymptotic normality result for $\nu(\hat{F}_n)$.

THEOREM 5.1. *Suppose that:*

(*i*) *The support of* F_0 *is a bounded interval* $I = [0, M]$, $G \ll F_0$, $F_0 \ll G$, *and* G *has density* g *with respect to Lebesgue measure,*

(*ii*) F_0, g, *and* ψ *satisfy*

$$I_\kappa^{-1} \equiv I_\kappa^{-1}(F_0, g, \psi) \equiv \int_I \frac{F_0(y)(1-F_0(y))}{g(y)} \psi^2(y)\,dy < \infty.$$

(*iii*) $(\psi/g) \circ F_0^{-1}$ *is bounded and is a Lipschitz function on* $[0, 1]$. *Then*

$$\sqrt{n}(\nu(\hat{F}_n) - \nu(F_0)) \to_d N(0, I_\kappa^{-1}).$$

Perhaps the most important special cases of the theorem are the moments of F: with $\psi(x) = rx^{r-1}$ for $r > 0$,

$$\nu(F) = \int_0^M rx^{r-1}(1 - F(x))\,dx = E_F X^r \equiv \nu_r(F).$$

Of course the mean is just the further special case $r = 1$.

COROLLARY 5.1. *If the hypotheses of theorem 5.1 are satisfied with* $\psi(x) = rx^{r-1}$, *then*

$$\sqrt{n}(\nu_r(\hat{F}_n) - \nu_r(F_0)) \to_d N(0, I_r^{-1})$$

where

$$I_r^{-1} \equiv r^2 \int_I \frac{F_0(y)(1-F_0(y))}{g(y)} y^{2(r-1)}\,dy < \infty.$$

PROOF. Letting $P \equiv P_{F_0,G} = \mathscr{L}(Y, \delta)$, write

$$\sqrt{n}(\nu(\hat{F}_n) - \nu(F_0)) = \sqrt{n} \int_I [(1 - \hat{F}_n(y)) - (1 - F_0(y))]\psi(y)\,dy$$

$$= \sqrt{n} \int \frac{1 - \hat{F}_n(y) - (1 - \delta)}{g(y)} \psi(y)\,dP(y, \delta)$$

$$= \sqrt{n} \int \frac{\delta - \hat{F}_n(y)}{g(F_0^{-1}(\hat{F}_n(y)))} \psi(F_0^{-1}(\hat{F}_n(y)))\, \mathrm{d}P(y, \delta)$$

$$+ \sqrt{n} \int (\delta - \hat{F}_n(y)) \left(\frac{\psi(y)}{g(y)} - \frac{\psi(F_0^{-1}(\hat{F}_n(y)))}{g(F_0^{-1}(\hat{F}_n(y)))} \right) \mathrm{d}P(y, \delta)$$

$$= \sqrt{n} \int \frac{\delta - \hat{F}_n(y)}{g(F_0^{-1}(\hat{F}_n(y)))} \psi(F_0^{-1}(\hat{F}_n(y)))\, \mathrm{d}P(y, \delta)$$

$$+ \sqrt{n} \int (F_0(y) - \hat{F}_n(y)) \left(\frac{\psi(y)}{g(y)} - \frac{\psi(F_0^{-1}(\hat{F}_n(y)))}{g(F_0^{-1}(\hat{F}_n(y)))} \right) \mathrm{d}G(y)$$

$$\equiv -\Delta_{1n} + \Delta_{2n}$$

We prove that

$$\Delta_{1n} = \sqrt{n} \int \frac{\delta - F_0(y)}{g(y)} \psi(y)\, \mathrm{d}(\mathbb{P}_n - P)(y, \delta) + o_p(1) \tag{14}$$

$$= -\sqrt{n}(\mathbb{P}_n - P)(\tilde{l}_\kappa) + o_p(1) \tag{15}$$

and $\Delta_{2n} = o_p(1)$. Then the result follows from the central limit theorem.

Let F_0^{-1} be the left-continuous inverse of F_0: $F_0^{-1}(u) \equiv \inf\{x : F_0(x) \geqslant u\}$. Then define $\gamma_0 = (\psi/g) \circ F_0^{-1}$, the composition of ψ divided by g with the inverse F_0^{-1} of F_0. By the key identity (4) established in section 2 we have

$$\int (\delta - \hat{F}_n(y)) \gamma_0(\hat{F}_n(y))\, \mathrm{d}\mathbb{P}_n(y, \delta) = 0.$$

Hence the first term

$$\Delta_{1n} = \sqrt{n} \int (\delta - \hat{F}_n(y)) \gamma_0(\hat{F}_n(y))\, \mathrm{d}(\mathbb{P}_n - P)(y, \delta)$$

$$= \sqrt{n} \int (\delta - F_0(y)) \gamma_0(F_0(y))\, \mathrm{d}(\mathbb{P}_n - P)(y, \delta)$$

$$+ \sqrt{n} \int (\delta - F_0(y))(\gamma_0(\hat{F}_n(y)) - \gamma_0(F_0(y)))\, \mathrm{d}(\mathbb{P}_n - P)(y, \delta)$$

$$- \sqrt{n} \int (\hat{F}_n(y) - F_0(y)) \gamma_0(\hat{F}_n(y))\, \mathrm{d}(\mathbb{P}_n - P)(y, \delta)$$

$$\equiv I_{1n} + I_{2n} + I_{3n}.$$

Let $M > 0$ be fixed, and let

$$\mathscr{F} = \{F : F \text{ is a distribution function on } [0, M]\}. \tag{16}$$

Consider the class of functions

$$\mathscr{H} = \{(F(y) - F_0(y)) \gamma_0(F(y)) : F \in \mathscr{F}\},$$

where $\mathscr{F}$ is defined in (16). First, the uniform covering entropy for $\mathscr{F}$ is bounded by $K(1/\epsilon)^{1+\tau}$ for every $\tau > 0$, this follows from DUDLEY (1987), theorem 5.1, page 1318, since $\mathscr{F}$ is contained in the convex hull of the VC graph class B of indicator function of right half lines with $D(\epsilon, B) < K\epsilon^{-2}$; see also DUDLEY (1987), example 5.9, page 1321. Although we did not need the refinement here, this bound actually remains true with $\tau = 0$ in view of a result of BALL and PAJOR (1990). This corresponds to the bracketing entropy bound for uniformly bounded monotone functions obtained by VAN DE GEER (1991); see VAN DE GEER (1993), corollary 2.3, page 19, for a statement. Moreover, by assumption (iii) γ_0 is bounded and Lipschitz on [0, 1]: in particular, there exists a finite constant K such that

$$|\gamma_0(u_1) - \gamma_0(u_2)| \leqslant K|u_1 - u_2|$$

for all $u_1, u_2 \in [0, 1]$. Therefore, for any $F_1, F_2 \in \mathscr{F}$,

$$|(F_1(y) - F_0(y))\gamma_0(F_1(y)) - (F_2(y) - F_0(y))\gamma_0(F_2(y))| \leqslant C|F_1(y) - F_2(y)|$$

for some constant C. It follows that the uniform entropy for $\mathscr{H}$ is also bounded by $K(1/\epsilon)^{1+\tau}$, $\tau > 0$. So $\mathscr{H}$ is a P-Donsker class by Pollard's theorem (POLLARD, 1982; see DUDLEY, 1984, theorem 11.31, page 117 and DUDLEY, 1987). Furthermore, since

$$\int [\hat{F}_n(y) - F_0(y)]^2 \, dP(y, \delta) = \int [\hat{F}_n(y) - F_0(y)]^2 \, dG(y) \to_p 0,$$

it follows (from the fact that $\mathscr{H}$ satisfying the Donsker property implies uniform asymptotic equicontinuity of the empirical process over $\mathscr{H}$, see e.g. DUDLEY (1984), theorem 4.1.1, page 27; or SHEEHY and WELLNER (1992), theorem 1.1; or VAN DER VAART and WELLNER (1994), theorem 1.4.6) that, $I_{3n} = o_p(1)$. To show that $I_{2n} = o_p(1)$, consider the class of functions

$$\tilde{\mathscr{H}} = \{(\delta - F_0(y))(\gamma_0(F(y)) - \gamma_0(F_0(y))): F \in \mathscr{F}\},$$

where $\mathscr{F}$ is defined in (16). Now the uniform entropy for $\tilde{\mathscr{H}}$ is bounded by $K(1/\epsilon)^{1+\tau}$, again using the Lipschitz property of γ_0, and it follows that $\tilde{\mathscr{H}}$ is P-Donsker, and in view of the $L_2(P)$ consistency of $\hat{F}_n$, this implies that $I_{2n} = o_p(1)$. Thus (14) is proved.

To handle Δ_{2n}, recall that $F_0^{-1} \circ F_0(x) = x$ a.e. F_0 (see e.g. SCHORACK and WELLNER, 1986, proposition 1.1.3, page 6). Since $G \ll F_0$, this implies that $F_0^{-1} \circ F_0(x) = x$ a.e. G. Therefore it follows that

$$\begin{aligned}|\Delta_{2n}| &= \sqrt{n}\left|\int (\hat{F}_n(y) - F_0(y))(\gamma_0(F_0(y)) - \gamma_0(\hat{F}_n(y))) \, dG(y)\right| \\ &\leqslant \sqrt{n}\,\|\hat{F}_n - F_0\|_{L_2(G)} \times \|\gamma_0 \circ \hat{F}_n - \gamma_0 \circ F_0\|_{L_2(G)} \\ &\leqslant K\sqrt{n}\,\|\hat{F}_n - F_0\|^2_{L_2(G)} \\ &= O_p(n^{-1/6})\end{aligned}$$

by van de Geer's rate result lemma 4.1. □

REMARK. It is possible to consider a wider class of functionals $v(F)$ satisfying

$$v(F) - v(F_0) = \int_I \dot{v}_{F_0}(x)\, \mathrm{d}(F - F_0)(x) + O(\|F - F_0\|^2_{L_2(G)}) \tag{17}$$

for some function $\dot{v}_{F_0} \in L_0^2(F_0)$; for the v of (3.7) this holds with $\dot{v}_{F_0}(x) = \Psi(x) - \int_I \Psi \, \mathrm{d}F_0$ and no remainder (or O) term. Further hypotheses on $\dot{v}_F$ are needed to handle these functionals in general.

The methods used in the proof of theorem 2.1 have been used by HUANG (1994) to prove the asymptotic normality of the maximum likelihood estimator of the regression parameters θ in the case of the Cox proportional hazards model with interval censoring. Similar methods are likely to be useful in other related problems.

Acknowledgements

The authors owe thanks to Sara van de Geer, Piet Groeneboom, and Aad van der Vaart for helpful discussions about rates of convergence, interval censoring, and related problems. The result of the present paper was presented by the authors at a Workshop on Semiparametric models organized by the Dutch Statistical Society and held in Utrecht, The Netherlands, on December 14, 1993.

References

BALL, K. and A. PAJOR (1990), The entropy of convex bodies with "few" extreme points, in: P. F. X. MÜLLER and W. SICHACKERMAYER (eds), *Geometry of Banach Spaces*, Proceedings of the conference held in Strobl, Austria 1989, *London Mathematical Society Lecture Note Series 158*, 25–32.

BICKEL, P. J., C. A. J. KLAASSEN, Y. RITOV, and J. A. WELLNER (1993), *Efficient and adaptive estimation for semiparametric models*, Johns Hopkins University Press, Baltimore.

BIRGÉ, L. and P. MASSART (1993), Rates of convergence for minimum contrast estimators, *Probability Theory and Related Fields 97*, 113–150.

DUDLEY, R. M. (1984), A course on empirical processes, École d'Été de Probabilités de St. Flour, *Lecture Notes in Mathematics 1097*, Springer Verlag, New York, 2–142.

DUDLEY, R. M. (1987), Universal Donsker classes and metric entropy, *Annals of Probability 15*, 1306–1326.

GEER, S. VAN DE (1991), The entropy bound for monotone functions, Technical Report 91-10, University of Leiden.

GEER, S. VAN DE (1993), Hellinger-consistency of certain nonparametric maximum likelihood estimators, *Annals of Statistics 21*, 14–44.

GESKUS, R. (1992), Efficient estimation of the mean for interval censoring case II, Technical Report 92–83, Department of Mathematics, Delft University of Technology.

GESKUS, R. and P. GROENEBOOM (1994), Estimation of smooth functionals for interval censoring, case 2, Technical Report 94-274, Department of Mathematics, Delft University of Technology.

GROENEBOOM, P. and J. A. WELLNER (1992), *Information bounds and nonparametric maximum likelihood estimation*, DMV Seminar Band 19, Birkhäuser, Basel.

HUANG, J. (1994a), Estimation in regression models with interval censoring, Ph.D. dissertation, Department of Statistics, University of Washington.

HUANG, J. (1994b), Efficient estimation for the Cox model with interval censoring, Technical Report, Department of Statistics, University of Washington.

LAAN, M. J. VAN DER (1993a), Efficient and inefficient estimation in semiparametric models, Thesis, Department of Mathematics, Uterecht, The Netherlands.
LAAN, M. J. VAN DER (1993b), General identity for linear parameters in convex models with application to efficiency of the (NP)MLE, Preprint No. 765, Department of Mathematics, University of Utrecht, The Netherlands.
LAAN M. J. VAN DER (1994), Proving efficiency of NPMLE and identities, Technical Report No. 44 Group in Biostatics, University of California, Berkeley.
POLLARD, D. B. (1982), A central limit theorem for empirical processes, *Journal of the Australian Mathematical Society A 33*, 235–248.
SHEEHY, A. and J. A. WELLNER (1992), Uniform Donsker classes of functions, *Annals of Probability 20*, 1983–2030.
SHORACK, G. R. and J. A. WELLNER (1986), *Empirical processes with applications to statistics*, Wiley, New York.
VAART, A. W. VAN DER (1991), On differentiable functions, *Annals of Statistics 19*, 178–205.
VAART, A. W. VAN DER and J. A. WELLNER (1994), *Weak convergence and empirical processes*, To appear.
WONG, W. H. and T. A. SEVERINI (1991), On maximum likelihood estimation in infinite dimensional parameter spaces, *Annals of Statistics 19*, 603–632.

Received: February 1994. Revised: August 1994.