Analysis of Censored Data
IMS Lecture Notes – Monograph Series (1995) Volume 27

Interval Censoring, Case 2: Alternative Hypotheses

Jon A. Wellner [1]

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

"Interval censoring case 2" involves observation times (U, V) with distribution H concentrated on the set $u \leq v$ and a time of interest X with distribution F. The goal is to estimate F based only on observation of i.i.d. copies of $(1_{[X \leq U]}, 1_{[U < X \leq V]}, U, V)$. Groeneboom (1991) initiated the study of the nonparametric maximum likelihood estimator $\widehat{F}_n$ of F; see Groeneboom and Wellner (1992), especially pages 43 - 50 and 100-108. Geskus (1992) and Geskus and Groeneboom (1994) have studied the estimation of smooth functionals (such as the mean of F) in case 2. Under hypotheses ensuring that the observations times U and V are close with (sufficiently) positive probability, Groeneboom (1991) showed that a one-step approximation $F_n^{(1)}$ to the nonparametric MLE satisfies

$$(n \log n)^{1/3}(F_n^{(1)}(t_0) - F(t_0)) \rightarrow_d 2\{\frac{3}{4} f_0^2(t_0)/h(t_0, t_0)\}^{1/3} Z$$

where Z is the last time where standard two-sided Brownian motion W minus the parabola $y(t) = t^2$ reaches its maximum. While it is conjectured in Groeneboom and Wellner (1992) that the nonparametric MLE $\widehat{F}_n$ has this same behavior, this conjecture is still unproved.

The goal of this paper is to explore alternative hypotheses under which U and V are *not* close with high probability. Under these alternative hypotheses, the one-step approximation to the nonparametric MLE will be shown to converge at rate $n^{-1/3}$ rather than $(n \log n)^{-1/3}$, much as in interval censoring case 1 (current status data). We will also briefly discuss the behavior of the one-step NPMLE with $k > 2$ observation points and estimators of smooth functionals.

[1] Research supported in part by National Science Foundation grant DMS-9306809, NIAID grant 2R01 AI291968-04, NATO NWO Grant B 61-238.

AMS 1980 subject classifications. Primary: 60F17, 62E20; secondary 60B12.

Key words and phrases. asymptotic distribution, case 2, empirical processes, hypotheses, interval censored data, mean, nonparametric maximum likelihood.

1. Interval Censoring: Models and Estimators. We begin with a review of interval censoring models starting with "case 1" or "current status data".

Suppose that $X \sim F_0$ is a "time of interest", and that $U \sim H$ is an "observation time". We will assume that X and U are independent random variables. Unfortunately we do not observe (X, U) but just $(1_{[X \leq U]}, U) \equiv (\Delta, U)$. Thus

$$(\Delta | U = u) \sim \text{Bernoulli}(F_0(u))$$

and if H has density function h with respect to Lebesgue measure, then the joint density of (Δ, U) is

$$p(\delta, u; F_0) = F_0(u)^{\delta}(1 - F_0(u))^{1-\delta} h(u)$$

for $\delta \in \{0, 1\}$. The goal is to estimate the distribution function F_0, or functions of F_0 such as the mean, based on observation of a sample $(\Delta_1, U_1), \ldots, (\Delta_n, U_n)$ i.i.d. as (Δ, U).

Another commonly arising observaton scheme involves two observation times, and hence is called "case 2" interval censoring in Groeneboom (1990) and Groeneboom and Wellner (1992) (which we henceforth refer to as "GW (1992)"). Again $X \sim F_0$ is a "time of interest", but now suppose that $(U, V) \sim H$ is independent of X where $P_H(U \leq V) = 1$. In this case we observe not (X, U, V) but just

$$(1_{[X \leq U]}, 1_{[U < X \leq V]}, 1_{[V < X]}, U, V) \equiv (\underline{\Delta}, U, V).$$

Clearly

$$(\underline{\Delta} | U = u, V = v) \sim \text{Multinomial}_3(1, (F_0(u), F_0(v) - F_0(u), 1 - F_0(v)))$$

and if H has density h with respect to Lebesgue measure on R^2, then the joint density of $(\underline{\Delta}, U, V)$ is given by

$$p(\underline{\delta}, u, v; F_0) = F_0(u)^{\delta_1}(F_0(v) - F_0(u))^{\delta_2}(1 - F_0(v))^{\delta_3} h(u, v)$$

where $\delta_i \in \{0, 1\}$ for $i = 1, 2, 3$ and $\delta_1 + \delta_2 + \delta_3 = 1$. For an application of this case 2 model to data involving AIDS survival times ($X =$ time from onset of AIDS to death) for 92 members of the U.S. Air Force, see Aragon and Eberly (1992). [This data set also suggests the need for regression methods for interval censored data. See Huang and Wellner (1994), Huang (1994a,b), Rabinowitz, Tsiatis, and Aragon (1993) for work in this direction.]

Cases 1 and 2 extend to observation schemes involving several observation times in a variety of ways. We describe only the natural and obvious "case k" here. Suppose that $X \sim F_0$ is again the "time of interest", $\underline{U} = (U_1, \ldots, U_k) \sim H$ are observation times with $U_{j-1} \leq U_j$ for $j = 1, \ldots, k$, and we observe $(\underline{\Delta}, \underline{U})$ where $\Delta_j \equiv 1_{(U_{j-1}, U_j]}(X)$, $j = 1, \ldots, k+1$ where $U_0 \equiv 0$, $U_{k+1} \equiv \infty$. Then

$$(\underline{\Delta}|\underline{U}) \sim \text{Multinomial}_{k+1}(1, (F_0(u_1), F_0(u_2) - F_0(u_1), \ldots, 1 - F_0(u_k))),$$

and if H has density h with respect to Lebesgue measure on the subset $\{u \in R^k : \; 0 \le u_1 \le \ldots \le u_k\} \subset R^k$, then the joint density of $(\underline{\Delta}, \underline{U})$ is given by

$$p(\underline{\delta}, \underline{u}; F_0) = \prod_{j=1}^{k+1} \{F(u_j) - F(u_{j-1})\}^{\delta_j} h(\underline{u})$$

where $\delta_j \in \{0,1\}$ for $j = 1, \ldots, k+1$, and $\sum_{j=1}^{k+1} \delta_j = 1$.

Other models for interval censoring are also of interest: see e.g. Rabinowitz, Tsiatis, and Aragon (1993).

Now we turn to a description of the Nonparametric Maximum Likelihood Estimators (NPMLE's) for these models.

For a problem slightly more general than case 1, the NPMLE of F_0 was described by Ayer, Brunk, Ewing, Reid, and Silverman (1955). The following characterization and computational method is from GW (1992), proposition 1.2, page 41: First, order the observation times as

$$U_{(1)} \le \ldots \le U_{(n)}$$

and let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ denote the corresponding Δ_i's. Plot $(i\,, \sum_{j=1}^{i} \Delta_{(j)})$, $i = 1, \ldots, n$ and $(0,0)$. Form the Greatest Convex Minorant (GCM) G^* of these points. Then the NPMLE $\widehat{F}_n$ of F_0 is given by: $\widehat{F}_n(U_{(i)})$ is the left-derivative of the function G^* at i, $i = 1, \ldots, n$. For example, if $n = 5$, $U_{(\cdot)} = (1.2, 1.8, 2.1, 3.0, 3.5)$, and $\Delta_{(\cdot)} = (1, 0, 1, 1, 0)$, then $\widehat{F}_n(1.2) = \widehat{F}_n(1.8) = 1/2$ and $\widehat{F}_n(2.1) = \widehat{F}_n(3.0) = \widehat{F}_n(3.5) = 2/3$. The NPMLE does not specify where to place the remaining mass, and we will leave this undefined.

Characterization of the NPMLE of F_0 in case 2 was accomplished by Groeneboom (1991), and is given in Groeneboom's part of GW (1992), pages 43 - 50. To state Groeneboom's characterization, we need the following motivation and notation. The part of the log-likelihood for F divided by n in case 2 is given by

$$\begin{aligned} l_n(F) &= \frac{1}{n} \sum_{i=1}^{n} \{\Delta_{1i} \log F(U_i) \\ &\qquad + \Delta_{2i} \log(F(V_i) - F(U_i)) + \Delta_{3i} \log(1 - F(V_i))\} \\ &= \mathbb{P}_n \Big(1_{[x \le u]} \log F(u) \\ &\qquad + 1_{[u < x \le v]} \log(F(v) - F(u)) + 1_{[v < x]} \log(1 - F(v)) \Big) \end{aligned} \tag{1.1}$$

where $\mathrm{P}_n = n^{-1}\sum_{i=1}^n \delta_{(X_i,U_i,V_i)}$ is the empirical measure of the (unobservable) triples (X_i, U_i, V_i), $i = 1,\ldots,n$. A process which records the cumulative sums of the first derivatives with respect to F of $l_n(F)$ is the process

$$(1.2)\quad W_F(t) = \int_{u\le t}\left\{\frac{1_{[x\le u]}}{F(u)} - \frac{1_{[u<x\le v]}}{F(v)-F(u)}\right\}d\mathrm{P}_n(x,u,v) + \int_{v\le t}\left\{\frac{1_{[u<x\le v]}}{F(v)-F(u)} - \frac{1_{[v<x]}}{1-F(v)}\right\}d\mathrm{P}_n(x,u,v),$$

and the process recording the sums of minus the second derivatives is

$$(1.3)\quad G_F(t) = \int_{u\le t}\left\{\frac{1_{[x\le u]}}{F(u)^2} + \frac{1_{[u<x\le v]}}{(F(v)-F(u))^2}\right\}d\mathrm{P}_n(x,u,v) + \int_{v\le t}\left\{\frac{1_{[u<x\le v]}}{(F(v)-F(u))^2} + \frac{1_{[v<x]}}{(1-F(v))^2}\right\}d\mathrm{P}_n(x,u,v).$$

We then define a process V_F by

$$V_F(t) = W_F(t) + \int_{[0,t]} F(s)dG_F(s).$$

Examination of the log-likelihood $l_n(F)$ reveals that some of the observation times U_i or V_i do not appear in the log-likelihood; to isolate those times that do appear, we introduce

$$J_n^{(1)} = \cup_{i=1}^n\{U_i : \ X_i \le U_i \ \text{or}\ U_i < X_i \le V_i\},$$

$$J_n^{(2)} = \cup_{i=1}^n\{V_i : U_i < X_i \le V_i \ \text{or}\ V_i < X_i\},$$

and

$$J_n = J_n^{(1)} \cup J_n^{(2)} \equiv \{T_1,\ldots,T_m\}.$$

We write $T_{(1)} \le \cdots \le T_{(m)}$ for the ordered T_i's, and let $\underline{\Delta}_{(1)},\ldots,\underline{\Delta}_{(m)}$ denote the corresponding vectors of indicators. Note that m is random and that $n \le m \le 2n$.

The following proposition characterizing the NPMLE $\widehat{F}_n$ of F_0 is a restatement of GW (1992), Proposition 1.4, page 49.

Proposition 1.1. Suppose that $T_{(1)}$ has a corresponding $\underline{\Delta}_{(1)} = (1,0,0)$, and $T_{(m)}$ has a corresponding $\underline{\Delta}_{(m)} = (0,0,1)$. Then $\widehat{F}_n$ is the NPMLE of F_0 if and only if $\widehat{F}_n$ is the left - derivative of the greatest convex minorant of the "self - induced cumulative sum diagram" formed by the points

$$P_j \equiv (G_{\widehat{F}_n}(T_{(j)}), V_{\widehat{F}_n}(T_{(j)})), \qquad j = 1,\ldots,m$$

and $P_0 \equiv (0,0)$.

While the above characterization involves a "self - induced" cumulative sum diagram, and hence is circular, the solution exists and is well-defined in view of considerations involving Fenchel duality; see GW (1992), Proposition 1.3, page 46. Moreover, the above characterization leads naturally to the following *iterative convex minorant algorithm*:

(i) Start with some inital guess $\widehat{F}_n^{(0)}$ of F_0; set $k = 0$.

(ii) Form the cumulative sum diagram $P^{(k+1)}$ defined by $P_0^{(k+1)} = (0,0)$, $P_j^{(k+1)} = (G_{\widehat{F}_n^{(k)}}(T_{(j)}), V_{\widehat{F}_n^{(k)}}(T_{(j)}))$, $j = 1, \ldots, m$.

(iii) Let $\widehat{F}_n^{(k+1)} \equiv$ left - derivative of the GCM of cumulative sum diagram $P^{(k+1)}$.

(iv) Change k to $k+1$ and go to (ii).

The convergence properties of this algorithm for a fixed finite n have been examined by Aragon and Eberly (1992). Although Aragon and Eberly (1992) claim to prove global convergence of their modified ICM algorithm, in fact their proof only shows convergence once the points of jump do not change anymore (see their discussion of "secondary reductions" on page 132 and note that the dimension r in their Theorem 2, page 136, is fixed), and this is unrealistic since the points of jump may change in very late stages of the algorithm. Jongbloed (1995) has suggested a modification of the naive ICM algorithm which guarantees its global convergence, thus improving considerably on the results of Aragon and Eberly (1992). He introduced a line search along the direction suplied by the ICM algorithm to obtain a closed algorithmic map by a variant of Armijo's rule. Jongbloed (1995) also makes a number of suggestions for computationally practical modifications of the ICM algorithm. In practice the ICM algorithm converges quite quickly, and is considerably faster than the EM-algorithm.

The pointwise (local) behavior of $\widehat{F}_n$ is still an open problem. GW (1992) instead have studied the one-step "surrogate" for $\widehat{F}_n$ obtained from one step of the above algorithm starting at the true distribution function F_0: thus we let $F_n^{(1)}$ denote $\widehat{F}_n^{(1)}$ when $\widehat{F}_n^{(0)} = F_0$. Note that this is *not* a true estimator since we do not know F_0, and hence cannot start the algorithm at F_0 when dealing with data. However $F_n^{(1)}$ is a useful theoretical construct which we believe has the same asymptotic behavior as the NPMLE $\widehat{F}_n$. This is what is termed the *working hypothesis* in GW (1992), page 89: if

$$\alpha_n(F_n^{(1)}(t) - F_0(t)) \to_d Z$$

as $n \to \infty$ for some sequence $\alpha_n \to \infty$, then

$$\alpha_n(\widehat{F}_n(t) - F_0(t)) \to_d Z\,.$$

In the following sections we will *not* confirm (or refute!) this "working hypothesis". But will study the behavior of $F_n^{(1)}$ under different hypotheses than those imposed in GW (1992).

2. Known Limit Theory at a Fixed Point: Cases 1 and 2. In this section we summarize the currently known results for the behavior of the NPMLE's $\widehat{F}_n$ at a fixed point under cases 1 and 2. The two results stated below are from GW (1992), section 5, pages 89 and 100, respectively.

In case 1 the behavior of $\widehat{F}_n$ itself is known.

Theorem 2.1. Suppose that $0 < F_0(t_0), H(t_0) < 1$ and suppose that F_0 and H are differentiable at t_0 with strictly positive derivatives $f_0(t_0)$ and $h(t_0)$ respectively. Let $\widehat{F}_n$ be the NPMLE of F_0. Then

$$n^{1/3}\left\{\widehat{F}_n(t_0) - F_0(t_0)\right\} / \left\{\frac{1}{2} f_0(t_0)/c_1(t_0)\right\}^{1/3} \to_d 2Z$$

where Z is the last time where standard two-sided Brownian motion minus the parabola $y(t) = t^2$ reaches its maximum, and where

$$c_1(t_0) \equiv \frac{h(t_0)}{F_0(t_0)} + \frac{h(t_0)}{1 - F_0(t_0)} = \frac{h(t_0)}{F_0(t_0)(1 - F_0(t_0))} . \tag{2.4}$$

Because the limit distribution described by the random variable Z above apparently arose for the first time in the work of Chernoff (1964) in connection with the estimation of the mode of a distribution, we propose to call the distribution of Z *Chernoff's distribution*, and we say that $n^{1/3}\left\{\widehat{F}_n(t_0) - F_0(t_0)\right\}$ is *asymptotically Chernoffian* times the constant

$$2\left\{\frac{1}{2} f_0(t_0)/c_1(t_0)\right\}^{1/3}.$$

The analytical properties of the distribution of Z have been completely determined by Groeneboom (1989); in particular, Groeneboom shows that the density f_Z of Z satisfies

$$f_Z(t) \sim 2^{5/3}|t| \exp(-\frac{2}{3}|t|^3 + 2^{1/3} a_1 |t|)/Ai'(a_1) \qquad \text{as} \qquad |t| \to \infty$$

where $a_1 = -2.3381...$ is the largest zero of the Airy function Ai and $Ai'(a_1) = 0.7022....$

The behavior of $\widehat{F}_n$ in case 2 is still unknown. what has been proved is the following result for the theoretical one-step $F_n^{(1)}$.

Theorem 2.2. Suppose that $0 < F_0(t_0), H(t_0, t_0) < 1$, and let $F_n^{(1)}$ be the estimator of F_0 obtained at the first step of the iterative convex minorant algorithm. Suppose that $f_0(t_0) > 0$, $h(t_0, t_0) > 0$, and

$$h(t, t) \equiv \lim_{v \downarrow t} h(t, v)$$

is continuous in t in a neighborhood of t_0. Then

$$(n \log n)^{1/3}\left\{F_n^{(1)}(t_0) - F_0(t_0)\right\}/\left\{\frac{1}{2}f_0(t_0)/b_2(t_0)\right\}^{1/3} \to_d 2Z$$

where $Z \sim$ Chernoff, and where

$$b_2(t_0) \equiv \frac{2}{3}h(t_0, t_0)/f_0(t_0). \tag{2.5}$$

What is happening in theorem 2.2 is that the positivity hypothesis on h at the point (t_0, t_0) implies that the pair (U, V) are close together with substantial probability. When U and V are close together and X falls between U and V, X is known quite accurately (because U and V are observed). Even though this occurs only occasionally, it occurs frequently enough under the positivity hypothesis in theorem 2.2 to cause an increase in the rate of convergence increases from the $n^{-1/3}$ rate in case 1 to $(n \log n)^{-1/3}$ in case 2 under this positivity hypothesis. (Note that X is never "closely bracketed" in case 1 as it can be in case 2.) Moreover, the observations with $\Delta = (0,1,0)$ dominate the large sample behavior of $F_n^{(1)}(t_0)$, and the other two types of observations (with $\Delta = (1,0,0)$ or $(0,0,1)$) do not contribute at all asymptotically.

In the following section our goal is to examine the behavior of $F_n^{(1)}$ under hypotheses which force the pairs (U, V) to have U sufficiently separated from V to maintain the $n^{-1/3}$ convergence rate as in case 1.

3. New Limit Theory for Case 2 at a Fixed Point. Now suppose that the joint distribution H of observation times (U, V) in case 2 has a density function which puts sufficiently little mass along the diagonal $u = v$ in the sense that the integrals

$$k_1(u) \equiv \int_u^M \frac{h(u,v)}{F_0(v) - F_0(u)} dv$$

and

$$k_2(v) \equiv \int_0^v \frac{h(u,v)}{F_0(v) - F_0(u)} du$$

are finite for u, v in a neighborhood of t_0. [On the other hand, note that if h is the uniform density on $\{(u,v) : 0 \le u \le v \le 1\}$ and F_0 has a density f_0 which is positive at u, then k_1 and k_2 are both infinity at u.] Here are two simple examples showing that this is possible.

Example 1. Suppose that $0 < \epsilon < 1$,

$$h(u,v) = \frac{2}{(1-\epsilon)^2} 1_{[0 \le u \le v - \epsilon \le 1 - \epsilon,\, u \le 1-\epsilon]},$$

and that F_0 is the uniform distribution on $[0,1]$. Then it is easily calculated that

$$k_1(u) = a_\epsilon \log\left(\frac{1-u}{\epsilon}\right) 1_{[0,1-\epsilon]}(u), \qquad k_2(v) = a_\epsilon \log\left(\frac{v}{\epsilon}\right) 1_{[\epsilon,1]}(v).$$

where $a_\epsilon \equiv 2(1-\epsilon)^{-2}$. In this case the marginal densities h_1, h_2 are given by

$$h_1(u) = a_\epsilon\{1-\epsilon-u\}1_{[0,1-\epsilon]}(u), \qquad h_2(v) = a_\epsilon\{v-\epsilon\}1_{[\epsilon,1]}(v).$$

Example 2. Suppose that for some $\alpha > 0$ we have

$$h(u,v) = [F_0(v) - F_0(u)]^\alpha f_0(u) f_0(v) 1_{[0\le u \le v \le M]}/c_\alpha(F_0)$$

where

$$c_\alpha(F_0) \equiv \int_0^M \int_0^v [F_0(v) - F_0(u)]^\alpha f_0(u) f_0(v)\, du\, dv .$$

Then it is easily calculated that

$$k_1(u) = \frac{1}{\alpha}\frac{f_0(u)}{c_\alpha(F_0)}(1-F_0(u))^\alpha, \qquad k_2(v) = \frac{1}{\alpha}\frac{f_0(v)}{c_\alpha(F_0)}F_0(v)^\alpha .$$

The marginal densities h_1, h_2 are given by

$$h_1(u) = \frac{1}{1+\alpha}\frac{f_0(u)}{c_\alpha(F_0)}(1-F_0(u))^{1+\alpha}, \qquad h_2(v) = \frac{1}{1+\alpha}\frac{f_0(v)}{c_\alpha(F_0)}F_0(v)^{1+\alpha} .$$

Of course it is *not* necessary to suppose that H and F are directly connected as in example 2; example 2 was formulated to obtain explicit formulas for k_1 and k_2. To reinforce this, here is one more example of the same type which does not yield explicit formulas for the k_i's, but still makes them finite at points u with $f_0(u) > 0$.

Example 3. Suppose that for some $\alpha > 0$ we have

$$h(u,v) = d_\alpha (v-u)^\alpha 1_{[0\le u\le v\le M]}$$

where $d_\alpha \equiv (\alpha+1)(\alpha+2)/M^{\alpha+2}$. Then $k_1(u)$ and $k_2(u)$ are finite if $f_0(u) > 0$; in this case the marginal densities h_1, h_2 are given by

$$h_1(u) = d_\alpha(1+\alpha)^{-1}(M-u)^{1+\alpha}1_{[0,M]}(u), \qquad h_2(v) = d_\alpha(1+\alpha)^{-1}v^{1+\alpha}1_{[0,M]}(v).$$

To formulate our theorems, we will need to assume a little bit of asymptotic negligibility as follows. First, for each fixed $\epsilon > 0$ define "tail" or "high - level contributions" versions of the functions k_1, k_2 by

$$k_1(u;\epsilon) \equiv \int_u^M \frac{h(u,v)}{F_0(v) - F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\epsilon]}\, dv$$

and

$$k_2(v;\epsilon) \equiv \int_0^v \frac{h(u,v)}{F_0(v)-F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\epsilon]} du\,.$$

Then we will require that

$$\alpha \int_{(t_0, t_0+t/\alpha]} k_i(u;\epsilon\alpha)du \to 0 \quad \text{as} \quad \alpha \to \infty \tag{3.6}$$

for each $\epsilon > 0$ and $i = 1, 2$. It is easily checked that (3.6) holds in examples 1 - 3.

We shall prove the following result for the 1-step estimator $F_n^{(1)}$; it parallels Groeneboom and Wellner (1992), theorem 5.3, page 100.

Theorem 3.1. Let F_0 and H have densities f_0 and h respectively, and suppose that h_1, h_2, k_1, and k_2 are continuous at t_0 and $f(t_0) > 0$. Suppose that (3.6) holds. Let $0 < F_0(t_0), H(t_0, t_0) < 1$, and let $F_n^{(1)}$ be the estimator of F_0, obtained at the first step of the iterative convex minorant algorithm. Then

$$n^{1/3}\left\{F_n^{(1)}(t_0) - F_0(t_0)\right\} \Big/ \left\{\frac{1}{2} f_0(t_0)/c_2(t_0)\right\}^{1/3} \to_d 2Z$$

where $Z \sim$ Chernoff and

$$c_2(t_0) \equiv \frac{h_1(t_0)}{F_0(t_0)} + k_1(t_0) + k_2(t_0) + \frac{h_2(t_0)}{1-F_0(t_0)}\,. \tag{3.7}$$

According to the "working hypothesis", formulated at the beginning of Chapter 5 of GW (1992), this leads us to believe that the NPMLE has the same limiting behavior at t_0 as $F_n^{(1)}$ under these hypotheses.

Note that the conclusion of Theorem 3.1 agrees with the result for case 1 in the sense that the corresponding constant is $c_1(t_0)$ given in (2.4) which should be compared to the present formula (3.7) for $c_2(t_0)$.

To give perhaps just a bit more explanation for the difference in the rates of convergence and the difference between $b_2(t_0)$ and $c_2(t_0)$ appearing in Theorems 2.2 and 3.1, the reader should compare the proof of Lemma 5.5 in GW (1992) with the proof of Lemma 6.1 given here in section 6. In particular, the variance calculation at the bottom of page 101 of GW (1992) deserves comparison to the (co)variance calculation here starting in the display below (6.16): In GW (1992) only two of the four terms at the bottom of page 101, the second and third, contribute (equally) to the limiting quantity $(2/3)\{h(t_0,t_0)/f(t_0)\}t$ appearing at the top of page 102; note that the first and fourth terms on page 101 (corresponding to $\underline{\Delta} = (1,0,0)$ and to $\underline{\Delta} = (0,0,1)$ respectively) do not contribute at all, and the normalizations are chosen so that the logarithmic factors arising from the integrals in the second and third terms are cancelled by the normalizations. In contrast, taking $s = t$ in the covariance calculation in the proof of Lemma 6.1 here, all four terms in the display following (6.16) contribute equally and lead to the expression for c_2.

Is $n^{-1/3}$ the optimal rate of convergence under the current hypotheses? We claim that it is. Gill and Levit (1995) used Van Trees inequality to establish lower bounds for estimation of $F_0(t_0)$. Their Theorem 3 shows that the optimal rate of convergence of any estimator of $F_0(t_0)$ is $(n \log n)^{-1/3}$ under Lipschitz hypotheses on F_0 and positivity of the observation density h along the diagonal as in Groeneboom's theorem 2.2. By modifying the proof of Gill and Levit (1995) (but maintaining the Lipschitz hypotheses on F_0), it is easily shown that the optimal rate of convergence is $n^{-1/3}$ under our separation hypothesis (3.6).

We close this section with a calculation of the constant $c_2(t_0)$ appearing in theorem 3.1 in both examples 1 and 2.

Example 1, Continued. In this case $c_2(t_0)$ is

$$c_2(t_0) = \begin{cases} \frac{2}{(1-\epsilon)^2}\left\{\frac{1-\epsilon}{t_0} - 1 + \log\left(\frac{1-t_0}{\epsilon}\right)\right\} & \text{if } 0 < t_0 \le \epsilon \\ \frac{2}{(1-\epsilon)^2}\left\{\frac{1-\epsilon}{t_0(1-t_0)} - 2 + \log\left(\frac{t_0(1-t_0)}{\epsilon^2}\right)\right\} & \text{if } \epsilon \le t_0 \le 1-\epsilon \\ \frac{2}{(1-\epsilon)^2}\left\{\frac{1-\epsilon}{1-t_0} - 1 + \log\left(\frac{t_0}{\epsilon}\right)\right\} & \text{if } 1-\epsilon \le t_0 \le 1 \end{cases} .$$

Example 2, Continued. In this case $c_2(t_0)$ is given by

$$c_2(t_0) = \frac{1}{c_\alpha} f_0(t_0) \left\{ \overline{F}_0^\alpha(t_0)\left[\frac{1}{1+\alpha}\frac{\overline{F}_0}{F_0}(t_0) + \frac{1}{\alpha}\right] + F_0^\alpha(t_0)\left[\frac{1}{1+\alpha}\frac{F_0}{\overline{F}_0}(t_0) + \frac{1}{\alpha}\right]\right\} .$$

It is instructive to plot examine plots of $c_1(t_0)$, $c_2(t_0)$, and $c_1(t_0)/c_2(t_0)$ for examples 1 and 2.

4. Interval censoring, Case k. Now our goal is to outline what happens when there are k different observation points $U_1, \ldots, U_k$. We suppose that $\underline{U} = (U_1, \ldots, U_k) \sim H$ with density h on $\{u \in R^k : 0 \le u_1 \le \ldots \le u_k \le M\}$. Suppose that $X \sim F$ on $[0, M]$. Set

$$\Delta_j = 1_{(U_{j-1}, U_j]}(X), \qquad j = 1, \ldots, k+1$$

and write $\underline{\Delta} = (\Delta_1, \ldots, \Delta_{k+1})$ where $U_0 \equiv 0$, $U_{k+1} \equiv M$. Then the joint density of $\underline{\Delta}$ and $\underline{U}$ is

$$p(\underline{\delta}, \underline{u}; F) = \prod_{j=1}^{k+1} \{F(u_j) - F(u_{j-1})\}^{\delta_j} h(\underline{u})$$

where $\delta_j \in \{0, 1\}$ for $j = 1, \ldots, k+1$, and $\sum_{j=1}^{k+1} \delta_j = 1$. Hence

$$\log p(\underline{\delta}, \underline{u}; F) = \sum_{j=1}^{k+1} \delta_j \log\{F(u_j) - F(u_{j-1})\} + \log h(\underline{u}),$$

and, with $F_\epsilon = (1-\epsilon)F + \epsilon G$,

$$\frac{\partial}{\partial\epsilon}\log p(\underline{\delta}, \underline{u}; F)|_{\epsilon=0} = \sum_{j=1}^{k+1} \delta_j \frac{1}{F(u_j) - F(u_{j-1})}(G-F)(u_{j-1}, u_j].$$

Now suppose we observe $(\underline{\Delta}_i, \underline{U}_i)$, $i = 1, \ldots, n$ i.i.d. with the same distribution as $(\underline{\Delta}, \underline{U})$. The log-likelihood of the data (divided by n) is

$$l_n(F) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k+1} \Delta_{i,j}\log(F(U_{i,j}) - F(U_{i,j-1})) \tag{4.8}$$

$$= \mathrm{P}_n\left(\sum_{j=1}^{k+1} 1_{[u_{j-1}<x\le u_j]}\log(F(u_j) - F(u_{j-1}))\right) \tag{4.9}$$

where P_n is the (unobservable) empirical measure of $(X_i, \underline{U}_i)$, $i = 1, \ldots, n$ on R^{k+1}. We regard $F(u_j)$, $j = 1, \ldots, k$ as "parameters" to be estimated. Thus the derivatives of the log - likelihood with respect to $F(u_j) \equiv \theta_j$ are

$$\frac{\partial}{\partial\theta_j} l_n(F) = \mathrm{P}_n\left(1_{[u_{j-1}<x\le u_j]}\frac{1}{F(u_j) - F(u_{j-1})} - 1_{[u_j<x\le u_{j+1}]}\frac{1}{F(u_{j+1}) - F(u_j)}\right).$$

Thus, paralleling GW (1992), we define

$$W_F(t) \equiv \sum_{j=1}^{k}\int_{u_j\le t}\left\{\frac{1_{[u_{j-1}<x\le u_j]}}{F(u_j) - F(u_{j-1})} - \frac{1_{[u_j<x\le u_{j+1}]}}{F(u_{j+1}) - F(u_j)}\right\} d\,\mathrm{P}_n(x, \underline{u})$$

and

$$G_F(t) \equiv \sum_{j=1}^{k}\int_{u_j\le t}\left\{\frac{1_{[u_{j-1}<x\le u_j]}}{[F(u_j) - F(u_{j-1})]^2} + \frac{1_{[u_j<x\le u_{j+1}]}}{[F(u_{j+1}) - F(u_j)]^2}\right\} d\,\mathrm{P}_n(x, \underline{u})\ .$$

Let $h_j(u_{j-1}, u_j)$ denote the marginal densities of the pairs (U_{j-1}, U_j), $j = 2, \ldots, k$. We first state a conjectured theorem concerning the one-step NPMLE $F_n^{(1)}$ corresponding to the type of hypotheses in GW (1992); in this first theorem at least some of the pairs (U_{j-1}, U_j) are close together with substantial probability – and these pairs dominate the asymptotics because of the resulting additional factor of $(\log n)^{1/3}$ in the convergence rate. Our second conjecture will impose hypotheses requiring separation of *all* the pairs (U_{j-1}, U_j), $j = 2, \ldots, k$, along the lines of the hypotheses imposed in section 2.

Let J be the subset of the indices $\{2, \ldots, k\}$ consisting of those satisfying $h_j(t_0, t_0) > 0$.

Conjectured Theorem 4.1. Suppose that $0 < F_0(t_0)$, $H(t_0, \ldots, t_0) < 1$, and let $F_n^{(1)}$ be the estimator of F_0 obtained at the first step of the iterative convex minorant algorithm. Suppose that $f_0(t_0) > 0$ and $h_j(t_0, t_0) > 0$ for $j \in J$ where J is nonvoid and

$$h_j(t,t) \equiv \lim_{v \downarrow t} h_j(t,v)$$

is continuous in t in a neighborhood of t_0 for all $j \in J$. Then

$$(n \log n)^{1/3}\Big\{F_n^{(1)}(t_0) - F_0(t_0)\Big\}\Big/\Big\{\frac{1}{2}f_0(t_0)/b(t_0)\Big\}^{1/3} \to_d 2Z$$

where $Z \sim$ Chernoff, and where

$$b(t_0) \equiv \frac{2}{3}\sum_{j \in J} h_j(t_0, t_0)/f_0(t_0). \tag{4.10}$$

When $k = 2$ and $J = \{2\}$, then conjectured theorem 4.1 reduces to the statement of GW (1992), Theorem 5.3, page 100. Note that the conclusion of the conjectured theorem 4.1 can be restated as

$$(n \log n)^{1/3}\Big\{F_n^{(1)}(t_0) - F_0(t_0)\Big\}\Big/\Big\{\frac{3}{4}f_0^2(t_0)/d(t_0)\Big\}^{1/3} \to_d 2Z$$

where

$$d(t_0) \equiv \sum_{j \in J} h_j(t_0, t_0). \tag{4.11}$$

To state our conjectured theorem for k observation points under hypotheses similar to those in section 2, we define

$$k_{1j}(u) \equiv \int_u^M \frac{h_j(u,v)}{F_0(v) - F_0(u)}dv, \qquad \text{and} \qquad k_{2j}(v) \equiv \int_0^v \frac{h_j(u,v)}{F_0(v) - F_0(u)}du$$

for $j = 2, \ldots, k$ where h_j denotes the joint density of (U_{j-1}, U_j), $j = 2, \ldots, k$ as before. We will suppose that all of the functions k_{mj}, $j = 2, \ldots, k$, $m = 1, 2$, are finite, and moreover that with

$$k_{1j}(u;\epsilon) \equiv \int_u^M \frac{h_j(u,v)}{F_0(v) - F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\epsilon]}dv$$

and

$$k_{2j}(v;\epsilon) \equiv \int_0^v \frac{h_j(u,v)}{F_0(v) - F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\epsilon]}du\,,$$

we have

$$\alpha \int_{(t_0, t_0 + t/\alpha]} k_{mj}(u; \epsilon\alpha)du \to 0 \qquad \text{as} \qquad \alpha \to \infty \tag{4.12}$$

for each $\epsilon > 0$ and $j = 2, \ldots, k$, $m = 1, 2$.

Conjectured Theorem 4.2. Let F_0 and H have densities f_0 and h respectively, and suppose that h_1, h_k, k_{1j}, and k_{2j}, $j = 2, \ldots, k$ are continuous at t_0 and $f(t_0) > 0$ where h_1, h_k are the marginal densities of U_1, U_k respectively. Suppose that (4.12) holds. Let $0 < F_0(t_0), H(t_0, t_0) < 1$, and let $F_n^{(1)}$ be the estimator of F_0, obtained at the first step of the iterative convex minorant algorithm. Then

$$n^{1/3}\Big\{F_n^{(1)}(t_0) - F_0(t_0)\Big\}\Big/\Big\{\frac{1}{2}f_0(t_0)/c_k(t_0)\Big\}^{1/3} \to_d 2Z$$

where $Z \sim$ Chernoff, and where

$$c_k(t_0) \equiv \frac{h_1(t_0)}{F_0(t_0)} + \sum_{j=2}^{k}\{k_{1j}(t_0) + k_{2j}(t_0)\} + \frac{h_k(t_0)}{1 - F_0(t_0)} \,. \tag{4.13}$$

5. Discussion and Further Problems. First a summary of the rationale for, and possible advantages of, the alternative hypotheses suggested here:

- Under the alternative hypotheses we obtain limit theorems for the NPMLE at a point (or at least the theoretical construct $F_n^{(1)}$) which are comparable with case 1, and give the possibility of addressing the question of how much is gained by two observation times over one observation time.

- Study of the properties of smooth functionals such as $E_{F_0}X$ under case 2 may be easier under the separation type hyptheses and certainly will be considerably easier under the "strict separation hypothesis" $P(V - U \geq \epsilon) = 1$.

- Realism: in practice, separation of the observation times U and V may be forced by practical or economic considerations.

- Mathematical completeness: we need to understand how these estimators behave on as much of the parameter space

  $$\Theta = \{(F_0, H) : \; F_0 \text{ a d.f. on } R^+, \; H \text{ a d.f. on } R^{+2\leq}\}$$

  as possible.

Despite the slow rates of convergence of the NPMLE or the one-step NPMLE in cases 1, 2, and k, smooth functionals such as means or other moment estimators, are sufficiently smooth to enjoy $n^{-1/2}$ rates of convergence; this has been shown in GW (1992) and Huang and Wellner

(1995) for case 1, and in Geskus (1992), Geskus and Groeneboom (1995a,b) for case 2 under a "separated" observation times formulation as investigated here. See Groeneboom (1995) for further work on the the behavior of smooth functionals for case 2 data without "separation" type hypotheses – where the question is still not quite resolved. Efficient estimates of the parametric part of the Cox proportional hazards model with case 1 data, and with case 2 data under a "strict separation hypothesis" as mentioned above, have been constructed by Huang (1996), and by Huang and Wellner (1995) respectively.

Here are a few problems suggested by the above development.

A. Does the MLE itself have the same behavior as $F_n^{(1)}$ under the hypotheses of either theorem 2.1 or theorem 3.1? In other words, *does the "working hypothesis" hold*?

B. What is the behavior of the NPMLE of the mean and other smooth functionals under these hypotheses? [I conjecture that it will be easier to prove. In fact, this has now been carried out by Geskus and Groeneboom (1995a,b).]

C. What is the rate of convergence for (F, H) pairs "between" the alternative hypotheses of theorem 3.1 and those of theorem 2.1? Is there a way of unifying the various cases by using a random - norming?

D. Are the conjectured theorems 4.1 and 4.2 true? Do they remain true for the NPMLE $\widehat{F}_n$ itself?

6. Proofs for Section 3. Proceeding as in GW (1992), we introduce the processes $W_n^{(0)}$ and $G_n^{(0)}$ defined by

$$W_n^{(0)} = W_{F_0}, \text{ and } G_n^{(0)} = G_{F_0},$$

where W_{F_0} and G_{F_0} are defined by GW (1992), page 45, (1.25), and page 49, (1.29), respectively. The process $V_n^{(0)}$ is defined by

$$V_n^{(0)}(t) = W_n^{(0)}(t) + \int_{[0,t]} \Big(F_0(t') - F_0(t_0)\Big)\, dG_n^{(0)}(t'), \qquad t \geq 0.$$

We have the following result for $V_n^{(0)}$ corresponding to GW (1992) lemma 5.5.

Lemma 6.1. Suppose that the hypotheses of theorem 3.1 hold, and define the process $U_n^{(0)}$ by

$$U_n^{(0)}(t) = n^{2/3}\Big\{V_n^{(0)}(t_0 + n^{-1/3}t) - V_n^{(0)}(t_0)\Big\}, \qquad t \in R,$$

where $U_n^{(0)}(t) = 0$, if $t \le -t_0 n^{-1/3}$. Then $U_n^{(0)}$ converges in distribution, in the topology of uniform convergence on compacta on the space of locally bounded real-valued functions on R, to the process U, defined by

$$U(t) = \sqrt{c_2(t_0)}W(t) + \tfrac{1}{2} f_0(t_0)c_2(t_0)t^2, \qquad t \in R,$$

where W is (standard) two-sided Brownian motion on R, originating from zero and $c_2(t_0)$ is as defined in (3.7).

Proof. We first show that the process

$$Z_n^{(0)}(t) \equiv n^{2/3}\Big\{W_n^{(0)}(t_0 + n^{-1/3}t) - W_n^{(0)}(t_0)\Big\}, \qquad t \ge 0 \tag{6.14}$$

converges, in the topology of uniform convergence on compacta, to the process

$$t \mapsto \sqrt{c_2(t_0)}W(t), \qquad t \ge 0. \tag{6.15}$$

To do this, we will use Kim and Pollard (1990), theorem 4.7, or equivalently lemmas 4.5 and 4.6.

We first verify the hypotheses of lemma 4.5: note that $Z_n^{(0)}(t) = n^{2/3}\mathrm{P}_n g(\cdot, t_0 + n^{-1/3}t)$ for the family of functions $\{g(\cdot, t_1)\}_{t_1 \ge t_0}$ defined by

$$\begin{aligned} g(x,u,v,t_1) \equiv\ & 1_{[t_0<u\le t_1]}1_{[x\le u]}\frac{1}{F_0(u)} - 1_{[t_0<u\le t_1]}1_{[u<x\le v]}\frac{1}{F_0(v)-F_0(u)} \\ & + 1_{[t_0<v\le t_1]}1_{[u<x\le v]}\frac{1}{F_0(v)-F_0(u)} - 1_{[t_0<v\le t_1]}1_{[v<x]}\frac{1}{1-F_0(v)}. \end{aligned} \tag{6.16}$$

Then $Pg(X,U,V,t_0,t_1) = 0$ for all t_0, t_1 so (iii) of lemma 4.5 holds easily. To check (ii), fix $0 \le s < t$. Then

$$\begin{aligned} Pg(\cdot, t_0 + s/\alpha)g(\cdot, t_0+t/\alpha) &= \int_{(t_0,t_0+s/\alpha]} \frac{1}{F_0(u)}dH_1(u) \\ &\quad + \int_{u\in(t_0,t_0+s/\alpha]}\int \frac{1}{F_0(v)-F_0(u)}dH(u,v) \\ &\quad + \int_{v\in(t_0,t_0+s/\alpha]}\int \frac{1}{F_0(v)-F_0(u)}dH(u,v) \\ &\quad + \int_{(t_0,t_0+s/\alpha]} \frac{1}{1-F_0(v)}dH_2(v) \\ &= \int_{(t_0,t_0+s/\alpha]} c_2(u)du \end{aligned}$$

where $c_2(u)$ is as defined in (3.7). Hence by continuity of h_1, h_2, k_1, k_2 (and hence also of c) in a neighborhood of t_0 it follows that

$$\alpha P g(\cdot, t_0 + s/\alpha)g(\cdot, t_0+t/\alpha) \to s c_2(t_0);$$

hence (ii) holds. To verify (iv) we compute

$$\begin{aligned}
&\alpha P g^2(\cdot, t_0 + t/\alpha) 1_{\{|g(\cdot,t_0+t/\alpha)|>\alpha\epsilon\}} \\
&\le \alpha\Big\{\int_{(t_0,t_0+t/\alpha]} \frac{1}{F_0(u)} 1_{[1/F_0(u)>\alpha\epsilon/4]} dH_1(u) \\
&\qquad + \int_{u\in(t_0,t_0+t/\alpha]}\int \frac{1}{F_0(v)-F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\alpha\epsilon/4]} dH(u,v) \\
&\qquad + \int_{v\in(t_0,t_0+t/\alpha]}\int \frac{1}{F_0(v)-F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\alpha\epsilon/4]} dH(u,v) \\
&\qquad + \int_{(t_0,t_0+t/\alpha]} \frac{1}{1-F_0(v)} 1_{[1/(1-F_0(v))>\alpha\epsilon/4]} dH_2(v)\Big\} \\
&\le \alpha\Big\{\int_{(t_0,t_0+t/\alpha]} \frac{1}{F_0(u)} dH_1(u) 1_{[1/F_0(t_0)>\alpha\epsilon/4]} \\
&\qquad + \int_{u\in(t_0,t_0+t/\alpha]} k_1(u;\alpha\epsilon/4)du \\
&\qquad + \int_{v\in(t_0,t_0+t/\alpha]} k_2(v;\alpha\epsilon/4)dv \\
&\qquad + \int_{(t_0,t_0+t/\alpha]} \frac{1}{1-F_0(v)} dH_2(v) 1_{[1/(1-F_0(t_0+t/\alpha))>\alpha\epsilon/4]}\Big\} \\
&\to 0
\end{aligned}$$

easily for the two end terms and by (3.6) for the two middle terms. Thus (iv) of Kim and Pollard's lemma 4.5 holds, and hence by lemma 4.5 finite-dimensional convergence of the processes $W_n^{(0)}$ holds. □

Now we want to apply Kim and Pollard (1990), lemma 4.6 to deduce tightness and hence weak convergence of the processes $W_n^{(0)}$. The classes of functions we need to consider are

$$\mathcal{G}_R = \{g(\cdot, t_1) : \ |t_1 - t_0| \le R\}, \qquad R > 0$$

where the functions $g(\cdot, t_1)$ are given for $t_1 \ge 0$ in (6.16); for $t_1 \le 0$ an obvious analogous expression holds. First note that an envelope function for $\mathcal{G}_R$ is

$$\begin{aligned}
G_R(x,u,v) \equiv\ & 1_{[t_0-R\le u\le t_0+R]}\Big\{1_{[x\le u]}\frac{1}{F_0(u)} + 1_{[u<x\le v]}\frac{1}{F_0(v)-F_0(u)}\Big\} \\
&+ 1_{[t_0-R<v\le t_0+R]}\Big\{1_{[u<x\le v]}\frac{1}{F_0(v)-F_0(u)} + 1_{[v<x]}\frac{1}{1-F_0(v)}\Big\}.
\end{aligned}$$

The classes $\mathcal{G}_R$ are clearly uniformly manageable for their envelopes (under the assumption that k_1, k_2 are finite) since they are of the form finite sum of indicator of an interval times a fixed square integrable function. Thus (i) of lemma 4.6 holds. To show that (ii) holds, we compute

$$PG_R^2 \le 4\Big\{\int_{t_0-R}^{t_0+R} \frac{1}{F_0(u)} dH_1(u) + \int_{t_0-R}^{t_0+R} k_1(u)du + \int_{t_0-R}^{t_0+R} k_2(v)dv + \int_{t_0-R}^{t_0+R} \frac{1}{1-F_0(v)} dH_2(v)\Big\}$$

which is easily seen to be $O(R)$ as $R \to 0$. Hence (ii) holds. Furthermore, for $t_1 < t_2$,

$$g(x,u,v,t_2) - g(x,u,v,t_1) \equiv 1_{[t_1<u\le t_2]}\Big\{1_{[x\le u]}\frac{1}{F_0(u)} - 1_{[u<x\le v]}\frac{1}{F_0(v)-F_0(u)}\Big\} + 1_{[t_1<v\le t_2]}\Big\{1_{[u<x\le v]}\frac{1}{F_0(v)-F_0(u)} - 1_{[v<x]}\frac{1}{1-F_0(v)}\Big\}.$$

Therefore

$$P|g(\cdot,t_2) - g(\cdot,t_1)| \le 4\Big\{\int_{t_1}^{t_2} \frac{1}{F_0(u)} dH_1(u) + \int_{t_1}^{t_2} k_1(u)du + \int_{t_1}^{t_2} k_2(v)dv + \int_{t_1}^{t_2} \frac{1}{1-F_0(v)} dH_2(v)\Big\}$$

which is of the order $t_2 - t_1$ for t_2, t_1 near t_0. Thus (iii) holds. Similarly, using the assumption (ii), for $\epsilon > 0$ there exists K so that

$$PG_R^2 1_{[G_R>K]} < \epsilon R$$

for R near 0; i.e. (iv) holds. Hence we conclude from Kim and Pollard (1990) lemma 4.6 that the stochastic equicontinuity condition holds, and by lemmas 4.5 and 4.6 together that the processes $W_n^{(0)}$ converge weakly in the topology of uniform convergence on compact to the process $\sqrt{c_2(t_0)}W(t)$ where W is two-sided Brownian motion starting from 0.

To complete the proof of lemma 6.1, it suffices to show that

$$n^{2/3}\int_{[t_0,t_0+n^{-1/3}t]} (F_0(u) - F_0(t_0))dG_n^{(0)}(u) \equiv T_n^{(0)}(t)$$

converges in probability, uniformly on compacts, to the deterministic function $f_0(t_0)c_2(t_0)t^2/2$. But

$$ET_n^{(0)}(t) = n^{2/3}\Big\{\int_{u\in[t_0,t_0+n^{-1/3}t]} (F_0(u)-F_0(t_0))\{\frac{h_1(u)}{F_0(u)} + k_1(u)\}du + \int_{v\in[t_0,t_0+n^{-1/3}t]} (F_0(v)-F_0(t_0))\{\frac{h_2(v)}{1-F_0(v)} + k_2(v)\}dv\Big\} \to \frac{1}{2}f_0(t_0)c_2(t_0)t^2$$

uniformly on compact t intervals in view of our hypotheses on f_0, h_i, k_i, $i = 1, 2$, while

$$Var(T_n^{(0)}(t)) = O(n^{-2/3})$$

uniformly in t in compacts. Hence the desired convergence holds, and the proof of lemma 6.1 is complete. □

Now set $a_0 \equiv F_0(t_0)$, and define, for $a > 0$,

$$T_n^{(0)}(a) \equiv \sup\{t \in R : \ V_n^{(0)}(t) - (a - a_0)G_n^{(0)}(t) \ \text{ is minimal}\}.$$

The second key lemma needed in proving theorem 3.1 is the analogue of GW (1992), Lemma 5.6, page 103. In fact we will not give the (rather long) proof here.

Lemma 6.2. Suppose the hypotheses of theorem 3.1 hold. Then for each $\epsilon > 0$ and $M_1 > 0$, there is an $M_2 > 0$ so that

$$\text{limsup}_{n\to\infty} P\Big\{ \sup_{|a|\le M_1} n^{1/3}|T_n^{(0)}(a_0 + n^{-1/3}a) - t_0| > M_2 \Big\} < \epsilon.$$

Proof of theorem 3.1. As in GW (1992) equation (5.16) we have

$$P(F_n^{(1)}(t_0) - F_0(t_0) > n^{-1/3}a) = P(T_n^{(0)}(a_0 + n^{-1/3}a) < t_0)$$

where a_0 and $T_n^{(0)}(a)$ were defined just before the statement of lemma 6.2. Now it follows from lemmas 6.1 and 6.2 that

$$\{n^{1/3}(T_n^{(0)}(a_0 + n^{-1/3}a) - t_0) : \ a \in R\}$$

converges in the Skorohod topology on $D(R)$ to $\{T(a) : \ a \in R\}$ where

$$T(a) \equiv \sup\{t \in R : \ U(t) - ac(t_0)t \ \text{ is minimal}\}.$$

Hence, with $d \equiv 1/f_0(t_0)$,

$$\begin{aligned} P(T_n^{(0)}(a_0 + n^{-1/3}a) < t_0) &= P(n^{1/3}(T_n^{(0)}(a_0 + n^{-1/3}a) - t_0) < 0) \\ &\to P(T(a) < 0) \\ &= P(T(a) - da < -da) \\ &= P(T(0) < -da) \end{aligned}$$

by the stationarity of $\{T(a) - da : \ a \in R\}$ proved in Groeneboom (1989); note that

$$T(a) = \sup\{t \in R : \ \sqrt{c_2(t_0)}W(t) + \frac{1}{2}f_0(t_0)c_2(t_0)(t - da)^2 \ \text{ is minimal}\}.$$

Now $T(0)$ is the last time when

$$W(c_2(t_0)t) + \frac{1}{2} f_0(t_0)c_2(t_0)t^2$$

is minimal. By Brownian scaling

$$W(c_2(t_0)t) + \frac{1}{2} f_0(t_0)c_2(t_0)t^2 =_d a(W(s) + s^2)$$

for $s = bt$, $b \equiv \{2f_0^2(t_0)c_2(t_0)\}^{1/3}/2$, and $a \equiv \sqrt{c_2(t_0)/b}$. Thus

$$\frac{1}{2} T(0)\{2f_0^2(t_0)c_2(t_0)\}^{1/3} \tag{6.17}$$

is the last time when $W(s)+s^2$ is minimum. By symmetry of the distribution of Brownian motion with respect to the time axis, this means that (6.17) has the same distribution as the last time Z when $W(s) - s^2$ is maximum. Thus

$$\begin{aligned} P(T(0) < -da) &= P(Z < -da\frac{1}{2}\{2f_0^2(t_0)c_2(t_0)\}^{1/3}) \\ &= P(Z < -a\frac{1}{2}\{2c_2(t_0)/f_0(t_0)\}^{1/3}) \\ &= P(Z > a\frac{1}{2}\{2c_2(t_0)/f_0(t_0)\}^{1/3}) \end{aligned}$$

by symmetry of the distribution of Z about 0. Hence we have

$$P(n^{1/3}(F_n^{(0)}(t_0) - F_0(t_0)) > a) \to P(2Z\{\frac{1}{2} f_0(t_0)/c_2(t_0)\}^{1/3} > a),$$

and the conclusion follows. □

ACKNOWLEDGEMENTS: The author owes thanks to Piet Groeneboom, Jian Huang, and Aad van der Vaart for helpful discussions about rates of convergence, interval censoring, and related problems.

REFERENCES

Aragon, J. and Eberly, D. (1992). On convergence of convex minorant algorithms for distribution estimation with interval-censored data. *J. Computational and Graphical Statistics* **1**, 129 - 140.

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.

Barlow, R.E., Bartholomew, D. J., Bremner, J. M., Brunk, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16**, 31 - 41.

Dudley, R.M. (1987). Universal Donsker classes and metric entropy. *Ann. Probability* **15**, 1306 - 1326.

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 Groeneboom, P. (1995a). Asymptotically optimal estimation of smooth functionals for interval censoring, part 1. *Statistica Neerl.* **49**, to appear.

Geskus, R. and Groeneboom, P. (1995b). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2. *Statistica Neerl.* **49**, to appear.

Gill, R. D. and Levit, B. Y. (1995). Applications of the Van Trees inequality: a Bayesian Cramér - Rao bound. *Bernoulli* **1**, ?? - ?? .

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, Lucien M. LeCam and Richard A. Olshen eds.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields* **79**, 327 - 368.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. *Technical Report 378*, Department of Statistics, Stanford University.

Groeneboom, P. (1995). Inverse problems in statistics. Proceedings of the St. Flour Summer School in Probability, 1994. *Lecture Notes in Math.* **??**, ??-??. Springer Verlag, Berlin.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Basel.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24**, to appear.

Huang, J. (1994b). Maximum likelihood estimation for proportional odds regression model with current status data. *Preprint*, University of Iowa.

Huang, J. and Wellner, J. A. (1994). Regression models with interval censoring. *Proceedings of the Kolmogorov Semester,* Euler Institute, St. Petersburg, Russia.

Huang, J. and Wellner, J. A. (1995a). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Statistica Neerlandica* **49**, to appear.

Huang, J. and Wellner, J. A. (1995b). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22**, 3 - 33.

Huang, J. and Wellner, J. A. (1995c). Efficient estimation for the proportional hazards model with "case 2" interval censoring. Submitted to *Biometrika.*

Huang, Y. and Zhang, C. H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22**, 1256 - 1274.

Jongbloed, G. (1995). Three Statistical Inverse Problems. Ph.D. Dissertation, Delft University.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191 - 219.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankyā Ser. A 31*, 23-36.

Prakasa Rao, B.L.S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Statist.* **41**, 507 - 519.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1993). Regression with interval censored data. *Technical Report*, Department of Biostatistics, Harvard School of Public Health

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Van der Vaart, A. W. and Wellner, J. A. (1995). *Weak Convergence and Empirical Processes.* Springer Verlag, New York, to appear.