The Annals of Statistics
2001, Vol. 29, No. 6, 1699–1731

LIKELIHOOD RATIO TESTS FOR MONOTONE FUNCTIONS

BY MOULINATH BANERJEE[1] AND JON A. WELLNER[2]

University of Michigan and University of Washington

We study the problem of testing for equality at a fixed point in the setting of nonparametric estimation of a monotone function. The likelihood ratio test for this hypothesis is derived in the particular case of interval censoring (or current status data) and its limiting distribution is obtained. The limiting distribution is that of the integral of the difference of the squared slope processes corresponding to a canonical version of the problem involving Brownian motion $+t^2$ and greatest convex minorants thereof. Inversion of the family of tests yields pointwise confidence intervals for the unknown distribution function. We also study the behavior of the statistic under local and fixed alternatives.

1. Introduction.

We shall consider likelihood ratio tests, and the corresponding confidence intervals, in a class of problems involving nonparametric estimation of a monotone function. The problem in each case involves testing the null hypothesis H_0 that the monotone function has a particular value at a fixed point t_0 in the domain of the function. Of course with each testing problem there is a related problem of finding confidence intervals. Here are some examples of the problems we have in mind.

EXAMPLE 1 (Monotone density function). Suppose that $X_1, X_2, \dots, X_n$ are i.i.d. random variables from the unknown density f on $[0, \infty)$ that is assumed to be decreasing (i.e., non-increasing). The maximum likelihood estimator $\hat{f}_n$ of f is the well-known Grenander estimator: it is the step function equal to the left-derivative of the least concave majorant of the empirical distribution function $\mathbb{F}_n$; see Grenander (1956), Prakasa Rao (1969) and Groeneboom (1985). For fixed $t_0 \in (0, \infty)$ and $\theta_0 \in (0, \infty)$, consider testing $H_0 : f(t_0) = \theta_0$ versus $H_1 : f(t_0) \neq \theta_0$. The corresponding interval estimation problem is to find a $1 - \alpha$ confidence interval for $f(t_0)$ for fixed $\alpha \in (0, 1)$.

EXAMPLE 2 (Interval censoring, current status data). Suppose that (X_i, T_i), $i = 1, \dots, n$, are i.i.d., where for each pair X_i and T_i are independent, and $X_i \sim F$ and $T_i \sim G$ where F and G are distribution functions on $[0, \infty)$. For each pair we observe $Y_i = (T_i, \Delta_i)$ where $\Delta_i = 1\{X_i \le T_i\}$. The goal is

Received August 2000; revised July 2001.
[1]Supported in part by NSF Grant DMS-00-71818.
[2]Supported in part by NSF Grant DMS-95-32039, NIAID Grant 2R01 AI291968-04 and the Stieltjes Institute.
AMS 2000 *subject classifications*. Primary 62G05; secondary 60G15, 62E20.
Key words and phrases. Asymptotic distribution, Brownian motion, constrained estimation, fixed alternatives, Gaussian process, greatest convex minorant, interval censoring, Kullback-Leibler discrepancy, least squares, local alternatives, likelihood ratio, monotone function, slope processes.

to make inference about the monotone (increasing) function F. The nonparametric maximum likelihood estimator $\mathbb{F}_n$ of F is well known; see, for example, Ayer, Brunk, Ewing, Reid and Silverman (1955) or Groeneboom and Wellner (1992). We are interested here in likelihood ratio tests of H_0: $F(t_0) = \theta_0$ versus H_1: $F(t_0) \neq \theta_0$ for $t_0 \in (0, \infty)$ and $\theta_0 \in (0, 1)$ fixed. The corresponding interval estimation problem is to find a $1 - \alpha$ confidence interval for $F(t_0)$ for fixed $\alpha \in (0, 1)$.

EXAMPLE 3 (Panel count data). Suppose that $N_1, \ldots, N_n$ are i.i.d. counting processes on $R^+ = [0, \infty)$ with mean function $\Lambda(t) \equiv EN_1(t)$. When the counting processes N_i are observed only at irregular (random) times $\{T_{ij}\}_{j=1}^{K_i}$, with perhaps a random number of observation times K_i for the ith individual, Wellner and Zhang (2000) have referred to this type of data as "panel count data," and have studied the nonparametric maximum likelihood estimator of the monotone functon Λ. Here our interest focuses on likelihood ratio tests of H_0: $\Lambda(t_0) = \theta_0$ versus H_1: $\Lambda(t_0) \neq \theta_0$ $t_0 \in (0, \infty)$ and $\theta_0 \in R^+$ fixed. The corresponding interval estimation problem is to find a $1 - \alpha$ confidence interval for $\Lambda(t_0)$ for fixed $\alpha \in (0, 1)$.

EXAMPLE 4 (Monotone hazard function with right-censored data). Suppose that (X_i, T_i), $i = 1, \ldots, n$ are i.i.d., and for each pair (X_i, T_i), X_i and T_i are independent with $X_i \sim F$ and $T_i \sim G$ where F and G are distribution functions on $[0, \infty)$. For each pair we observe $Y_i = (X_i \wedge T_i, \Delta_i)$ where $\Delta_i = 1\{X_i \leq T_i\}$. Moreover, suppose that it is known that the distribution F has monotone increasing hazard rate $\lambda = f/(1-F)$. The goal is to make inference about the monotone (increasing) function λ. The nonparametric maximum likelihood estimator $\hat{\lambda}_n$ of λ has been studied by Huang and Zhang (1994), Huang and Wellmer (1995), and, in the uncensored case, by Prakasa Rao (1969). We are interested here in likelihood ratio tests of H_0: $\lambda(t_0) = \theta_0$ versus H_1: $\lambda(t_0) \neq \theta_0$ for $t_0 \in (0, \infty)$ and $\theta_0 \in R^+$ fixed. The corresponding interval estimation problem is to find a $1 - \alpha$ confidence interval for $\lambda(t_0)$ for fixed $\alpha \in (0, 1)$.

EXAMPLE 5 (Monotone regression function). Suppose that (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d. with $Y_i = r(X_i) + \varepsilon_i$ where r: $[0, 1] \to \mathbb{R}$ is an unknown monotone function and ε_i are i.i.d. Gaussian random variables with mean zero, finite variance, and independent of X_i. The least squares estimator of r was studied by Brunk (1970) (under more generality than the above assumptions). We are interested here in likelihood ratio tests of H_0: $r(t_0) = \theta_0$ versus H_1: $r(t_0) \neq \theta_0$ for $t_0 \in (0, \infty)$ and $\theta_0 \in \mathbb{R}$ fixed. The corresponding interval estimation problem is to find a $1 - \alpha$ confidence interval for $r(t_0)$ for fixed $\alpha \in (0, 1)$.

For further examples of this type, see Groeneboom and Wellner (2001).

A common theme in all of these examples is that (under modest assumptions) $n^{1/3}$ times the difference between the maximum likelihood estimator at a point t_0 and the true function at the same point converges in distribution to a constant (depending on the particular problem) times the distribution of the location of the minimum of two-sided Brownian motion plus a parabola. Another equivalent description of the asymptotic distribution is as a constant (the same constant as before divided by 2) times the slope at zero of the greatest convex minorant of two-sided Brownian motion plus a parabola. Since we have neither a $\sqrt{n}$ convergence rate nor a Gaussian limiting distribution for the *MLE* in any of these problems, we do *not* expect a limiting χ^2 distribution for the likelihood ratio statistic, as would be expected in regular parametric and certain semiparametric settings [see e.g. Murphy and Van der Vaart (1997) for the latter].

For example, in Example 2, assuming that F and G have positive densities f and g respectively at t_0, it is known that

$$n^{1/3}(\mathbb{F}_n(t_0) - F(t_0)) \xrightarrow{d} \left\{ \frac{F(t_0)(1 - F(t_0))f(t_0)}{2g(t_0)} \right\}^{1/3} 2\mathbb{Z} \tag{1.1}$$

where $\mathbb{Z} \equiv \text{argmin}(W(t) + t^2)$ and W is two-sided Brownian motion starting from 0; see Groeneboom and Wellner (1992).

Instead, one would expect the limiting distribution to be described by some functional of two sided Brownian motion (in conformity with the limiting distribution of the MLE). This is indeed the case. The limiting distribution of the likelihood ratio statistic is, instead of χ_1^2, a fixed universal distribution described briefly as follows: Let G be the greatest convex minorant of $W(t) + t^2$ for a two-sided Brownian motion W, and let $\mathbb{S}$ denote the corresponding process of slopes. Similarly, let G_0 be the process which, for $t \geq 0$, is the greatest convex minorant of $W(t) + t^2$ for $t \geq 0$ subject to the constraint that its slopes stay greater than or equal to zero, and, for $t < 0$, is the greatest convex minorant of $W(t) + t^2$ for $t < 0$ subject to the constraint that its slopes stay less than or equal to zero. Let $\mathbb{S}_0$ denote the corresponding process of slopes. Then the limiting distribution we expect for the the likelihood ratio statistics in Examples 1-5 is exactly that of

$$\mathbb{D} \equiv \int \left\{ \mathbb{S}^2(t) - \mathbb{S}_0^2(t) \right\} dt. \tag{1.2}$$

We deal here in complete detail with the interval censoring model discussed above in Example 2. We show that the likelihood ratio statistic in this problem does indeed have limiting distribution given by $\mathbb{D}$ in (1.2) under the null hypothesis. Note that this is, as for the usual χ^2 limit obtained in regular parametric problems, universal: it does not depend on θ_0, t_0, or any of the parameters of the particular problem. The universality of the limiting distribution is useful not only in devising an asymptotic test for the null hypothesis, but is also important in constructing approximate level $1-\alpha$ confidence sets for the parameter of interest. Note that to construct an approximate confidence interval for $F(t_0)$ from (1.1), we must contend with the awkward problem of estimating the unknown parameter $\{F(t_0)(1 - F(t_0))f(t_0)/(2g(t_0))\}^{1/3}$ appearing on the right side. This entails smoothing to estimate $f(t_0)$ and $g(t_0)$.

By forming the likelihood ratio statistic (which entails estimation of the distribution function under the null hypothesis with no smoothing involved), we get a universal limiting distribution, thereby completely avoiding the smoothing issue. This is a strong advantage of the likelihood ratio method of constructing confidence sets.

For use of likelihood ratio methods in a related problem involving monotone functions, see Wu, Woodroofe and Mentz (2001).

The rest of the paper is organized as follows: Section 2 gives statements of our results for the interval censoring problem, including the limiting distribution of the likelihood ratio statistic under both the null hypothesis and local (contiguous) alternatives, and the consistency of the test under fixed alternatives. Section 3 gives the distribution of $\mathbb{D}$ as estimated by Monte-Carlo methods. Section 4 shows how we can use the results of Sections 2 and 3 to obtain confidence intervals for $F(t_0)$ in the context of Example 2. In Section 5 we give a brief discussion of further results, a heuristic discussion of why we expect the same limiting null distribution in Examples 1 and 3-5, and open problems. Proofs or proof sketches for the results in Sections 2 and 4 are given in Section 6.

2. The interval censoring problem: statements of results.

2.1. *The model.* The density of the pair (T, Δ) with respect to the measure $G \times$ Counting measure on the product space $\mathbb{R}^+ \times \{0, 1\}$ is given by

$$p_F(t, \delta) = F(t)^{\delta}(1 - F(t))^{1-\delta}.$$

Hence the log-likelihood for n observations is given by:

$$\begin{aligned}
&\log L_n(F, Y_1, Y_2, \dots, Y_n)\\
&\quad = \sum_{i=1}^{n} \{\Delta_i \log F(T_i) + (1 - \Delta_i) \log(1 - F(T_i))\} \\
&\quad = \sum_{i=1}^{n} \{\Delta_{(i)} \log F(T_{(i)}) + (1 - \Delta_{(i)}) \log(1 - F(T_{(i)}))\}
\end{aligned} \tag{2.1}$$

where $T_{(1)}, \dots, T_{(n)}$ are the ordered T_i's and $\Delta_{(i)}$ is the indicator corresponding to $T_{(i)}$. Let $\mathbb{F}_n$ denote the MLE of F under no constraints, and let $\mathbb{F}_n^0$ be the MLE of F under the constraint that the value of F at the point t_0 equals θ_0. The unconstrained MLE $\mathbb{F}_n$ is well characterized in this situation; see e.g. Groeneboom and Wellner (1992). Note that from the expression for L_n it is clear that both $\mathbb{F}_n$ and $\mathbb{F}_n^0$ are determined uniquely only up to their values at the observed T_i 's (of course $\mathbb{F}_n^0$ is determined at the point t_0).

2.2. *The estimators and the likelihood ratio.* The likelihood ratio in this problem is given by

$$\lambda_n = \frac{\sup_F L_n(F, Y_1, Y_2, \dots, Y_n)}{\sup_{F(t_0)=\theta_0} L_n(F, Y_1, Y_2, \dots, Y_n)} = \frac{L_n(\mathbb{F}_n, Y_1, Y_2, \dots, Y_n)}{L_n(\mathbb{F}_n^0, Y_1, Y_2, \dots, Y_n)}.$$

The log-likelihood ratio statistic, namely twice the log-likelihood ratio, can therefore be writtten as:

$$2\log\lambda_n = 2\log L_n(\mathbb{F}_n, Y_1, Y_2, \ldots, Y_n) - 2\log L_n(\mathbb{F}_n^0, Y_1, Y_2, \ldots, Y_n) \tag{2.2}$$

In order to compute the likelihood ratio statistic we need to characterize the MLE's, $\mathbb{F}_n$ (the unconstrained MLE) and $\mathbb{F}_n^0$ (the constrained MLE).

Characterization and computation of the unconstrained maximum likelihood estimator $\mathbb{F}_n$ is well understood; see e.g. Groeneboom and Wellner (1992), pages 35–50. Our main object here is to briefly give a characterization of the constrained maximum likelihood estimator. Because these finite sample results are important for an understanding of the corresponding asymptotic versions of the problem, we give statements of them here. For the unconstrained problem $0 \le F(T_{(1)}) \le F(T_{(2)}) \le \cdots \le F(T_{(n)}) \le 1$, and hence it suffices to find $0 \le w_1 \le w_2 \le \cdots \le w_n \le 1$ so as to maximize:

$$\phi(w) = \sum_{i=1}^{n} \{\Delta_{(i)} \log(w_i) + (1-\Delta_{(i)}) \log(1-w_i)\}.$$

For the constrained maximization problem we want to maximize the likelihood (2.1) over the class of distributions F satisfying $F(t_0) = \theta_0$. Recall that $0 \le \theta_0 \le 1$. Let m be such that $T_{(m)} \le t_0 \le T_{(m+1)}$. For any F in the above class we then have $F(T_{(m)}) \le \theta_0 \le F(T_{(m+1)})$ and denoting $F(T_{(i)})$ by w_i as before the problem reduces to maximizing

$$\begin{aligned}\phi(w_1, w_2, \ldots, w_n) &= \sum_{i=1}^{m} \{\Delta_{(i)} \log(w_i) + (1-\Delta_{(i)}) \log(1-w_i)\} \\ &\qquad + \sum_{i=m+1}^{n} \{\Delta_{(i)} \log(w_i) + (1-\Delta_{(i)}) \log(1-w_i)\} \\ &\equiv \phi_L(w_L) + \phi_R(w_R)\end{aligned}$$

over the set

$$0 \le w_1 \le w_2 \le \cdots \le w_m \le \theta_0 \le w_{m+1} \le \cdots \le w_n \le 1.$$

Note that m itself is random and that m/n tends to $G(t_0)$ almost surely. It suffices to maximize ϕ_L and ϕ_R separately. In fact one only needs to address the problem of maximizing ϕ_L, since maximizing ϕ_R can be reduced to a corresponding "left" maximization problem. Note that to maximize ϕ_L we need to address the following problem: Given indicators $\Delta_{(1)}, \ldots, \Delta_{(m)}$, for some m we need to find $0 \le w_1 \le w_2 \le \cdots \le w_m \le \theta_0 < 1$ so as to maximize $\phi_L(w_L)$.

Necessary and sufficient conditions characterizing the unconstrained and constrained maximizing vectors $\hat{w}$ and $\hat{w}^0$ are given by the following theorem:

THEOREM 2.1 (Characterization of the unconstrained and constrained MLEs). *Suppose that* $\Delta_{(1)} = 1$ *and* $\Delta_{(n)} = 0$. *Then* $\hat{w}$ *maximizes* ϕ *over* w *satisfying* $0 < w_1 \le w_2 \le \cdots \le w_n < 1$ *if and only if the following two con-*

ditions are satisfied (and further the maximizer $\hat{w}$ is uniquely determined by these two conditions):

$$\sum_{j\le i}\left\{\frac{\Delta_{(j)}}{\hat{w}_j}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j}\right\}\ge 0,\qquad i=1,2,\dots,n \tag{2.3}$$

and

$$\sum_{i=1}^{n}\left\{\frac{\Delta_{(j)}}{\hat{w}_j}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j}\right\}\hat{w}_j=0\,. \tag{2.4}$$

Furthermore $\hat{w}_L\equiv(\hat{w}_1^0,\dots,\hat{w}_m^0)$ maximizes ϕ_L over w satisfying $0<w_1\le w_2\le\cdots\le w_m\le\theta_0$ if and only if the following two conditions are satisfied (and further the maximizer $\hat{w}_L$ is uniquely determined by these two conditions):

$$\sum_{j\le i}\left\{\frac{\Delta_{(j)}}{\hat{w}_j^0}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j^0}\right\}\ge 0,\qquad i=1,2,\dots,m; \tag{2.5}$$

$$\sum_{i=1}^{m}\left\{\frac{\Delta_{(j)}}{\hat{w}_j^0}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j^0}\right\}\hat{w}_j^0=\theta_0\sum_{i=1}^{m}\left\{\frac{\Delta_{(j)}}{\hat{w}_j^0}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j^0}\right\}. \tag{2.6}$$

PROOF. The first part follows from Groeneboom and Wellner (1992), pages 39 -40. To prove the second part, let $S_0=0$ and

$$S_i\equiv\sum_{j\le i}\left\{\frac{\Delta_{(j)}}{\hat{w}_j^0}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j^0}\right\},\qquad i=1,\dots,m.$$

Now by concavity of ϕ_L and convexity of $V_L\equiv\{w\in\mathbb{R}^m\colon 0\le w_1\le\ldots\le w_m\le\theta_0\}$, $\hat{w}_L$ maximizes ϕ_L over V_L if and only if

$$\frac{d}{dt}\phi_L((1-t)\hat{w}_L+tw)\Big|_{t=0}=\sum_{j=1}^{m}\left\{\frac{\Delta_{(j)}}{\hat{w}_j^0}-\frac{1-\Delta_{(j)}}{1-\hat{w}_j^0}\right\}(w_j-\hat{w}_j^0) \tag{2.7}$$

$$=-\sum_{i=1}^{m}S_i\{w_{i+1}-w_i-(\hat{w}_{i+1}^0-\hat{w}_i^0)\} \tag{2.8}$$

$$\le 0\qquad\text{for all } w\in V_L\,; \tag{2.9}$$

where (2.8) holds by summation by parts with $\hat{w}_{m+1}^0=w_{m+1}=\theta_0$. When $\Delta_{(1)}=1$ (so $\hat{w}_1^0>0$), let $w=\hat{w}_L-\varepsilon\underline{1}_i\in V_L$ where $\underline{1}_i=(1,\dots,1,0,\dots,0)$ is the vector with 1 in the first i positions and 0 in the remaining $m-i$ positions. Taking this choice of w in (2.7) shows that $S_i\ge 0$; that is, (2.5) holds. Taking $w_i=\theta_0$ for all i in (2.8) shows that $\sum_{i=1}^m S_i(\hat{w}_{i+1}^0-\hat{w}_i^0)\le 0$, and together with $S_i\ge 0$ and $\hat{w}_{i+1}^0-\hat{w}_i^0\ge 0$ this yields

$$\sum_{i=1}^{m}S_i(\hat{w}_{i+1}^0-\hat{w}_i^0)=0. \tag{2.10}$$

But (2.10) is equivalent to (2.6) via summation by parts. For sufficiency, note that (2.5) and (2.6) imply that (2.9) holds. For further details see Banerjee (2000) and Banerjee and Wellner (2000). Alternatively, see Van Eeden (1957a), Van Eeden (1957b), and Barlow, Bartholomew, Bremner and Brunk (1972), pages 56–57. □

For both the unconstrained and the constrained MLE there are important geometric interpretations which we now briefly describe: Define $H^\star : [0, n] \to R$ as

$$H^\star(t) = \sup\left\{H(t) : H(i) \le \sum_{j\le i} \Delta_{(j)} \text{ for } i = 0, 1, \dots, n; H(0) = 0; H \text{ is convex}\right\}.$$

Here by convention $\Delta_{(0)} = 0$. The function $H^\star$ is the *Greatest Convex Minorant* (GCM) of the points $(i, \sum_{j\le i} \Delta_{(j)})$ on $[0, n]$. In other words it is the greatest convex function on $[0, n]$ whose graph lies below that obtained by joining the points $(i, \sum_{j\le i} \Delta_{(j)})$ successively by means of straight lines (the pointwise supremum of a collection of convex functions gives a convex function). Alternatively $H^\star$ can also be thought of as the greatest convex minorant of the left-continuous function $\tilde{H}$ which assumes the value 0 at the point 0 and on the interval $(i-1, i]$ assumes the value $\sum_{j\le i} \Delta_{(j)}$. Now let $\hat{w}_i$ be the left derivative of $H^\star$ at i. Then $(\hat{w}_1, \dots, \hat{w}_n)$ is the unique vector maximizing $\phi(w)$ subject to the monotonicity constraints. For a proof see Groeneboom and Wellner (1992), page 41. Furthermore, an explicit expression for $\hat{w}$ is given by the following "max-min" formula:

$$\hat{w}_m = \max_{i\le m} \min_{k\ge m} \frac{\sum_{i\le j\le k} \Delta_{(j)}}{k - i + 1}.$$

The geometric interpretation of the solution $\hat{w}_L$ is easily obtained in parallel to the discussion of the unconstrained solution: Define $H_L^\star : [0, m] \to R$ as

$$H_L^\star(t) = \sup\left\{H(t) : H(i) \le \sum_{j\le i} \Delta_{(j)} \text{ for } i = 0, 1, \dots, m; \right.$$
$$\left. H(0) = 0; H \text{ is convex}\right\}.$$

Here by convention $\delta_{(0)} = 0$. The function $H_L^\star$ is the *Greatest Convex Minorant* (GCM) of the points $(i, \sum_{j\le i} \Delta_{(j)})$ on $[0, m]$. In other words it is the greatest convex function on $[0, m]$ whose graph lies below that obtained by joining the points $\{(i, \sum_{j\le i} \Delta_{(j)})\}_{j=1}^m$ successively by means of straight lines Now let $\tilde{w}_i$ be the left derivative of $H_L^\star$ at i, and set $\hat{w}_i^0 = \min\{\tilde{w}_i, \theta_0\}$. Then $(\hat{w}_1^0, \hat{w}_2^0, \dots, \hat{w}_m^0)$ is the unique vector maximizing $\phi_L(w)$ subject to the monotonicity constraints and $w_m \le \theta_0$. In words, we form the greatest convex minorant of the cumulative sum diagram on $[0, m]$ (corresponding to $T_{(j)}$'s less than or equal to θ_0), and find the left derivatives thereof; when these slopes exceed θ_0 we simply truncate them to θ_0. We will use the notation $\tilde{\mathbb{F}}_n(t)$ for the function obtained from $\tilde{\mathbb{F}}_n(T_{(i)}) = \tilde{w}_i$, $i = 1, \dots, m$, and $\mathbb{F}_n^0(t)$ for the function obtained from

$\mathbb{F}_n^0(T_{(i)})$, $i = 1, \ldots, m$, with corresponding natural definitions to the right of t_0. (For the interval $[T_{(m)}, t_0]$, we define both $\tilde{\mathbb{F}}_n$ and $\mathbb{F}_n^0$ by extending the value from $T_{(m)}$ and jumping to θ_0 if necessary in the case of $\mathbb{F}_n^0$.)

To maximize ϕ_R we are addressing the same problem as in maximizing ϕ_L except for the fact that now the class of vectors over which we maximize satisfies $0 < \theta_0 \le w_{m+1} \le w_{m+2} \le \cdots \le w_n \le 1$. Now, setting $u_j = 1 - w_{n+1-j}$ and $\gamma_{(j)} = 1 - \Delta_{(n+1-j)}$ we note that maximizing $\phi_R(w_R)$ subject to $0 < \theta_0 \le w_{m+1} \le w_{m+2} \le \cdots \le w_n \le 1$ is the same as maximizing $\tilde{\phi}_R(u_1, u_2, \ldots, u_{n-m})$, where

$$\tilde{\phi}_R(u_1, u_2, \ldots, u_{n-m}) = \sum_{i=1}^{n-m} \{\gamma_{(i)} \log u_i + (1 - \gamma_{(i)}) \log(1 - u_i)\}$$

subject to $0 \le u_1 \le u_2 \le \cdots \le u_{n-m} \le 1 - \theta_0 < 1$. Once the maximizer $\hat{u}$ has been obtained, the corresponding $\hat{w}_R$ can be recovered from the relation $w_j = 1 - u_{n+1-j}$. So the constrained maximizer $\mathbb{F}_n^0$ is evaluated in two pieces; $(\hat{w}_1, \ldots, \hat{w}_m)$ and $(\hat{w}_{m+1}, \ldots, \hat{w}_n)$ being obtained separately. The key is the geometric picture of the solution $\hat{w}_R$: form the greatest convex minorant of the cumulative sum diagram $\{(i, \sum_{j \le i} \Delta_{(j)})\}_{i=m}^n$ on $[m, n]$, and the corresponding left derivatives. If these drop below θ_0, we simply replace them by θ_0.

These characterizations are illustrated by Figures 1 and 2. [These figures were generated from $F =$Exponential(1), $G =$Uniform(0, 3) and $n = 30$. For the constrained case we chose t_0 satisfying $F(t_0) = 2/3$. The estimators $\mathbb{F}_n$ and $\mathbb{F}_n^0$ turned out to be as follows:

$\mathbb{F}_n$: $2/5(1-5), 5/11(6-16), 4/5(17-21), 1(22-30)$.

$\mathbb{F}_n^0$: $2/5(1-5), 1/2(6-11), 2/3(12-16), 4/5(17-21), 1(21-30)$.]

2.3. *Asymptotic properties of the estimators.* To describe the asymptotic properties of the unconstrained and constrained estimators $\mathbb{F}_n$ and $\mathbb{F}_n^0$, we first describe several processes connected with the natural limiting problem. Let W denote a standard two-sided Brownian motion process starting from zero, and for positive constants a and b, define the process $X_{a,b}$ by $X_{a,b}(t) \equiv aW(t) + bt^2$. The greatest convex minorant $G_{a,b}$ of $X_{a,b}$ on R is characterized by the following theorem.

THEOREM 2.2 (Greatest convex minorant of $X_{a,b}$). *The greatest convex minorant $G_{a,b}$ of $X_{a,b}$ exists and is characterized by the following conditions*:

(i) *The function $G_{a,b}$ is everywhere below the function $X_{a,b}$*:

$$G_{a,b}(t) \le X_{a,b}(t) \qquad \text{for all } t \in \mathbb{R}. \tag{2.11}$$

(ii) *$G_{a,b}$ has a monotone (right) derivative $g_{a,b}$.*

(iii) *The function $G_{a,b}$ and its (right) derivative $g_{a,b}$ satisfy*

$$\int_R \{X_{a,b}(t) - G_{a,b}(t)\}\, dg_{a,b}(t) = 0. \tag{2.12}$$

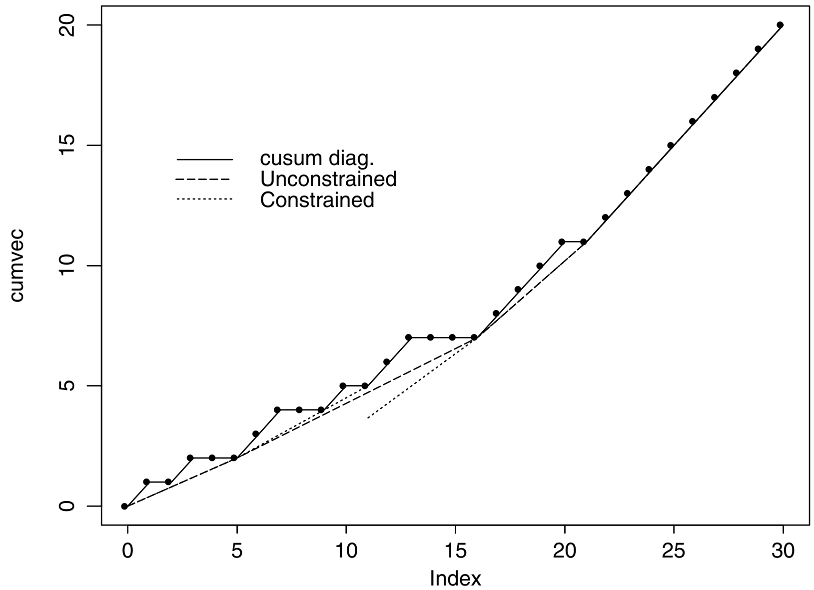

FIG. 1. *Cumulative sum diagram with left, right, and global Greatest convex minorants.*

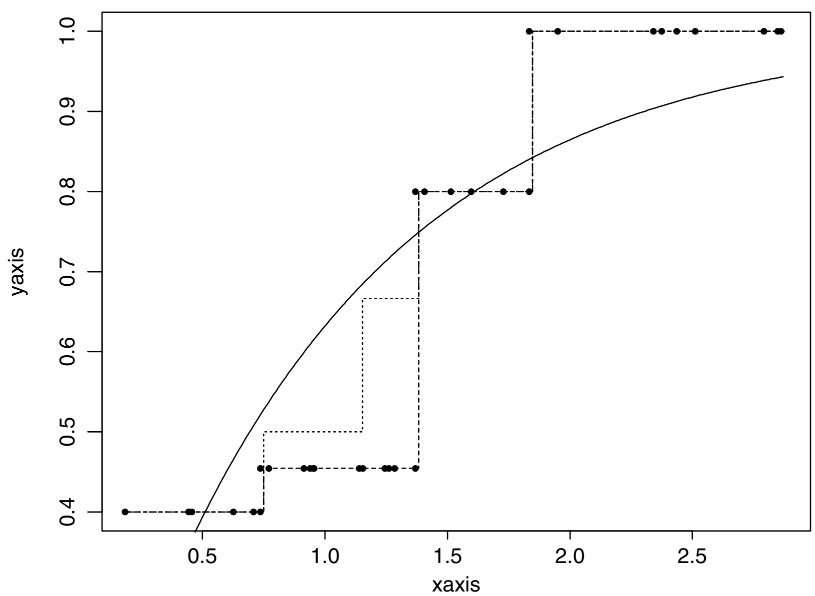

FIG. 2. *The unconstrained estimator* $\mathbb{F}_n$ *and constrained estimator* $\mathbb{F}_n^0$. *Solid line*: $1 - \exp(-x)$. *Big dots*: *unconstrained estimator*. *Small dots*: *constrained estimator*.

The picture of $G_{a,b}$ and $g_{a,b}$ which emerges from Theorem 2.2 is as follows: The greatest convex minorant process $G_{a,b}$ is piecewise linear with changes of slope at isolated points where it touches $X_{a,b}$; thus the slope process $g_{a,b}$ is piecewise constant, with jumps only at points s where $G_{a,b}(s) = X_{a,b}(s)$.

The slope process $g_{a,b}$ can be viewed as the unconstrained estimator of the monotone function $2bt$ based on observation of $X_{a,b}$ (which we can think of as $dX_{a,b}(t) = 2bt + a\,dW(t)$). On the other hand, we can consider a constrained estimator $g^0_{a,b}$ of $2bt$ based on observation of $X_{a,b}$ which uses the knowledge that the "true" monotone function is zero at $t = 0$. The corresponding "constrained convex minorants" of $X_{a,b}$ are characterized in the following theorem:

THEOREM 2.3 (Constrained greatest convex minorants of $X_{a,b}$). *The constrained greatest convex minorants $G^0_{a,b}$ of $X_{a,b}$ exist and are characterized by the following conditions*:

(i) *The function $G^0_{a,b}$ is everywhere below the function $X_{a,b}$*:

$$G^0_{a,b}(t) \le X_{a,b}(t) \qquad \text{for all } t \in \mathbb{R}. \tag{2.13}$$

(ii) *$G^0_{a,b}$ has a monotone* (*right*) *derivative $g^0_{a,b}$ satisfying $g^0_{a,b}(0) = 0$.*

(iii) *The function $G^0_{a,b}$ and its* (*right*) *derivative satisfy*

$$\int_R \left\{ X_{a,b}(t) - G_{a,b}(t) \right\} dg^0_{a,b}(t) = 0\,. \tag{2.14}$$

The picture of $G^0_{a,b}$ and $g^0_{a,b}$ which emerges from Theorem 2.3 parallels the situation for the constrained estimator in Theorem 2.1 and is as follows: for $t \le 0$ we form the greatest convex minorant $\tilde{G}_L(t)$ of the process $X_{a,b}(t)$, $t \le 0$; when its corresponding slope process $\tilde{g}_L(t)$ exceeds zero, we replace the slopes by 0 (and replace $\tilde{G}_L$ by the appropriate constant value from there to $t = 0$). Similarly, for $t > 0$ we form the greatest convex minorant $\tilde{G}_R(t)$ of the process $X_{a,b}(t)$, $t > 0$; when its corresponding slope process $\tilde{g}_R(t)$ decreases below zero as t decreases to 0, we replace the slopes by 0 (and replace $\tilde{G}_R$ by the appropriate constant value from there to $t = 0$). The resulting process is $G^0_{a,b}$ with slope process $g^0_{a,b}$. Note that $g^0_{a,b}(0) = 0$ and, from results of Groeneboom (1983), $g^0_{a,b}$ is continuous at 0 almost surely, while $G^0_{a,b}$ has a jump discontinuity at 0.

Figures 3 and 4 illustrate Theorems 2.2 and 2.3.

Now we can describe the joint limiting distributions of the unconstrained and unconstrained estimators $\mathbb{F}_n$ and $\mathbb{F}^0_n$. Here is our basic assumption:

ASSUMPTION A. *Suppose that F and G are fixed distributions with continuous* (*Lebesgue*) *densities f and g in a neighborhood of the fixed point t_0 with $F(t_0) \in (0, 1)$, $0 < f(t_0) < \infty$ and $0 < g(t_0) < \infty$.*

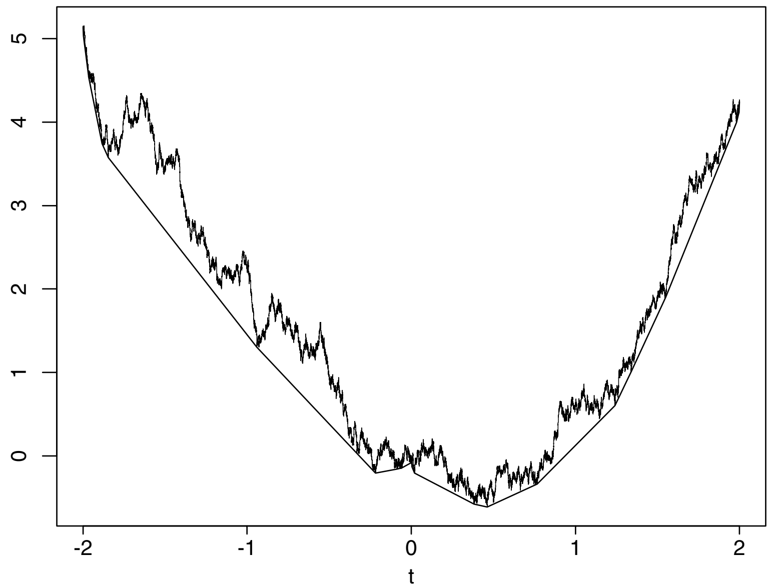

FIG. 3. *The one-sided convex minorants* $\tilde{G}_L$ *and* $\tilde{G}_R$ *and* $W(t) + t^2$.

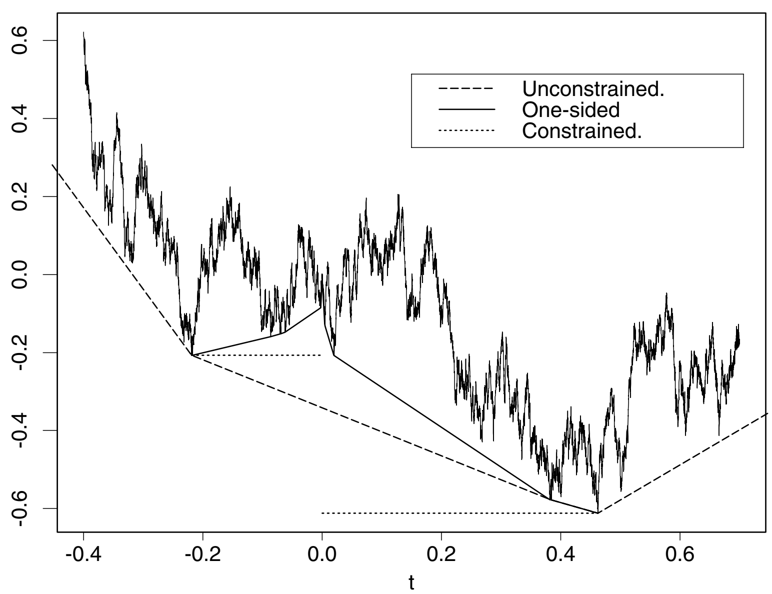

FIG. 4. *Close-up view of* $G_{1,1}$, $\tilde{G}_{L,R}$, $G^0_{1,1}$ *and* $W(t) + t^2$.

THEOREM 2.4 (Asymptotic distributions for the estimators). A. (*At a point* $t \neq t_0$.) *Suppose that* F *and* G *have positive continuous densities* f *and* g *respectively at* $t \neq t_0$). *Then*

$$\left(n^{1/3}(\mathbb{F}_n(t) - F(t)), (n^{1/3}(\mathbb{F}_n^0(t) - F(t))\right) \stackrel{d}{\to} (\mathbb{T}, \mathbb{T})$$

where

$$\mathbb{T} \equiv \left(\frac{4F(t)(1 - F(t)f(t)}{g(t)}\right)^{1/3} \arg\min\{W(h) + h^2\}\,. \tag{2.15}$$

Consequently

$$n^{1/3}\left(\mathbb{F}_n(t) - \mathbb{F}_n^0(t)\right) \stackrel{p}{\to} 0. \tag{2.16}$$

B. (*In* $n^{-1/3}$ *neighborhoods of* t_0.) *Suppose Assumption* A *holds. Define processes* X_n *and* Y_n *by*

$$X_n(t) = n^{1/3}(\mathbb{F}_n(t_0 + tn^{-1/3}) - F(t_0))$$

and

$$Y_n(t) = n^{1/3}(\mathbb{F}_n^0(t_0 + tn^{-1/3}) - F(t_0))\,.$$

Then the finite dimensional marginals of the processes $(X_n(t), Y_n(t))$, *converge to the finite dimensional marginals of the process* $(1/g(t_0))(g_{a,b}(t), g^0_{a,b}(t))$ *where* $a \equiv \sqrt{F(t_0)(1 - F(t_0))g(t_0)}$, $b \equiv f(t_0)g(t_0)/2$, *and the slope processes* $g_{a,b}$ *and* $g^0_{a,b}$ *are described in Theorems* 2.2 *and* 2.3. *Furthermore, for* $p \geq 1$,

$$(X_n(t), Y_n(t)) \stackrel{d}{\to} (1/g(t_0))\,(g_{a,b}(t), g^0_{a,b}(t))$$

in $\mathscr{L}^p[-K, K] \times \mathscr{L}^p[-K, K]$, *for each* $K > 0$.

2.4. *The likelihood ratio statistic under* H_0. Now we can state the main theorem of this paper concerning the asymptotic distribution of the likelihood ratio statistic $2\log(\lambda_n)$ given in (2.2) under the null hypothesis. For the particular values $a = b = 1$, the corresponding slope processes $g_{1,1}$ and $g^0_{1,1}$ in Theorems 2.2 and 2.3 will be denoted by $g_{1,1} \equiv \mathbb{S}$ and $g^0_{1,1} \equiv \mathbb{S}^0$.

THEOREM 2.5 (Asymptotic distribution of $2\log\lambda_n$ under H_0). *Suppose that Assumption* A *holds. Suppose that* F *satisfies the null hypothesis* $H_0 : F(t_0) = \theta_0$. *Then the likelihood ratio statistic* $2\log(\lambda_n)$ *given in* (2.2) *satisfies*

$$2\log(\lambda_n) \stackrel{d}{\to} \int \left((\mathbb{S}(z))^2 - (\mathbb{S}^0(z))^2\right) dz \equiv \mathbb{D}$$

where $g_{1,1} \equiv \mathbb{S}$ *and* $g^0_{1,1} \equiv \mathbb{S}^0$ *are as defined in Theorems* 2.2 *and* 2.3.

2.5. *The likelihood ratio statistic under local (contiguous) alternatives.* We need some further assumptions to handle local alternatives:

Suppose that $\{F_n\}$ is a sequence of continuous distribution functions satisfying the following conditions:

ASSUMPTION B(1). For some $c > 0$, $F_n(t) = F(t)$ for all t with $|t - t_0| \geq cn^{-1/3}$.

ASSUMPTION B(2). The functions $A_n(t) = n^{1/3}(F_n(t) - F(t))$ satisfy

$$A_n(t_0 + n^{-1/3}z) \equiv B_n(z) \to B(z) \equiv f(t_0)K(z)$$

uniformly for $z \in [-c, c]$. [Thus B and K are continuous functions on $[-c, c]$ and both B and K vanish on $(-c, c)^c$.]

THEOREM 2.6. *Suppose that F and G satisfy Assumption* A, *and $\{F_n\}$ is a sequence of distribution functions satisfying Assumptions* B(1) *and* B(2). *Consider the sequence of probability measures $\{P^n_{F_n,G}\}$ and $\{P^n_{F,G}\}$. Then, under $P^n_{F,G}$, the local log-likelihood ratio*

$$\log(L_n(F_n)/L_n(F)) \xrightarrow{d} N(-\sigma^2/2, \sigma^2)$$

where

$$\sigma^2 = \frac{g(t_0)}{F(t_0)(1 - F(t_0))}\int B^2(z)\,dz = \frac{f^2(t_0)g(t_0)}{F(t_0)(1 - F(t_0))}\int K^2(z)\,dz. \tag{2.17}$$

This, in particular, implies that the sequence $\{P^n_{F_n,G}\}$ and the sequence $\{P^n_{F,G}\}$ are mutually contiguous.

To state our main result concerning the behavior of the likelihood ratio statistic under local alternatives requires some further notation. First we define

$$\Psi(z) = \begin{Bmatrix} g(t_0)\displaystyle\int_0^{z\wedge c} B(y)\,dy, & z \geq 0 \\ -g(t_0)\displaystyle\int_{z\vee -c}^0 B(y)\,dy, & z < 0 \end{Bmatrix}$$
$$= \begin{Bmatrix} f(t_0)g(t_0)\displaystyle\int_0^{z\wedge c} K(y)\,dy, & z \geq 0 \\ -f(t_0)g(t_0)\displaystyle\int_{z\vee -c}^0 K(y)\,dy, & z < 0 \end{Bmatrix}.$$

Clearly Ψ is continuous and constant outside of $[-c, c]$. Also $\Psi(0) = 0$. Now consider the processes

$$X_{a,b,\Psi}(z) \equiv aW(z) + bz^2 + \Psi(z);$$

we will be primarily interested in this process for Ψ defined above,

$$a = \sqrt{F(t_0)\,(1 - F(t_0))g(t_0)}, \qquad b = f(t_0)\,g(t_0)/2,$$

and for the "canonical parameters" $a = 1$, $b = 1$, and Ψ replaced by $(b/a)^{4/3}\Psi(\cdot/(b/a)^{2/3})$.

Our first limiting result under the local alternatives concerns the behavior of the processes X_n and Y_n as in Theorem 2.4.

THEOREM 2.7. *Suppose that the distribution functions F, G satisfy Assumption* A, *and the sequence of distribution functions $\{F_n\}$ satisfies Assumptions* B(1) *and* B(2). *Then the finite dimensional marginals of the process $(X_n(t), Y_n(t))$, considered as a process in the space $\mathscr{L}^p[-K, K] \times \mathscr{L}^p[-K, K]$ converge to the finite dimensional marginals of the process $(1/g(t_0))\,(g_{a,b,\Psi}(t), g^0_{a,b,\Psi}(t))$ under the sequence of (contiguous) alternatives $\{P^n_{F_n, G}\}$. Furthermore it is also the case that under this sequence, for any $p \geq 1$,*

$$\big(X_n(t), Y_n(t)\big) \stackrel{d}{\to} \big(1/g(t_0)\big)\,(g_{a,b,\Psi}(t), g^0_{a,b,\Psi}(t))$$

in $\mathscr{L}^p[-K, K] \times \mathscr{L}^p[-K, K]$, for each $K > 0$.

With Theorem 2.7 in hand, we can state our result concerning the asymptotic behavior of the likelihood ratio statistics under local alternatives:

THEOREM 2.8. *Suppose that the hypotheses of Theorem* 2.6 *hold. Then, under the local alternatives $\{P^n_{F_n, G}\}$, the likelihood ratio statistics converge in distribution as follows:*

$$2\log\lambda_n \stackrel{d}{\to} \frac{1}{g(t_0)F(t_0)(1-F(t_0))}\int_{D_{a,b,\Psi}} \big((g_{a,b,\Psi}(z))^2 - (g^0_{a,b,\Psi}(z))^2\big)\,dz \tag{2.18}$$

$$\stackrel{d}{=} \int \big((g_{1,1,\Psi^0_{a,b}}(z))^2 - (g^0_{1,1,\Psi^0_{a,b}}(z))^2\big)\,dz. \tag{2.19}$$

where

$$\Psi^0_{a,b}(t) \equiv (b/a)^{4/3}\Psi((a/b)^{2/3}t).$$

2.6. *The likelihood ratio statistic under a fixed alternative.* In Section 2.4 we stated our main result for the asymptotic distribution of the log-likelihood ratio when the underlying distribution belongs to the null hypothesis, or in other words satisfies $F(t_0) = \theta_0$. Here we study the behavior of the log-likelihood ratio when the true distribution is in the alternative hypothesis. Hence $F(t_0) \neq \theta_0$. We will assume that t_1 satisfies $F(t_1) = \theta_0$ (and that this point is unique).

THEOREM 2.9 (Asymptotic behavior of $2\log\lambda_n$ under fixed alternatives). *Suppose that $F(t_0) \neq \theta_0$, and there is a neighborhood A of t_0 such that F and G are continuously differentiable on A with densities f and g respectively, and $f(t_0)$ and $g(t_0)$ are both positive. Moreover, suppose that there is* *some open interval* (c, d) *with* $[t_0 \wedge t_1, t_0 \vee t_1] \subset (c, d)$ *and each* $t \in (c, d)$ *is a support point of* G. *Then*

$$\frac{2}{n}\log\lambda_n \xrightarrow{p} 2K(P_{F,G}, P_{H,G}) > 0 \tag{2.20}$$

$$= 2\inf\{K(P_{F,G}, P_{U,G})\colon U \text{ a d.f. with } U(t_0) = \theta_0\} \tag{2.21}$$

where $K(P, Q) = E_P \log(dP/dQ)$ *is the Kullback-Leibler discrepancy between* P *and* Q, *and the distribution function* H *is described as follows*:

$$H(t) = \begin{cases} F(t) \vee \theta_0, & t \geq t_0, \\ F(t) \wedge \theta_0, & t < t_0. \end{cases} \tag{2.22}$$

Theorem 2.9 yields consistency of the likelihood ratio test based on the asymptotic critical values coming from Theorem 2.5: that is, let d_α satisfy $P(\mathbb{D} \geq d_\alpha) = \alpha$ for $0 < \alpha < 1$, and suppose that we reject H_0 when $2\log\lambda_n > d_\alpha$.

COROLLARY 2.1. *If the hypotheses of Theorem* 2.9 *hold*, *then the likelihood ratio test given by* (2.2) *is consistent*; *that is*,

$$P_{F,G}(2\log\lambda_n \geq d_\alpha) \to 1.$$

3. The limiting distribution under H_0: results via simulations. To carry out the tests described in Section 2 or find confidence sets based on the likelihood ratio statistic, we need to know the distribution of $\mathbb{D}$ described in Theorem 2.5, or at least a few selected quantiles thereof. Although it may be possible to use the methods and techniques of Groeneboom (1983) and Groeneboom (1988) to find this distribution analytically, we will leave this problem for future research. Here we give estimates of the distribution of $\mathbb{D}$ by two different methods.

SIMULATION METHOD 1. The first method involves simply estimating the distribution of $\mathbb{D}$ by using Theorem 2.5: we simply compute the log-likelihood ratio statistic many times $M = 10^4$ for a large sample size $n = 10^4$. In the particular two cases we chose, the distribution F was Exponential(1) or Weibull with shape parameter 2 and scale parameter 1, while the distribution G was Uniform(0, 2), and we chose $t_0 = \log(2)$ or $t_0 = (\log(2))^{1/2}$, so that $\theta_0 = 1/2$ in both cases. Table 1 gives the values of the various constants involved in the two situations studied; we present results in Figure 5 for only the Exponential case.

The resulting empirical distribution of the $M = 10^4$ values of the statistic $2\log\lambda_n$ for the exponential case is shown in Figure 5, together with the empirical distribution from method 2 as explained below.

SIMULATION METHOD 2. In this method we generated discrete approximations to the Brownian motion process W by summing independent standard

TABLE 1

F	t_0	θ_0	$f(t_0)$	$g(t_0)$	a	b
Exponential(1)	log(2)	.5	.5	.5	.3536	.1250
Weibull(1, 2)	$(\log(2))^{1/2}$	.5	.8326	.5	.3536	.2081

normal random variables $\{Z_j, Z'_j\}$ and forming the corresponding partial sum processes

$$W_m(t) \equiv m^{-1/2}\left\{1\{t \geq 0\} \sum_{j=1}^{[mt]} Z_j + 1\{t < 0\} \sum_{j=1}^{[m(-t)]} Z'_j\right\}$$

for $m = 10^4$ and $-2 \leq t \leq 2$. We then generated the process $\mathbb{X}_m(t) \equiv W_m(t)+t^2$ on a grid with step size $\Delta = .0002 = 1/m$ and found the greatest convex minorant G_m, constrained (one-sided) greatest convex minorant(s) G_m^0, and the corresponding slope processes g_m and g_m^0. We then computed the corresponding value of the random variable $\mathbb{D}_m$, repeating this processes $M = 3 \times 10^4$ times. The resulting empirical distribution of all $M = 3 \times 10^4$ values of $\mathbb{D}_m$ is shown in Fig. 5.

Based on these estimators of the distribution of $\mathbb{D}$, our corresponding estimators of several selected quantiles of $F_{\mathbb{D}}$ are shown in Table 2. The last (fifth) column of the table gives an estimate of the standard deviation of the

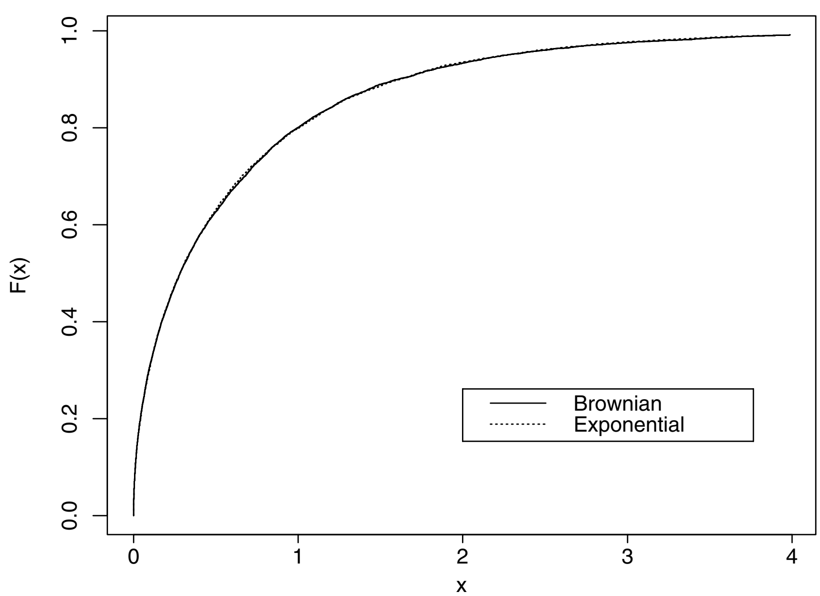

FIG. 5. *Empirical Distributions, Methods* 1 *and* 2, $F = $ *Exponential*.

corresponding empirical quantiles (method 2) in column four. We favor (and have been using) the quantile estimates produced via method 2. For more complete tables, see Banerjee (2001) or Banerjee and Wellner (2000).

4. Pointwise confidence intervals for $F(t_0)$. To form confidence sets for $F(t_0)$, we proceed by inverting the likelihood ratio tests for different values of θ. That is, let $\lambda_n(\theta)$ denote the likelihood ratio for testing $H_0 : F(t_0) = \theta$ versus $H_1 : F(t_0) \neq \theta$. For $0 < \alpha < 1$, let d_α be the upper α quantile of the distribution of $\mathbb{D}$: $P(\mathbb{D} > d_\alpha) = \alpha$. Then an approximate $1 - \alpha$ confidence set $C_{n,\alpha}$ for $F(t_0)$ is given by

$$C_{n,\alpha} \equiv \{\theta: 2\log\lambda_n(\theta) \le d_\alpha\}. \tag{4.1}$$

Suppose that F is the true distribution function and $\theta_0 \equiv F(t_0)$ is the true value of F at t_0. Then the following proposition guarantees that the coverage probability of the sets $C_{n,\alpha}$ is approximately $1 - \alpha$:

PROPOSITION 4.1. *Suppose that F and G have densities f and g which are positive and continuous in a neighborhood of t_0. Then*

$$P_{F,G}(F(t_0) \in C_{n,\alpha}) \to P(\mathbb{D} \le d_\alpha) = 1 - \alpha$$

as $n \to \infty$.

PROOF. Note that

$$P_{F,G}(F(t_0) \in C_{n,\alpha}) = P_{F,G}(2\log\lambda_n(\theta_0) \le d_\alpha) \to P(\mathbb{D} \le d_\alpha) = 1 - \alpha$$

by Theorem 2.5. □

The following proposition guarantees that the sets $C_{n,\alpha}$ are closed intervals bounded away from 0 and 1 if we observe a failure to the left of t_0 and a censored point to the right of t_0.

PROPOSITION 4.2. *Fix $\alpha \in (0, 1)$. If $\sum_{i=1}^{m} \Delta_{(i)} \ge 1$ and $\sum_{i=m+1}^{n}(1 - \Delta_{(i)}) \ge 1$, then the set $C_{n,\alpha}$ defined in* (4.1) *is a closed bounded interval contained in* $(0, 1)$.

For a proof of Proposition 4.2 and a study of the finite sample properties of the confidence intervals, see Banerjee (2000). We illustrate the formation of the confidence sets and Proposition 4.2 in Fig. 6 [in which $n = 3000$, $d_{.05} = 2.269$ from Table 2, and the "true" $F(t_0) = .5$].

5. Discussion: further results and open problems. There are a number of interesting further results and open problems connected with the methods and approaches of this paper. The following paragraphs discuss several of these.

A. Analytic structure of the distribution of $\mathbb{D}$? In Section 3 we presented Monte-Carlo estimates of the distribution of $\mathbb{D}$. It would be very interesting to

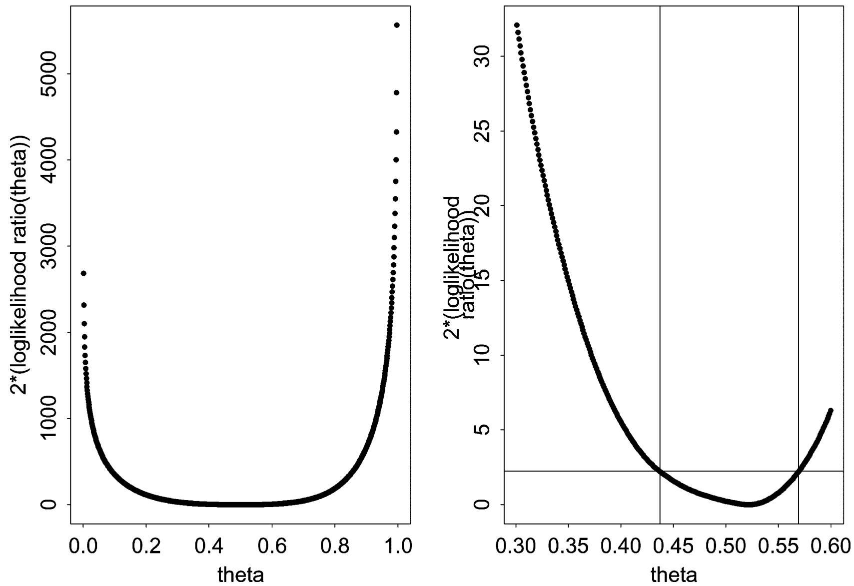

FIG. 6. *Plots of* $\theta \mapsto 2\log\lambda_n(\theta)$ *and resulting* 95% *confidence interval.*

characterize the distribution of $\mathbb{D}$ analytically. This will undoubtedly involve the methods used in both Groeneboom (1983) and Groeneboom (1988).

B. Testing at $k > 1$ *points*? If we consider testing H_0: $F(t_1) = \theta_1, \dots, F(t_k) = \theta_k$ for different time points $t_1 < \cdots < t_k$, $\theta_1 < \cdots < \theta_k$, and $k \geq 2$, then it follows from methods similar to those used here that (under the assumption that F and G have continuous positive derivatives f and g respectively at all t_i, $i = 1, \dots, k$)

$$2\log(\lambda_n) \xrightarrow{d} \mathbb{D}_k \tag{5.1}$$

TABLE 2
Estimated quantiles $\hat{x}_p$ *of the distribution of* $\mathbb{D}$

p	$\hat{x}_p$, Method 1 F = Exponential	$\hat{x}_p$, Method 1 F = Weibull	$\hat{x}_p$, Method 2	Method 2 $\hat{S}_p$
.25	0.06594	0.06204	0.06402	0.00161
.50	0.28310	0.27522	0.28506	0.00361
.75	0.81148	0.79449	0.80694	0.00806
.80	1.00587	0.96737	0.98729	0.00943
.85	1.24393	1.22480	1.22756	0.01178
.90	1.61669	1.61514	1.60246	0.01650
.95	2.24792	2.26465	2.26916	0.02374
.99	3.75947	4.02426	3.83630	0.05471

where $\mathbb{D}_k \stackrel{d}{=} Y_1 + \dots + Y_k$ and the Y_j's are i.i.d. as $\mathbb{D}$. This is somewhat analogous to the familiar results concerning limiting χ_k^2 distributions for log-likelihood ratio tests in regular parametric cases. For a proof of (5.1), see Banerjee (2000).

C. Union-intersection tests: supremum of LR statistics? For completely observed data, the supremum of point-wise (binomial-) likelihood ratio tests has been studied by Berk and Jones (1979) and Owen (1995). The confidence bands constructed by Owen by inverting these tests have some very desirable properties. It would be very interesting to study the asymptotic distribution of the supremum of the log-likelihood ratio statistics $2 \log \lambda_n(F_0(t))$ as a process in t with a view toward construction of confidence bands for F by inversion of the tests.

D. Other problems of this (*monotone function*) *type*? In the introduction we introduced several other problems of the same basic type studied here. In each of these problems, the unconstrained estimator is defined in terms of the slopes of the greatest convex minorant of a certain cumulative-sum diagram, and the limiting distribution of the unconstrained estimator at a fixed point is (under a positive curvature assumption) that of a constant (typically $a\,(b/a)^{1/3}$ divided by a constant from the localization of the $x-$axis for the cumulative-sum diagram) times the slope at zero of the greatest convex minorant of the sum of two-sided Brownian motion and a parabola. We conjecture that in problems 1 and 3–5 the behavior of the natural constrained estimators will behave in a way which is (asymptotically at least) the same as the constrained estimators in our current Example 2, and hence that the asymptotic distribution of the likelihood ratio statistic is the same as that obtained in Theorem 2.5 for Example 2. Verification of this conjecture will depend on careful analyses of the constrained estimators in the various examples; we have not yet completed such a detailed study in any of these examples, but have begun a detailed study of Example 1. It would be very interesting to have some unified approach to all of these various problems.

E. Other problems of related type. Groeneboom, Jongbloed and Wellner (2000b) have obtained limiting distributions for the estimation of convex functions. What are the corresponding results for log-likelihood ratio statistics in that (and other related) cases?

6. Proofs for Section 2.

6.1. *Proofs for Subsection* 2.3. Theorems 2.2 and 2.3 can be viewed as natural "continuous" extensions of the two parts of Theorem 2.1 respectively. The proofs proceed by considering the corresponding unconstrained and constrained optimization problems connected with estimation of a montone function based on observation of the Gaussian processes $\{X_{a,b}(t)\colon t \in [-c, c]\}$, and then passing to the limit as $c \to \infty$ exactly as in Groeneboom, Jongbloed and Wellner (2000a) in the case of convex function estimation. We will not give these proofs here; they are given in some detail in Wellner (2001). Instead we will focus on the proofs of Theorems 2.4–2.9.

Two slightly different approaches have been developed for proving results such as Theorem 2.4.

The first of these, a type of continuous mapping approach, was initiated by Prakasa Rao (1969), used by Brunk (1970) and developed further by Leurgans (1982), and Huang and Zhang (1994); in particular, see Leurgans (1982) [Theorem 2.1, page 289] and Huang and Zhang (1994) [Lemma 4, page 1265]. This approach has the merit of conceptual simplicity: the limiting distribution is obtained by performing the same operations [namely taking (left-)derivatives of the greatest convex minorant] on a limiting process corresponding to the finite-sample cumulative-sum diagram which are used to form the estimators.

The second approach, developed in Groeneboom (1985), Groeneboom (1988) and Kim and Pollard (1990), proceeds by "switching relations" which relate the estimators to the maximum of a certain process, and then appeal to an argmax-continuous mapping theorem. Systematic use of the switching relationships allowed Groeneboom to study the distribution theory of the processes $G_{a,b}$ and the slope process $g_{a,b}$ in great detail; see Groeneboom (1988). The resulting limiting distribution is, by virtue of a corresponding switching relationship for the limiting process, the same as that obtained by the continuous mapping approach.

In any case, these types of results have become standard: see Kim and Pollard (1990), Huang and Zhang (1994) and Huang and Wellner (1995), so we will not present detailed proofs here. For complete proofs of Theorems 2.4 and 2.7 by way of switching relations and the argmax continuous mapping theorem, see Banerjee (2000), Banerjee and Wellner (2000) and Banerjee and Wellner (2001).

Here we give a heuristic sketch of the proofs of Theorems 2.4 and 2.7. For $0 \le t < \infty$ set

$$V(t) = P\Delta 1\{T \le t\} = \int_0^t F\, dG, \qquad G(t) = P1\{T \le t\},$$

so that

$$\frac{dV}{dG}(t) = F(t), \text{ or } V'(t) = F(t)\, g(t)$$

if G has density g with respect to Lebesgue measure. Let $\mathbb{P}_n$ be the empirical measure of the pairs $(\Delta_1, T_1), \dots, (\Delta_n, T_n)$. The empirical counterparts of the functions V and G are defined, for $0 \le t < \infty$, by

$$\mathbb{V}_n(t) = \mathbb{P}_n \Delta 1\{T \le t\} = n^{-1} \sum_{i=1}^n \Delta_i 1\{T_i \le t\},$$

$$\mathbb{G}_n(t) = \mathbb{P}_n 1\{T \le t\} = n^{-1} \sum_{i=1}^n 1\{T_i \le t\}.$$

The estimators $\mathbb{F}_n$ and $\mathbb{F}_n^0$ are defined in terms of slopes of various greatest convex minorants of $\{(\mathbb{G}_n(t), \mathbb{V}_n(t)) : 0 \le t < \infty\}$, as explained in Section 2.2.

Now for fixed $t \in (0,\infty)$ and $0 < K < \infty$ we define localized versions $\{(\mathbb{G}_n^{loc}(t,h), \mathbb{V}_n^{loc}(t,h)) : \; h \in [-K, K]\}$ of the cumulative sum diagram at a fixed $t \in (0,\infty)$ as follows:

$$\begin{aligned}\mathbb{G}_n^{loc}(t,h) &\equiv n^{1/3}(\mathbb{G}_n(t+n^{-1/3}h) - \mathbb{G}_n(t)),\\ \mathbb{V}_n^{loc}(t,h) &\equiv n^{1/3}\big\{n^{1/3}\big(\mathbb{V}_n(t+n^{-1/3}h) - \mathbb{V}_n(t)\big)\\ &\qquad - n^{1/3}(\mathbb{G}_n(t+n^{-1/3}h) - \mathbb{G}_n(t))F(t)\big\}.\end{aligned}$$

Note that

$$E\mathbb{G}_n^{loc}(t,h) = n^{1/3}(G(t+n^{-1/3}h) - G(t)) \to hg(t)$$

if $g = G'$ exists at t, while $Var(\mathbb{G}_n^{loc}(t,h)) = O(n^{-2/3})$, so that $\mathbb{G}_n^{loc}(t,h) \xrightarrow{p} hg(t)$. Furthermore

$$\begin{aligned}\mathbb{V}_n^{loc}(t,h) &= n^{2/3}(\mathbb{P}_n - P)(\Delta - F(t))(1_{[0,\, t+n^{-1/3}h]}(T) - 1_{[0,\, t]}(T))\\ &\quad + \; n^{2/3}P(\Delta - F(t))(1_{[0,\, t+n^{-1/3}h]}(T) - 1_{[0,\, t]}(T))\\ &\Rightarrow aW(h) + bh^2 \equiv X_{a,\, b}(h)\end{aligned} \tag{6.1}$$

where $a = \sqrt{F(t)(1-F(t))g(t)}$, $b = f(t)g(t)/2$, W is a two-sided Brownian motion starting from zero, and the weak convergence is in $l^\infty([-K, K])$ for each $0 < K < \infty$; see, for example, Van der Vaart and Wellner (1996), page 299.

When $t \ne t_0$, part A of Theorem 2.4 follows (at least heuristically) by the "slope of Greatest Convex Minorant continuous mapping theorem" of Prakasa Rao (1969) and Huang and Zhang [(1994), Lemma 4, page 1265] upon noting that asymptotically the constraint at t_0 has no effect at $t \ne t_0$.

Similarly, when $t = t_0$, part B of Theorem 2.4 follows from (6.1) at $t = t_0$ (so a and b are as in (6.1) with $t = t_0$), and the "slope of Greatest Convex Minorant continuous mapping theorem" of Prakasa Rao (1969) and Huang and Zhang (1994). Note that t_0 has become 0 on the localized time scale, while the constraint on slopes ($= \theta_0$ at t_0) has become $= 0$ at $h = 0$. In this case the constraint at t_0 matters and the limiting process for the constrained estimator is as described in Theorem 2.3. The joint convergence in $\mathscr{L}^p[-K, K] \times \mathscr{L}^p[-K, K]$ follows immediately from the finite-dimensional convergence since the processes are monotone, as was noted by Huang and Zhang (1994), Corollary 2, page 1260.

To lay the groundwork for Theorem 2.7, we rewrite the localized process $\mathbb{V}_n^{loc}(t_0, h)$ as

$$\begin{aligned}\mathbb{V}_n^{loc}(t_0,h) &= n^{2/3}(\mathbb{P}_n - P_{F_n,\, G})(\Delta - F(t_0))\big(1_{[0,\, t_0+n^{-1/3}h]}(T) - 1_{[0,\, t_0]}(T)\big)\\ &\quad + n^{2/3}(P_{F_n,\, G} - P_{F,\, G})\big(\Delta - F(t_0)\big)\big(1_{[0,\, t_0+n^{-1/3}h]}(T) - 1_{[0,\, t_0]}(T)\big)\end{aligned} \tag{6.2}$$

$$+ n^{2/3} P_{F,G}(\Delta - F(t_0))\big(1_{[0,t_0+n^{-1/3}h]}(T) - 1_{[0,t_0]}(T)\big)$$
$$\Rightarrow aW(h) + \Psi(h) + bh^2 \equiv X_{a,b,\Psi}(h)$$

under $P_{F_n,G}$ by an adaption of the null hypothesis proof together with the fact that

$$\begin{aligned}
n^{2/3}&(P_{F_n,G} - P_{F,G})(\Delta - F(t_0))(1_{[0,t_0+n^{-1/3}h]}(T) - 1_{[0,t_0]}(T))\\
&= n^{2/3} \int_{t_0}^{t_0+n^{-1/3}h} \big\{(F_n(s) - F(t_0)) - (F(s) - F(t_0))\big\}\, dG(s)\\
&= \int_0^h n^{1/3}\big(F_n(t_0 + n^{-1/3}z) - F\big(t_0 + n^{-1/3}z\big)\big) g\big(t_0 + n^{-1/3}z\big)\, dz\\
&\to g(t_0) \int_0^h B(z)\, dz = \Psi(h)
\end{aligned}$$

uniformly on compact subsets by Assumptions B(1) and B(2). Then Theorem 2.7 follows by appeal to the "slope of greatest convex minorant" continuous mapping theorem. □

Another proof of (6.2) proceeds from joint convergence of $\mathbb{V}_n^{loc}$ and the local likelihood ratio $\log(L_n(F_n)/L_n(F))$ together with an application of (a general version of) Le Cam's third lemma; see Banerjee and Wellner (2001) for a proof organized this way.

6.2. *Proofs for Subsection* 2.4.

PROOF OF THEOREM 2.5. In what follows we denote the set on which $\mathbb{F}_n$ and $\mathbb{F}_n^0$ differ by D_n. We first note that:

$$\log L_n(\mathbb{F}_n) - \log L_n(\mathbb{F}_n^0) = n \int_{D_n} \big(K(\mathbb{F}_n(t), \theta_0) - K(\mathbb{F}_n^0(t), \theta_0)\big)\, d\mathbb{G}_n(t), \tag{6.3}$$

where

$$K(p, \theta_0) = p \log \frac{p}{\theta_0} + (1 - p) \log \frac{1 - p}{1 - \theta_0}.$$

We first sketch the proof of the identity (6.3). From the characterizations of $\mathbb{F}_n$ and $\mathbb{F}_n^0$ it follows that these are constant on blocks, and on each block $\mathbb{F}_n$ and $\mathbb{F}_n^0$ are equal to the average of the Δ_i's on that block, or, in the case of $\mathbb{F}_n^0$, constant and equal to θ_0 on the entire block. Using these facts together with elementary algebra yields (6.3). Thus the likelihood ratio statistic is

$$2 \log \lambda_n = 2n\mathbb{P}_n\big\{\big(K(\mathbb{F}_n(T), \theta_0) - K(\mathbb{F}_n^0(T), \theta_0)\big)1_{D_n}(T)\big\}.$$

Now set $\Gamma(a, x) = a \log(x) + (1 - a) \log(1 - x)$, and note that

$$K(\mathbb{F}_n(T), \theta_0) = \Gamma(\mathbb{F}_n(T), \mathbb{F}_n(T)) - \Gamma(\mathbb{F}_n(T), \theta_0).$$

Expanding $\Gamma(\mathbb{F}_n(T), \theta_0)$ around $\mathbb{F}_n(T)$ gives

$$\begin{aligned} K(\mathbb{F}_n(T), \theta_0) = &-\Gamma^{'}(\mathbb{F}_n(T), \mathbb{F}_n(T))(\theta_0 - \mathbb{F}_n(T)) \\ &- \tfrac{1}{2}\Gamma^{''}(\mathbb{F}_n(T), \mathbb{F}_n(T))(\theta_0 - \mathbb{F}_n(T))^2 \\ &- \tfrac{1}{6}\Gamma^{'''}(\mathbb{F}_n(T), \mathbb{F}_n^{\star}(T))(\theta_0 - \mathbb{F}_n(T))^3, \end{aligned}$$

where $\mathbb{F}_n^{\star}(T)$ is an intermediate point between $\mathbb{F}_n(T)$ and θ_0 and

$$\Gamma^{'}(\mathbb{F}_n(T), \mathbb{F}_n(T)) = \frac{\mathbb{F}_n(T)}{\mathbb{F}_n(T)} - \frac{1-\mathbb{F}_n(T)}{1-\mathbb{F}_n(T)} = 0,$$

$$\Gamma^{''}(\mathbb{F}_n(T), \mathbb{F}_n(T)) = -\frac{\mathbb{F}_n(T)}{\mathbb{F}_n(T)^2} - \frac{1-\mathbb{F}_n(T)}{(1-\mathbb{F}_n(T))^2} = -\frac{1}{\mathbb{F}_n(T)\,(1-\mathbb{F}_n(T))}$$

and

$$\Gamma^{'''}(\mathbb{F}_n(T), \mathbb{F}_n^{\star}(T)) = 2\left(\frac{\mathbb{F}_n(T)}{(\mathbb{F}_n^{\star}(T))^3} - \frac{1-\mathbb{F}_n(T)}{(1-\mathbb{F}_n^{\star}(T))^3}\right).$$

Thus,

$$\begin{aligned} K(\mathbb{F}_n(T), \theta_0) = &\frac{1}{2}\,\frac{1}{\mathbb{F}_n(T)\,(1-\mathbb{F}_n(T))}\,(\mathbb{F}_n(T) - \theta_0)^2 \\ &+ \frac{1}{6}\Gamma^{'''}(\mathbb{F}_n(T), \mathbb{F}_n^{\star}(T))(\mathbb{F}_n(T) - \theta_0)^3. \end{aligned}$$

Similarly,

$$\begin{aligned} K(\mathbb{F}_n^0(T), \theta_0) = &\frac{1}{2}\frac{1}{\mathbb{F}_n^0(T)\,(1-\mathbb{F}_n^0(T))}(\mathbb{F}_n^0(T) - \theta_0)^2 \\ &+ \frac{1}{6}\Gamma^{'''}(\mathbb{F}_n^0(T), \mathbb{F}_n^{\star\star}(T))(\mathbb{F}_n(T) - \theta_0)^3 \end{aligned}$$

where $\mathbb{F}_n^{\star\star}(T)$ is an intermediate point between $\mathbb{F}_n^0(T)$ and θ_0. Thus,

$$\begin{aligned} 2\log\,\lambda_n &= 2\,n\,\mathbb{P}_n\big\{\big(K(\mathbb{F}_n(T), \theta_0) - K(\mathbb{F}_n^0(T), \theta_0)\big)\,1_{D_n}(T)\big\} \\ &= n\,\mathbb{P}_n\bigg(\frac{1}{\mathbb{F}_n(T)\,(1-\mathbb{F}_n(T))}(\mathbb{F}_n(T) - \theta_0)^2 \\ &\qquad - \frac{1}{\mathbb{F}_n^0(T)\,(1-\mathbb{F}_n^0(T))}(\mathbb{F}_n^0(T) - \theta_0)^2\bigg)1_{D_n}(T) \\ &\quad + \frac{n}{6}\,\mathbb{P}_n\Big(\Gamma^{'''}(\mathbb{F}_n(T), \mathbb{F}_n^{\star}(T))(\mathbb{F}_n(T) - \theta_0)^3 \\ &\qquad - \;\Gamma^{'''}(\mathbb{F}_n^0(T), \mathbb{F}_n^{\star\star}(T))(\mathbb{F}_n(T) - \theta_0)^3\Big)1_{D_n}(T) \\ &= S_n + R_n. \end{aligned} \tag{6.4}$$

We now introduce the local variable h as before through the relation $h = n^{1/3}\,(T - t_0)$, and denote the transformed difference set in terms of the local variable by $\tilde{D}_n \equiv n^{1/3}(D_n - t_0)$. The processes X_n and Y_n are as before. Now it is easily shown that:

(a) For every $\varepsilon > 0$, there exists a $K_\varepsilon > 0$ such that

$$\liminf_n P(\tilde{D}_n \subset [-K_\varepsilon, K_\varepsilon]) > 1 - \varepsilon.$$

(b) For every $\varepsilon > 0$ and $M > 0$, there exists a $B > 0$ such that

$$\limsup_n P\Big(\sup_{z \in [-M, M]} |X_n(z)| > B\Big) \le \varepsilon$$

and

$$\limsup_n P\Big(\sup_{z \in [-M, M]} |Y_n(z)| > B\Big) \le \varepsilon.$$

The above results along with the facts that $\mathbb{F}_n$ and $\mathbb{F}_n^0$ converge almost surely to F uniformly on some interval around t_0, that F is continuous, that $\mathbb{F}_n^\star(T)$ is intermediate between $\mathbb{F}_n(T)$ and θ_0 and $\mathbb{F}_n^{\star\star}(T)$ is intermediate between $\mathbb{F}_n^0(T)$ and θ_0 and that $\mathbb{F}_n(T)$ and $\mathbb{F}_n^0(T)$ are eventually bounded away from 0 and 1 with arbitraily high probability, entails that we can write,

$$\begin{aligned} R_n = {} & \frac{n}{6}\mathbb{P}_n\Big(\Gamma'''(\mathbb{F}_n(T), \theta_0)\big(\mathbb{F}_n(T) - \theta_0\big)^3 - \Gamma'''\big(\mathbb{F}_n^0(T), \theta_0\big)\big(\mathbb{F}_n(T) - \theta_0\big)^3\Big)1_{D_n}(T) \\ & + o_p(1). \end{aligned}$$

The first term on the right side of the above display can be shown to be $O_p(n^{-1/3})$ and hence is certainly $o_p(1)$ showing that $R_n = o_p(1)$. Then from (6.4) it follows that we only need to find the asymptotic distribution of

$$\begin{aligned} S_n = n\mathbb{P}_n\bigg(& \frac{1}{\mathbb{F}_n(T)(1 - \mathbb{F}_n(T))}(\mathbb{F}_n(T) - \theta_0)^2 - \frac{1}{\mathbb{F}_n^0(T)(1 - \mathbb{F}_n^0(T))}(\mathbb{F}_n^0(T) - \theta_0)^2\bigg) \\ & \times 1_{D_n}(T). \end{aligned}$$

But

$$\begin{aligned} S_n &= n\,\mathbb{P}_n\bigg(\frac{1}{\theta_0(1 - \theta_0)}\big\{(\mathbb{F}_n(T) - \theta_0)^2 - (\mathbb{F}_n^0(T) - \theta_0)^2\big\}\bigg) + o_p(1) \\ &= \tilde{S}_n + o_p(1). \end{aligned} \tag{6.5}$$

This follows from the facts that

$$a_n \equiv n\,\mathbb{P}_n\bigg(\bigg(\frac{1}{\mathbb{F}_n(T)(1 - \mathbb{F}_n(T))} - \frac{1}{\theta_0(1 - \theta_0)}\bigg)(\mathbb{F}_n(T) - \theta_0)^2\bigg)1_{D_n}(T)$$

and

$$b_n \equiv n\,\mathbb{P}_n\bigg(\bigg(\frac{1}{\mathbb{F}_n^0(T)(1 - \mathbb{F}_n^0(T))} - \frac{1}{\theta_0(1 - \theta_0)}\bigg)(\mathbb{F}_n^0(T) - \theta_0)^2\bigg)1_{D_n}(T)$$

are both $o_p(1)$. In the case of a_n this can be seen as follows. Write

$$a_n = n\mathbb{P}_n\,\tilde{f}_n = n(\mathbb{P}_n - P)\,\tilde{f}_n + nP\tilde{f}_n.$$

Now

$$n(\mathbb{P}_n - P)\tilde{f}_n = n^{1/3}(\mathbb{P}_n - P)\bigg(\bigg(\frac{1}{\mathbb{F}_n(T)(1-\mathbb{F}_n(T))} - \frac{1}{\theta_0(1-\theta_0)}\bigg) \times \big(n^{1/3}(\mathbb{F}_n(T) - \theta_0)\big)^2\bigg)$$
$$= n^{1/3}(\mathbb{P}_n - P)\tilde{g}_n.$$

Now, $\tilde{g}_n$ eventually belongs to a uniformly bounded Donsker class of functions with arbitrarily high probability, whence it follows that $n\,(\mathbb{P}_n - P)\,\tilde{f}_n = o_p(1)$. Also,

$$nP\tilde{f}_n = \int_{\tilde{D}_n}\bigg(\frac{1}{\mathbb{F}_n(t_n(z))(1-\mathbb{F}_n(t_n(z)))} - \frac{1}{\theta_0(1-\theta_0)}\bigg)X_n^2(z)g(t_n(z))\,dz,$$

where $t_n(z) = t_0 + n^{-1/3}\,z$. The boundedness in probability of X_n on $\tilde{D}_n$ and the uniform convergence of $\mathbb{F}_n(t_n(z))$ to θ_0 on $\tilde{D}_n$ and the fact that $\tilde{D}_n$ is eventually in a compact set, then entail that the expression in the above display is $o_p(1)$. Thus a_n is $o_p(1)$.

Now write $\tilde{S}_n$ (refer to 6.5) as

$$\tilde{S}_n = n\mathbb{P}_n\tilde{u}_n = n(\mathbb{P}_n - P)\tilde{u}_n + nP\tilde{u}_n = o_p(1) + nP\tilde{u}_n.$$

That $n\,(\mathbb{P}_n - P)\,\tilde{u}_n$ is $o_p(1)$ can be established as before by arguing that $n^{2/3}\,\tilde{u}_n$ is (eventually) in a Donsker class of functions with arbitrarily high probability. It remains to tackle $n\,P\,\tilde{u}_n$, which can be written as

$$\begin{aligned}
nP\tilde{u}_n &= n\frac{1}{\theta_0(1-\theta_0)}P\Big\{(\mathbb{F}_n(T) - \theta_0)^2 - (\mathbb{F}_n^0(T) - \theta_0)^2\Big\}1_{D_n}(T)\\
&= \frac{1}{\theta_0(1-\theta_0)}\int_{\tilde{D}_n}\big(X_n^2(z) - Y_n^2(z)\big)g(t_n(z))\,dz\\
&= \frac{g(t_0)}{\theta_0\,(1-\theta_0)}\int_{\tilde{D}_n}\big(X_n^2(z) - Y_n^2(z)\big)\,dz + o_p(1)\\
&= L_n + o_p(1).
\end{aligned}$$

We will show that

$$L_n \xrightarrow{d} L_{a,b} \equiv \frac{1}{g(t_0)\theta_0\,(1-\theta_0)}\int_{D_{a,b}}\big(g_{a,b}^2(z) - (g_{a,b}^0(z))^2\big)\,dz\,,$$

where $D_{a,b}$ is the set on which $g_{a,b}$ and $g^0_{a,b}$ differ. To this end it clearly suffices to show that

$$\int_{\tilde{D}_n} \big(X_n^2(z) - Y_n^2(z)\big)\,dz \overset{d}{\to} \int_{D_{a,b}} \frac{(g_{a,b}^2(z) - (g^0_{a,b}(z))^2)}{g^2(t_0)}\,dz. \tag{6.6}$$

To this end, we invoke the following lemma from Prakasa Rao (1969).

LEMMA 6.1. *Suppose that* $\{X_{n\varepsilon}\}$, $\{Y_n\}$ *and* $\{W_\varepsilon\}$ *are three sets of random variables such that*:

(i) $\lim_{\varepsilon\to 0} \limsup_{n\to\infty} P(X_{n\varepsilon} \neq Y_n) = 0$;
(ii) $\lim_{\varepsilon\to 0} P(W_\varepsilon \neq Y) = 0$;
(iii) *For every* $\varepsilon > 0$, $X_{n\varepsilon} \overset{d}{\to} W_\varepsilon$ *as* $n \to \infty$.

Then $Y_n \overset{d}{\to} Y$ *as* $n \to \infty$.

Using result (a) together with the above lemma and Theorem 2.4, by choosing $Y_n \equiv \int_{\tilde{D}_n} (X_n^2(z) - Y_n^2(z))\,dz$,

$$X_{n\varepsilon} \equiv \int_{[-K_\varepsilon, K_\varepsilon]} \big(X_n^2(z) - Y_n^2(z)\big)\,dz,$$

$$W_\varepsilon \equiv \int_{[-K_\varepsilon, K_\varepsilon]} \frac{\big(g_{a,b}^2(z) - (g^0_{a,b}(z))^2\big)}{g^2(t_0)}\,dz$$

and

$$Y \equiv \int_{D_{a,b}} \frac{\big(g_{a,b}^2(z) - (g^0_{a,b}(z))^2\big)}{g^2(t_0)}\,dz,$$

the convergence in distribution in (6.6) follows in a straightforward manner.

It remains to prove that

$$L_{a,b} \overset{d}{\equiv} \int_D \big(\mathbb{S}^2(y) - \mathbb{S}_0^2(y)\big)\,dy \equiv \mathbb{D};$$

this gives the key universality of the limiting distribution promised in the introduction. This proceeds by Brownian scaling. The first step is to note that

$$X_{a,b}(t) \overset{d}{=} a(a/b)^{1/3} X_{1,1}((b/a)^{2/3} t) \equiv a(a/b)^{1/3} X((b/a)^{2/3} t) \tag{6.7}$$

as a process indexed by $t \in \mathbb{R}$. This implies that

$$(G_{a,b}(t), G^0_{a,b}(t)) \overset{d}{=} a(a/b)^{1/3}\big(G_{1,1}((b/a)^{2/3} t), G^0_{1,1}((b/a)^{2/3} t)\big), \tag{6.8}$$

as processes, which in turn yields

$$\begin{aligned}\big(g_{a,b}(t), g^0_{a,b}(t), D_{a,b}\big) &\overset{d}{=} a(b/a)^{1/3}\big(g_{1,1}((b/a)^{2/3} t), g^0_{1,1}((b/a)^{2/3} t), \\ &\qquad (a/b)^{2/3} D_{1,1}\big) \\ &\equiv a(b/a)^{1/3}\big(\mathbb{S}((b/a)^{2/3} t), \mathbb{S}^0\big((b/a)^{2/3} t\big), (a/b)^{2/3} D\big)\end{aligned} \tag{6.9}$$

as processes indexed by $t \in \mathbb{R}$. Thus by straightforward calculation it follows that

$$\begin{aligned} L_{a,b} &= \frac{1}{g(t_0)F(t_0)(1-F(t_0))} \int_{D_{a,b}} \left(g_{a,b}^2(z) - (g_{a,b}^0(z))^2\right) dz \\ &\stackrel{d}{=} \frac{1}{a^2} \int_{(a/b)^{2/3} D_{1,1}} a^2 (b/a)^{2/3} \left(g_{1,1}^2((b/a)^{2/3} z) - (g_{1,1}^0((b/a)^{2/3} z))^2\right) dz \\ &= \int_D \left(\mathbb{S}^2(y) - \mathbb{S}_0^2(y)\right) dy \equiv \mathbb{D}, \end{aligned}$$

completing the proof. □

6.3. *Proofs for Subsection* 2.5.

PROOF OF THEOREM 2.6. The local log-likelihood ratio for the interval censoring problem, is by expanding around F,

$$\begin{aligned} \log L_n(F_n) - \log L_n(F) &= n\mathbb{P}_n\left\{\Delta \log \frac{F_n}{F}(T) + (1-\Delta)\log\frac{1-F_n}{1-F}(T)\right\} \\ &= n\mathbb{P}_n\left\{\psi(\Delta, T; F_n) - \psi(\Delta, T; F)\right\} \\ &= n\mathbb{P}_n\left\{\psi'(\Delta, T; F)(F_n - F)(T)\right\} \\ &\quad + \frac{1}{2} n\mathbb{P}_n\{\psi''(\Delta, T; F)(F_n - F)^2(T)\} \\ &\quad + \frac{1}{6} n\mathbb{P}_n\{\psi'''(\Delta, T; F_n^\star)(F_n - F)^3(T)\} \\ &\equiv I_n + II_n + III_n \end{aligned}$$

where

$$\begin{aligned} \psi(\Delta, T; F) &\equiv \Delta \log F(T) + (1-\Delta)\log(1 - F(T)), \\ \psi'(\Delta, T; F) &= \frac{\Delta}{F(T)} - \frac{1-\Delta}{1-F(T)}, \\ \psi''(\Delta, T; F) &= -\frac{\Delta}{F^2(T)} - \frac{1-\Delta}{(1-F(T))^2}, \\ \psi'''(\Delta, T; F) &= 2\left(\frac{\Delta}{F^3(T)} - \frac{1-\Delta}{(1-F(T))^3}\right), \end{aligned}$$

and $F_n^*(T)$ is an intermediate point between $F_n(T)$ and $F(T)$. Note that

$$E\{\psi'(\Delta, T; F) | T\} = 0 \tag{6.10}$$

while

$$-E\{\psi''(\Delta, T; F)\,|\,T\} = \frac{1}{F(T)} + \frac{1}{1-F(T)} = \frac{1}{F(T)(1-F(T))}.$$

Consider now, the term I_n. We have

$$\begin{aligned} I_n &= n\,\mathbb{P}_n\biggl(\biggl(\frac{\Delta}{F(T)} - \frac{1-\Delta}{1-F(T)}\biggr)(F_n - F)(T)\biggr) \\ &= n^{2/3}(\mathbb{P}_n - P)\biggl(\biggl(\frac{\Delta - F(T)}{F(T)(1-F(T))}\biggr)A_n(T)\biggr) \\ &= \sqrt{n}(\mathbb{P}_n - P)(s_n) \end{aligned}$$

where

$$s_n(\Delta, T) = n^{1/6}\frac{\Delta - F(T)}{F(T)(1-F(T))}A_n(T).$$

Now $E(s_n(\Delta, T)) = 0$,

$$\begin{aligned} \mathrm{Var}(s_n(\Delta, T)) &= P(s_n^2(\Delta, T)) \\ &= n^{1/3}\int \frac{A_n^2(t)}{F(t)(1-F(t))}\,dG(t) \\ &= \int \frac{B_n^2(z)}{F(t_0 + n^{-1/3}z)(1 - F(t_0 + n^{-1/3}z))}\,g(t_0 + n^{-1/3}z)\,dz \\ &\to \sigma^2 \end{aligned} \tag{6.11}$$

by Assumptions A, B(1) and B(2) where σ^2 is as defined in (2.17). Moreover, with

$$M_n \equiv \sup_{t:\ |t-t_0|\le cn^{-1/3}}\biggl\{\frac{1}{F(t)(1-F(t))}\biggr\} \to \frac{1}{F(t_0)(1-F(t_0))} < \infty,$$

we find that, for each $\varepsilon > 0$ we have

$$\begin{aligned} E\{s_n^2 1_{[|s_n|\ge\sqrt{n}\varepsilon]}\} &\le M_n n^{1/3}\int_{[t:\ |A_n(t)|\ge\varepsilon n^{1/3}/M_n]} A_n^2(t)\,dG(t) \\ &\le M_n \int_{[z:\ |B_n(z)|\ge\varepsilon n^{1/3}/M_n]} B_n^2(z)g(t_0 + n^{-1/3}z)\,dz \\ &\to 0 \end{aligned}$$

by Assumptions A, B(1), and B(2) again. Thus the Lindeberg condition holds, and it follows from the Lindeberg–Feller central limit theorem that

$$I_n \xrightarrow{d} N(0, \sigma^2).$$

We now treat the term II_n. Note that

$$II_n = -\frac{1}{2} n^{1/3} \mathbb{P}_n\left(\left[\frac{\Delta}{F^2(T)} + \frac{1-\Delta}{(1-F(T))^2}\right] A_n^2(T)\right).$$

Thus we have

$$\begin{aligned} E(II_n) &= -\frac{1}{2} n^{1/3} P\left(\left[\frac{\Delta}{F^2(T)} + \frac{1-\Delta}{(1-F(T))^2}\right] A_n^2(T)\right) \\ &= -\frac{1}{2} n^{1/3} \int \frac{A_n^2(t)}{F(t)(1-F(t))}\, dG(t) \\ &= -\frac{1}{2} \int \frac{B_n^2(z)}{F(t_0+n^{-1/3}z)(1-F(t_0+n^{-1/3}z))} g(t_0+n^{-1/3}z)\, dz \\ &\to -\frac{\sigma^2}{2} \end{aligned}$$

as in (6.11), and, moreover,

$$\begin{aligned} \mathrm{Var}(II_n) &\le n^{-1/3} P\left\{\left(\frac{\Delta}{F^2(T)} + \frac{1-\Delta}{(1-F)^2(T)}\right)^2 A_n^4(T)\right\} \\ &\le n^{-1/3} \int \left(\frac{1}{F^3(t)} + \frac{1}{(1-F(t))^3}\right) A_n^4(t)\, dG(t) \\ &\le (2M_n)^2 n^{-2/3} \int \frac{B_n^4(z)}{F(t_0+n^{-1/3}z)(1-F(t_0+n^{-1/3}z))} g(t_0+n^{-1/3}z)\, dz \\ &= O(n^{-2/3}). \end{aligned}$$

Hence it follows that $II_n \xrightarrow{p} -(1/2)\sigma^2$.

It remains to deal with III_n. Note that for each n, if $F_n^\star(T)$ lies between $F(T)$ and $F_n(T)$, then

$$\left|F_n^\star(T) - F(T)\right| \le \left|F_n(T) - F(T)\right|.$$

Denote the set $[t_0 - cn^{-1/3}, t_0 + cn^{1/3}]$ by D_n. Then

$$\sup_{D_n} \left|F_n^\star(T) - F(T)\right| \le \sup_{D_n} \left|F_n(T) - F(T)\right| = n^{-1/3} \sup_{[-c,c]} B_n(z) \to 0,$$

so that $F_n^\star$ converges uniformly to F on the line (recall that $F_n^\star$ and F coincide outside D_n).

Using this it follows easily that $E(III_n) = O(n^{-1/3})$ and $\mathrm{Var}(III_n) = O(n^{-4/3})$, and consequently $III_n \xrightarrow{p} 0$.

It follows that

$$\log L_n(F_n) - \log L_n(F) \xrightarrow{d} N(-\sigma^2/2, \sigma^2);$$

hence the sequence of alternatives $\{P^n_{F_n,G}\}$ and $\{P^n_{F,G}\}$ are mutually contiguous, by a direct application of Le Cam's first lemma [see, e.g., Van der Vaart and Wellner (1996), page 404]. □

Proof of Theorem 2.8. From the proof of Theorem 2.5 we have the following representation of the likelihood ratio statistic under the null hypothesis:

$$2\log\lambda_n = \frac{g(t_0)}{F(t_0)(1-F(t_0))}\int_{\tilde{D}_n}\big(X_n^2(z) - Y_n^2(z)\big)\,dz + o_p(1) \equiv L_n + o_p(1).$$

Since terms that are $o_p(1)$ under $P^n_{F,G}$ continue be $o_p(1)$ under $P^n_{F_n,G}$ by contiguity (which follows from Theorem 2.6), it follows that the same representation holds under $\{P^n_{F_n,G}\}$, and it suffices to find the asymptotic distribution of L_n under $\{P^n_{F_n,G}\}$. That L_n converges in distribution under $\{P^n_{F_n,G}\}$ to the right side of (2.18) follows from Theorem 2.7 together with Lemma 6.1 by steps similar to the proof of Theorem 2.5.

The equality in distribution given by (2.19) follows from scaling arguments similar to those used in Theorem 2.5. □

6.4. *One Proof for Subsection* 2.6.

Proof of Part of Theorem 2.9. The convergence in probability in (2.20) is proved using consistency results of Schick and Yu (1999) for the unconstrained estimator $\mathbb{F}_n$ together with corresponding results for the constrained estimator $\mathbb{F}_n^0$ and Glivenko-Cantelli class arguments; see Banerjee (2000) for the details. Here we will just prove the equality in (2.21).

By straightforward calculation, the limit Kullback–Leibler discrepancy $K(P_{F,G}, P_{H,G})$ in (2.20) is given by

$$K(P_{F,G}, P_{H,G}) = P_{F,G}\left[\Delta\,\log\frac{F}{H}(T) + (1-\Delta)\log\frac{1-F}{1-H}(T)\right].$$

To show that (2.21) holds, we only need to show that for any distribution function U satisfying $U(t_0) = \theta_0$,

$$\mathrm{Diff}(U,H) \equiv K(P_{F,G}, P_{U,G}) - K(P_{F,G}, P_{H,G}) \ge 0.$$

But we can write

$$\begin{aligned}
\mathrm{Diff}(U,H) &= P_{F,G}\left[\Delta\log\frac{F}{U}(T) + (1-\Delta)\log\frac{1-F}{1-U}(T)\right] \\
&\qquad - P_{F,G}\left[\Delta\log\frac{F}{H}(T) + (1-\Delta)\log\frac{1-F}{1-H}(T)\right] \\
&= P_{F,G}\left[\Delta\log\frac{H}{U}(T) + (1-\Delta)\log\frac{1-H}{1-U}(T)\right] \\
&= \int\left(F(t)\,\log\frac{H}{U}(t) + (1-F(t))\log\frac{1-H}{1-U}(t)\right)dG(t)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{[t_0, t_1]^c} \left(H(t) \log \frac{H}{U}(t) + (1 - H(t)) \log \frac{1-H}{1-U}(t) \right) dG(t) \\
&\quad + \int_{[t_0, t_1]} \left(F(t) \log \frac{H}{U}(t) + (1 - F(t)) \log \frac{1-H}{1-U}(t) \right) dG(t) \\
&= \int \left(H(t) \log \frac{H}{U}(t) + (1 - H(t)) \log \frac{1-H}{1-U}(t) \right) dG(t) \\
&\quad + \int_{[t_0, t_1]} \left\{ (F(t) - H(t)) \log \frac{H}{U}(t) \right. \\
&\qquad\qquad \left. + (H(t) - F(t)) \log \frac{1-H}{1-U}(t) \right\} dG(t) \\
&\equiv K(P_{H,G}, P_{U,G}) + S
\end{aligned}
$$

where we have used the fact that H and F coincide outside the interval $[t_0, t_1]$ (or $[t_1, t_0]$ if $t_1 < t_0$]). Regarding the previous display, note that $K(P_{H,G}, P_{U,G})$ is always nonnegative (by Jensen's inequality). To show that the second term S is nonnegative, we show that the integrand is nonnegative. This follows easily because on $[t_0, t_1]$, $H(t) = \theta_0$ identically whereas $F(t) \leq \theta_0$ so that $F(t) - H(t) \leq 0$. Since $U(t_0) = \theta_0$, $U(t) \geq \theta_0$ on $[t_0, t_1]$, showing that $\log(H(t)/U(t)) \leq 0$. But then

$$(F(t) - H(t)) \log \frac{H}{U}(t) \geq 0.$$

Similarly, on $[t_0, t_1]$

$$(H(t) - F(t)) \log \frac{1-H}{1-U}(t) \geq 0.$$

This shows that $K(P_{F,G}, P_{U,G}) - K(P_{F,G}, P_{H,G}) \geq 0$, and hence (2.21) holds. □

Acknowledgments. We owe thanks to Piet Groeneboom for many conversations about estimation of monotone functions, and to Michael Perlman for the suggestion to consider testing equality at $k > 1$ points as discussed briefly in Section 5.

Thanks also go to the referees and an Associate Editor for several suggestions which improved the presentation and exposition. In particular the identity (6.3) was suggested by one of the referees.

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.

BANERJEE, M. (2000). Likelihood Ratio Inference in Regular and Nonregular Problems. Ph.D. dissertation, Univ. Washington.

BANERJEE, M. and WELLNER, J. A. (2000). Likelihood ratio tests for monotone functions. Technical Report 377, Dept. Statistics, Univ. Washington.

BANERJEE, M. and WELLNER, J. A. (2001). Tests and confidence intervals for monotone functions: further developements. Technical report, Dept. Statistics, Univ. Washington. In preparation.

BARLOW, R. E., BARTHOLOMEW, BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions* Wiley, New York.

BERK, R. H. and JONES, D. H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59.

BRUNK, H. D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 177–195. Cambridge Univ. Press.

GRENANDER, U. (1956). On the theory of mortality measurement, Part II. *Skand. Actuar.* **39** 125–153.

GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. LeCam and R. A. Olshen, eds.) **2** 529–555. Wadsworth, Belmont, CA.

GROENEBOOM, P. (1988). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109.

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2000a). A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. Technical Report 369, Dept. Statistics, Univ. Washington.

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2000b). Estimation of convex functions: characterizations and asymptotic theory. Technical Report 372, Dept. Statistics, Univ. Washington.

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser, Boston.

GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400.

HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22** 3–33.

HUANG, Y. and ZHANG, C. H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

LEURGANS, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Ann. Statist.* **10** 287–296.

MURPHY, S. and VAN DER VAART, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25** 1471–1509.

OWEN, A. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* **90** 516–521.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhȳa. Ser. A* **31** 23–36.

SCHICK, A. and YU, Q. (1999). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.* **27** 45–55.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

VAN EEDEN, C. (1957a). Maximum likelihood estimation of partially or completely ordered parameters, I. *Proc. K. Ned. Akad. Wet.* **60**; *Indag. Math.* **19** 128–136.

VAN EEDEN, C. (1957b). Maximum likelihood estimation of partially or completely ordered parameters, II. *Proc. K. Ned. Akad. Wet.* **60**; *Indag. Math.* **19** 201–211.

WELLNER, J. A. (2001). Gaussian white noise models: a partial review and results for monotone functions. In *IMS Lecture Notes and Monograph Series Volume in Honor of W. J. Hall*. To appear.

WELLNER, J. A. and ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814.

WU, W. B., WOODROOFE, M. and MENTZ, G. (2001). Isotonic regression: another look at the change point problem. Technical report, Univ. Michigan. Available at www.stat.lsa.umich.edu/~michaelw/.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
4062, FRIEZE BUILDING
105 S. STATE STREET
ANN ARBOR, MICHIGAN 48109
E-MAIL: moulib@umich.edu

UNIVERSITY OF WASHINGTON
DEPARTMENT OF STATISTICS
BOX 354322
SEATTLE, WASHINGTON 98195-4322
E-MAIL: jaw@stat.washington.edu