The International Journal of Biostatistics

Volume 1, *Issue* 1 2005 *Article* 3

Score Statistics for Current Status Data: Comparisons with Likelihood Ratio and Wald Statistics

Moulinath Banerjee* Jon A. Wellner†

*University of Michigan, moulib@umich.edu

†University of Washington, jaw@stat.washington.edu

Score Statistics for Current Status Data: Comparisons with Likelihood Ratio and Wald Statistics*

Moulinath Banerjee and Jon A. Wellner

Abstract

In this paper we introduce three natural "score statistics" for testing the hypothesis that F(t_0)takes on a fixed value in the context of nonparametric inference with current status data. These three new test statistics have natural interpretations in terms of certain (weighted) L_2 distances, and are also connected to natural "one-sided" scores. We compare these new test statistics with the analogue of the classical Wald statistic and the likelihood ratio statistic introduced in Banerjee and Wellner (2001) for the same testing problem. Under classical "regular" statistical problems the likelihood ratio, score, and Wald statistics all have the same chi-squared limiting distribution under the null hypothesis. In sharp contrast, in this non-regular problem all three statistics have different limiting distributions under the null hypothesis. Thus we begin by establishing the limit distribution theory of the statistics under the null hypothesis, and discuss calculation of the relevant critical points for the test statistics. Once the null distribution theory is known, the immediate question becomes that of power. We establish the limiting behavior of the three types of statisticsb under local alternatives. We have also compared the power of these five different statistics via a limited Monte-Carlo study. Our conclusions are: (a) the Wald statistic is less powerful than the likelihood ratio and score statistics; and (b) one of the score statistics may have more power than the likelihood ratio statistic for some alternatives.

KEYWORDS: asymptotic distribution, Brownian motion, contiguous alternatives, greatest convex minorant, likelihood ratio statistic, log-likelihood ratio, score statistic, Wald statistic

*We owe thanks to Piet Groeneboom for proposing the L_2 statistic T_{n,1} and for many helpful discussions. This research was supported by NSF grant DMS-0306235 (Banerjee) and by NSF grant DMS-0203320 and NIAID grant 2R01 AI292968-04 (Wellner).

1 Introduction

The problem of estimating a monotone function arises frequently in statistics. In what follows we study a very particular example of this class of problems, the case of "current status data".

Current Status Data: Let $(X_1, T_1), (X_2, T_2), \ldots, (X_n, T_n)$ be n i.i.d. pairs of random variables. For each i, $X_i \sim F$, $T_i \sim G$ and X_i is independent of T_i. Here F and G are continuous distributions concentrated on the positive half-line. In the interval censoring model, we do not observe the actual failure times, the X_i's. All that we observe for the i'th individual is the vector (Δ_i, T_i) where $\Delta_i = 1\{X_i \le T_i\}$ is the indicator of a failure. We are interested in inference about F, based on the "current status" data $\{(\Delta_i, T_i),\ i = 1, \ldots, n\}$. More specifically, here we focus on inference for $F(t_0)$, the value of F at a fixed point t_0.

For applications involving current status data see Keiding et al. (1996).

In this paper we introduce a natural "score (or Rao) statistic" for testing $F(t_0) = \theta_0$ and study several variants of this statistic. We derive the asymptotic distribution of the new statistics under the null hypothesis and local alternatives, and give some results concerning the power of these statistics relative to the likelihood ratio statistic studied in Banerjee and Wellner (2001) and a natural "Wald" type statistic. The three types of statistics are introduced in section 2. Behavior of the statistics under the null hypothesis and local alternatives is discussed in sections 3 and 4; first order behavior of the statistics under fixed alternatives is given in section 5. Section 6 summarizes the results of the theory presented together with preliminary Monte-Carlo studies. One definite conclusion is that the Wald test is dominated by the likelihood ratio and score tests. Proofs of the results in sections 2-5 are sketched in section 7. More detailed proofs are given in the companion technical report, Banerjee and Wellner (2003).

2 The Likelihood Ratio, Score and Wald Statistics

In this section we introduce the new "score statistics" for testing

$$H_0 : F(t_0) = \theta_0 \qquad \text{versus} \qquad H_1 : F(t_0) \neq \theta_0\,;$$

here $0 < t_0 < \infty$ and $0 < \theta_0 < 1$ are both fixed. We denote the distribution of (Δ, T) under (F, G) by $P_{F,G}$. We assume throughout that the following assumption holds:

Assumption A: Both F and G are continuously differentiable in a neighborhood of t_0 with Lebesgue densities f and g respectively; also $0 < f(t_0), g(t_0)$ and $0 < F(t_0) < 1$.

The log-likelihood $\log L_n(F)$ based on n i.i.d. observations $(\Delta_1, T_1), (\Delta_2, T_2), \ldots, (\Delta_n, T_n)$ is then given by

$$\log L_n(F) = n\mathbb{P}_n \left(\Delta \log F(T) + (1-\Delta)\log(1-F(T))\right) , \tag{2.1}$$

where $\mathbb{P}_n$ is the empirical measure of the observations $\{(\Delta_i, T_i)_{i=1}^n\}$. The likelihood ratio statistic for testing $F(t_0) = \theta_0$ is given by

$$\begin{aligned} 2\log(\lambda_n) &= 2\left(\log L_n(\mathbb{F}_n) - \log L_n(\mathbb{F}_n^0)\right) \\ &= 2n\mathbb{P}_n \left\{\Delta \log \frac{\mathbb{F}_n(T)}{\mathbb{F}_n^0(T)} + (1-\Delta)\log\frac{1-\mathbb{F}_n(T)}{1-\mathbb{F}_n^0(T)}\right\} \end{aligned}$$

where $\mathbb{F}_n$ is the unconstrained MLE and $\mathbb{F}_n^0$ is the constrained MLE under the null hypothesis (based on the interval censored data). The behavior of the likelihood ratio statistic under the null hypothesis and local alternatives was derived in Banerjee and Wellner (2001) and will be compared here to the new score statistics.

A natural and very intuitive way of measuring the distance between $\mathbb{F}_n$, the unconstrained MLE of F and $\mathbb{F}_n^0$, the constrained MLE of F is to compute $\text{Diff}_n = \sum_{i=1}^n (\mathbb{F}_n(T_i) - \mathbb{F}_n^0(T_i))^2$. Since $\mathbb{F}_n$ and $\mathbb{F}_n^0$ only differ in shrinking $(n^{-1/3})$ neighborhoods of the point of interest t_0, it is only observation times T_i in those neighborhoods that contribute to the sum. Thus, one could propose a test based on Diff_n. Note that $\text{Diff}_n = n\int (\mathbb{F}_n(t) - \mathbb{F}_n^0(t))^2 \, d\mathbb{G}_n(t)$ where $\mathbb{G}_n$ is the empirical distribution function of the observation times and hence measures the square of the L_2 distance between $\mathbb{F}_n$ and $\mathbb{F}_n^0$ with respect to the empirical measure of the observation times. We will use scaled/weighted versions of Diff_n to form our test statistics for testing H_0 against H_1. These are of the following types:

$$(i) \qquad T_{n,1} = \sum_{i=1}^n \frac{\left(\mathbb{F}_n(T_i) - \mathbb{F}_n^0(T_i)\right)^2}{\theta_0\,(1-\theta_0)} = n\int \frac{(\mathbb{F}_n(t) - \mathbb{F}_n^0(t))^2}{\theta_0(1-\theta_0)} d\mathbb{G}_n(t)\,,$$

$$(ii) \qquad T_{n,2} = \sum_{i=1}^n \frac{\left(\mathbb{F}_n(T_i) - \mathbb{F}_n^0(T_i)\right)^2}{\mathbb{F}_n(T_i)\,(1-\mathbb{F}_n(T_i))} = n\int \frac{(\mathbb{F}_n(t) - \mathbb{F}_n^0(t))^2}{\mathbb{F}_n(t)(1-\mathbb{F}_n(t))} d\mathbb{G}_n(t)\,,$$

$$(iii) \qquad T_{n,3} = \sum_{i=1}^n \frac{\left(\mathbb{F}_n(T_i) - \mathbb{F}_n^0(T_i)\right)^2}{\mathbb{F}_n^0(T_i)\,(1-\mathbb{F}_n^0(T_i))} = n\int \frac{(\mathbb{F}_n(t) - \mathbb{F}_n^0(t))^2}{\mathbb{F}_n^0(t)(1-\mathbb{F}_n^0(t))} d\mathbb{G}_n(t)\,.$$

We will show that these three score statistics $T_{n,j}$, $j = 1, 2, 3$, are asymptotically equivalent under both the null hypothesis and under local alternatives. Under the null hypothesis, these three score statistics have the same limiting null distribution (which is independent of the underlying parameters in the problem). On the other hand, unlike the situation in regular statistical problems in which the likelihood ratio, score (or Rao), and Wald statistics all have the same asymptotic chi-square distribution under the null hypothesis and the same non-central chi-square distribution under local alternatives, here the score statistics, Wald statistic, and likelihood ratio statistic all have *different limiting distribution* under the null hypothesis, and these differences persist in the limiting distributions under local alternatives.

While each of the score statistics can be regarded as an L_2 statistic, our reason for calling them "score statistics" is explained below.

The L_2 statistics as score statistics: The log-likelihood, based on n i.i.d. observations in the current status model, is given by (2.1) where F is the survival time distribution. Now consider the following perturbation of F in the direction of some other distribution function H on $[0, \infty)$:

$$F_{\epsilon,H} = (1 - \epsilon)\, F + \epsilon\, H\,,$$

where $\epsilon \geq 0$. This gives a one-dimensional parametric submodel of the original non-parametric model, passing through F and the (one-sided) score at F is then given by

$$\frac{\partial}{\partial\, \epsilon} \log L_n(F_{\epsilon,H})\ |_{\epsilon=0} \quad = \quad n\, \mathbb{P}_n \left(\frac{\Delta - F(T)}{F(T)\,(1 - F(T))}\, (H - F)(T) \right) .$$

Our score statistics in the interval censoring problem arise from suitable choices of H and F: we obtain $S_{n,1}$, our first score statistic, by setting $F = \mathbb{F}_n$, and $H = \mathbb{F}_n^0$; and $S_{n,2}$, our second score statistic, by taking $F = \mathbb{F}_n^0$, and $H = \mathbb{F}_n$. Thus, with $S_{n,1}$, we perturb the unconstrained MLE in the direction of the constrained MLE, and for $S_{n,2}$ we do the reverse. Since $\mathbb{F}_n$ maximizes the log-likelihood it is easy to see that $S_{n,1} \leq 0$. Thus

$$S_{n,1} = n\, \mathbb{P}_n \left(\frac{\Delta - \mathbb{F}_n(T)}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\, (\mathbb{F}_n^0 - \mathbb{F}_n)(T) \right) , \tag{2.2}$$

and

$$S_{n,2} = n\, \mathbb{P}_n \left(\frac{\Delta - \mathbb{F}_n^0(T)}{\mathbb{F}_n^0(T)\,(1 - \mathbb{F}_n^0(T))}\, (\mathbb{F}_n - \mathbb{F}_n^0)(T) \right) . \tag{2.3}$$

(In the context of regular parametric models $S_{n,1}$ is usually zero.) The following proposition connects $S_{n,1}$ and $S_{n,2}$ to both $T_{n,2}$ and $2 \log \lambda_n$: in a sense the scores $S_{n,1}$ and $S_{n,2}$ are the building blocks behind both these statistics.

Proposition 2.1 *Let $S_{n,1}$ and $S_{n,2}$ be defined as above and let the survival time distribution F belong to H_0. Then*

$$\begin{aligned} T_{n,2} &= S_{n,1} + S_{n,2} + O_p(n^{-1/6})\,, \\ 2\log\lambda_n &= S_{n,2} - S_{n,1} + o_p(1)\,. \end{aligned}$$

Thus $T_{n,2}$, which is one of the versions of the L_2 statistic, is asymptotically equivalent, under the null, to the sum of two one-sided score statistics, and the likelihood ratio statistic is asymptotically equivalent to the difference of these same two one-sided score statistics. We therefore refer to $T_{n,2}$ as a "score statistic". Since $T_{n,1}$ and $T_{n,3}$ are versions of the L_2 statistic which are asymptotically equivalent to $T_{n,2}$ we shall also regard them as "score statistics". The behavior of the one-sided score statistics under both the null hypothesis and contiguous alternatives will be discussed elsewhere.

The Wald statistic: It is known that

$$n^{1/3}\,(\mathbb{F}_n(t_0) - F(t_0)) \to_d \left(\frac{F(t_0)\,(1 - F(t_0))\,f(t_0)}{2\,g(t_0)}\right)^{1/3} 2\mathbb{Z}\,; \tag{2.4}$$

where $\mathbb{Z} = \operatorname{argmin}\{W(t) + t^2\}$ and W is a two-sided Brownian motion process starting from 0; see e.g. Groeneboom and Wellner (1992), Theorem 5.1, page 89. Thus a natural analogue W_n of the classical Wald statistic is defined by

$$W_n = \frac{n^{2/3}\,\left(\mathbb{F}_n(t_0) - F(t_0)\right)^2}{[\mathbb{F}_n(t_0)\,(1 - \mathbb{F}_n(t_0))\,\hat{f}_n(t_0)/2\,\hat{g}_n(t_0)]^{2/3}}$$

where $\hat{f}_n(t_0)$ and $\hat{g}_n(t_0)$ are consistent estimators of $f(t_0)$ and $g(t_0)$ respectively. Thus the Wald statistic is simply a scaled version of the squared distance between the MLE of F at the point t_0 and the true value $F(t_0)$. The scaling ensures that the limit distribution of the Wald statistic is free of underlying parameters. Note that, as opposed to the likelihood ratio and the score statistics, the Wald statistic requires explicit estimation of the (nuisance) parameters $f(t_0)$ and $g(t_0)$.

3 Limit Theory Under the Null Hypothesis

In this section we focus on the asymptotic behavior of the likelihood ratio, score, and Wald statistics under the null hypothesis $H_0 : F(t_0) = \theta_0$. We now introduce some notation: For constants $a, b > 0$ and $t \in \mathbb{R}$, let $X_{a,b}(t) \equiv$

$aW(t) + bt^2$, where $W(t)$ is standard two-sided Brownian motion starting from 0. Let $G_{a,b}$ denote the Greatest Convex Minorant (GCM) of the process $aW(t) + bt^2$. The GCM $G_{a,b}$ is well-characterized: it is a piecewise linear function that touches the paths of the process $X_{a,b}$ at the points where $G_{a,b}$ changes slope; furthermore the number of changes of slope of $G_{a,b}$ in any compact interval is finite. The right derivative process of $G_{a,b}$ is denoted by $g_{a,b}$. On the other hand $g^0_{a,b}$ is the process of right derivatives of $G^0_{a,b}$, the process of the constrained one-sided GCMs of $X_{a,b}(t)$. In other words, $G^0_{a,b}$ is the process, which for $t \geq 0$ is the GCM of $X_{a,b}(t)$ for $t \geq 0$, but subject to the constraint that its slopes do not fall below 0, and for $t < 0$ is the GCM of $X_{a,b}(t)$ for $t < 0$ but subject to the constraint that its slopes stay less than or equal to 0. See Groeneboom (1985), Groeneboom (1989), Banerjee (2000), Banerjee and Wellner (2001), and Wellner (2003) for more detailed descriptions of these processes.

Theorem 3.1 *Under the null hypothesis $F(t_0) = \theta_0$ and under assumption* **A**

$$\begin{aligned} 2\log\lambda_n &\to_d \int \left\{(g_{1,1}(z))^2 - (g^0_{1,1}(z))^2\right\} dz \equiv \mathbb{D}\,, \\ W_n &\to_d g^2_{1,1}(0) \stackrel{d}{=} 4\mathbb{Z}^2\,, \\ T_{n,2} &\to_d \int \left\{g_{1,1}(z) - g^0_{1,1}(z)\right\}^2 dz \equiv \mathbb{T}\,. \end{aligned}$$

Furthermore $T_{n,1}$ and $T_{n,3}$ are asymptotically equivalent to $T_{n,2}$ under H_0.

The proof of Theorem 3.1 is sketched in Section 5.

We now briefly discuss the distributions of $\mathbb{D}$, $\mathbb{T}$, and $4\mathbb{Z}^2$. Although the distributions of $\mathbb{D}$ and $\mathbb{T}$ are not yet known analytically, they can easily be estimated by Monte-Carlo methods. We have computed approximations to the distributions of $\mathbb{D}$ and $\mathbb{T}$ by constructing discrete approximations to Brownian motion on $[-3, 3]$ with a mesh size of .0002. For a discussion of this method, see Banerjee (2000) or Banerjee and Wellner (2001). Estimates of selected quantiles of the distribution of $\mathbb{D}$ based on a sample size of 3×10^4 along with the associated variability are provided in Banerjee and Wellner (2001).

The distribution of $\mathbb{Z}$ has been characterized analytically by Groeneboom (1985), Groeneboom (1989). The distribution function and quantiles of the distribution of $\mathbb{Z}$ have been computed by Groeneboom and Wellner (2001). Table 1 gives quantiles of $4\mathbb{Z}^2 = g_{1,1}(0)^2$ computed numerically from the density of $\mathbb{Z}$ along with the estimated quantiles of $\mathbb{D}$ and $\mathbb{T}$ obtained via discrete approximations to Brownian motion and Monte-Carlo. Even though the distributions of $\mathbb{D}$ and $\mathbb{T}$ are not analytically characterized, we do have a stochastic ordering result – namely that $\mathbb{D}$ is stochastically larger than $\mathbb{T}$.

Proposition 3.1

$$\mathbb{T} = \int \left(g_{1,1}(z) - g^0_{1,1}(z)\right)^2 \, dz \leq \int \left((g_{1,1}(z))^2 - (g^0_{1,1}(z))^2 \right) \, dz = \mathbb{D}\,.$$

Figure 1 shows the empirical distributions of $\mathbb{D}$ and $\mathbb{T}$ (based on the finite-grid approximations) and the distribution functions of $4\mathbb{Z}^2$ and χ^2_1 (computed numerically from the exact densities). The picture clearly suggests a stochastic ordering among three of these four random variables. Based on these numerical results we conjecture that $\mathbb{D} < 4\mathbb{Z}^2$ (stochastically). Although the distribution of χ^2_1 is above that of $4\mathbb{Z}^2$ for most of the range, it follows from Groeneboom (1989), Corollary 3.4, page 94, that the tails of the density of $\mathbb{Z}$ decay as $\exp(-(2/3)z^3)$, and hence it follows that the tail of the distribution of $4\mathbb{Z}^2$ decays as $\exp(-z^{3/2}/12)$. But the tail of χ^2_1 decays as $\exp(-z/2)$, and hence the tail of χ^2_1 should cross the tail of $4\mathbb{Z}^2$. In fact this crossing occurs somewhat before the .99 quantile as can be seen from Table 1.

Table 1: Estimated quantiles of the limit distributions $\mathbb{T}$ and $\mathbb{D}$ under the null compared to exact quantiles of χ_1^2 and $4\mathbb{Z}^2$

p	$F_{\mathbb{T}}^{-1}(p)$	$F_{\mathbb{D}}^{-1}(p)$	$F_{\chi_1^2}^{-1}(p)$	$F_{4\,\mathbb{Z}^2}^{-1}(p)$
.05	0.002412	0.002419	0.003932	0.004353
.10	0.009346	0.009847	0.015791	0.017475
.15	0.019687	0.022302	0.035766	0.039566
.20	0.034019	0.040279	0.064185	0.070965
.25	0.053003	0.063874	0.101531	0.112176
.30	0.075214	0.092615	0.148472	0.163893
.35	0.099452	0.128481	0.205900	0.227038
.40	0.132190	0.173011	0.274996	0.302833
.45	0.167592	0.224970	0.357317	0.392880
.50	0.210248	0.284706	0.454936	0.499308
.55	0.258662	0.351822	0.570652	0.624969
.60	0.322268	0.432697	0.708326	0.773794
.65	0.391991	0.527391	0.873457	0.951336
.70	0.480502	0.650435	1.074194	1.165780
.75	0.588160	0.802636	1.323304	1.429840
.80	0.716030	0.986587	1.642374	1.764830
.85	0.886630	1.230060	2.072251	2.210680
.90	1.139059	1.610734	2.705543	2.856650
.95	1.590908	2.286922	3.841459	3.985460
.99	2.758688	3.865057	6.6349	6.62198

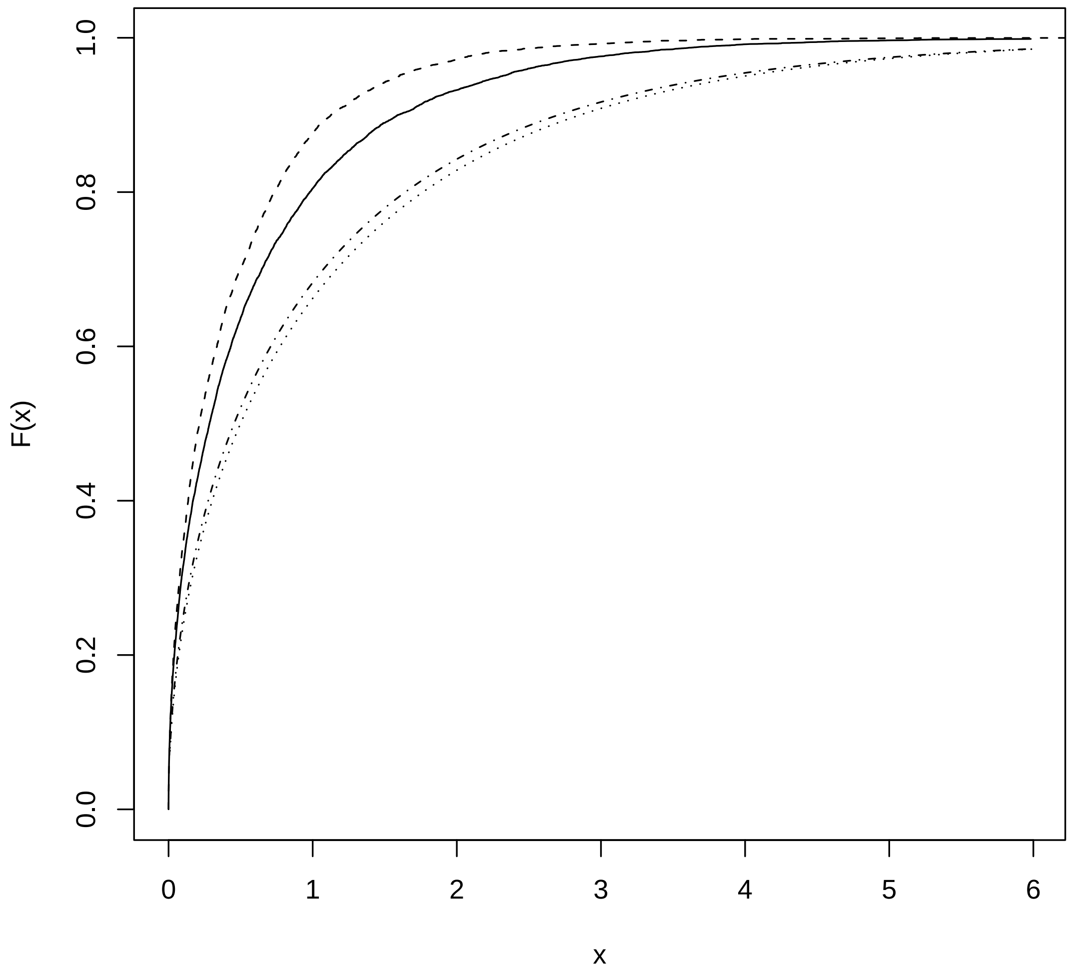

Figure 1: The distribution functions of $\mathbb{T}$ (dashed), $\mathbb{D}$ (solid), χ_1^2 (dots and dashes), and $4\mathbb{Z}^2$ (dots).

4 Limit Theory Under Local Alternatives

Our next theorem concerns the behavior of the three statistics under contiguous alternatives. Suppose that $\{F_n\}$ is a sequence of continuous

distribution functions satisfying the following conditions:
B(1). For some $c > 0$, $F_n(t) = F(t)$ for all t with $|t - t_0| \geq cn^{-1/3}$.
B(2). The functions $A_n(t) = n^{1/3}(F_n(t) - F(t))$ satisfy

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

uniformly for $z \in [-c, c]$.

Thus B is a continuous function on $[-c, c]$ and B vanishes on $(-c, c)^c$. It is shown in Theorem 2.6 of Banerjee and Wellner (2001) that $\{P^n_{F_n,G}\}$ and $\{P^n_{F,G}\}$ are mutually contiguous through a LAN (Local Asymptotic Normality) result for the local log-likelihood ratio $\log(L_n(F_n)/L_n(F))$. Now, let

$$\Psi(z) \;=\; \left\{ \begin{array}{ll} g(t_0) \int_0^{z\wedge c} B(y)\, dy\,, & z \geq 0 \\ -g(t_0) \int_{z\vee -c}^0 B(y)\, dy\,, & z < 0 \end{array} \right\}, \tag{4.1}$$

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

and let ϕ be defined by

$$\phi(t) = (b^{1/3}/a^{4/3})\Psi((b/a)^{-2/3}\, t) \tag{4.2}$$

where a and b are given by

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

Finally define the process $g_{a,b,\Psi}$ as the slope process of $G_{a,b,\Psi}$, the greatest convex minorant of $X_{a,b,\Psi}$, and the process $g^0_{a,b,\Psi}$ as the slope process of $G^0_{a,b,\Psi}$, where $G^0_{a,b,\Psi}$ is defined in terms of $X_{a,b,\Psi}$ in the exact same way as $G^0_{a,b}$ is defined in terms of $X_{a,b}$.

Theorem 4.1 *Suppose that the distribution functions F, G satisfy assumption* **A**, *and the sequence of distribution functions $\{F_n\}$ satisfies* **B(1)** *and* **B(2)**. *Then, under the local alternatives $\{P^n_{F_n,G}\}$,*

$$2\log\lambda_n \;\rightarrow_d\; \int \left((g_{1,1,\phi}(z))^2 - (g^0_{1,1,\phi}(z))^2\right) dz \equiv \mathbb{D}_\phi\,, \tag{4.3}$$

$$W_n \;\rightarrow_d\; g_{1,1,\phi}(0)^2 \equiv 4\mathbb{Z}^2_\phi\,, \tag{4.4}$$

$$T_{n,2} \;\rightarrow_d\; \int \left(g_{1,1,\phi}(z) - g^0_{1,1,\phi}(z)\right)^2 dz \equiv \mathbb{T}_\phi \tag{4.5}$$

where ϕ is as defined in (4.2). Furthermore, $T_{n,1}$ and $T_{n,3}$ are asymptotically equivalent to $T_{n,2}$ under the given sequence of contiguous alternatives and hence have the same asymptotic distribution.

Theorem 4.1 yields the limiting power of the tests under sequences $\{F_n\}$ satisfying $\mathbf{B(1)}$ and $\mathbf{B(2)}$. We conjecture that the distributions of $\mathbb{D}_\phi$, $\mathbb{Z}^2_\phi$, and $\mathbb{T}_\phi$ are continuous under the hypotheses of Theorem 4.1. We also conjecture that the limiting power of all three tests is strictly greater than the level (or size) α when $\phi'(0) \neq 0$: with $\alpha \in (0,1)$, if d_α, t_α, and z_α satisfy $P(\mathbb{D} > d_\alpha) = \alpha$, $P(4\mathbb{Z}^2 > 4z_\alpha) = \alpha$, and $P(\mathbb{T} > t_\alpha) = \alpha$ respectively, then $P(\mathbb{D}_\phi > d_\alpha) > \alpha$, $P(4\mathbb{Z}^2_\phi > 4z_\alpha) > \alpha$, and $P(\mathbb{T}_\phi > t_\alpha) > \alpha$. Unfortunately we have not succeeded in proving these conjectures, and furthermore the expressions for the limit variables $\mathbb{D}_\phi$, $\mathbb{T}_\phi$, and $\mathbb{Z}^2_\phi$ do not (yet) seem to yield insight into the comparative behavior of the power of the three tests. For this reason, we also consider the behavior of the tests under fixed alternatives and describe some preliminary Monte-Carlo studies.

5 Limit Theory Under Fixed Alternatives

The following theorem summarizes the behavior of the likelihood ratio, score, and Wald statistics under a fixed alternative.

Theorem 5.1 *Suppose that $F(t_0) \neq \theta_0$, that there is a unique point t_1 satisfying $F(t_1) = \theta_0$, and there is a neighborhood $\mathcal{N}$ of t_0 such that F and G are continuously differentiable on $\mathcal{N}$ 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 $[t_1 \wedge t_0, t_1 \vee t_0] \subset (c,d)$ and each $t \in (c,d)$ is a support point of G. Let the distribution function H be defined by*

$$H(t) = \begin{cases} F(t) \vee \theta_0\,, & t \geq t_0 \\ F(t) \wedge \theta_0\,, & t < t_0\,. \end{cases} \tag{5.1}$$

Then the following hold:

(i) $n^{-1} 2\log\lambda_n \to_p 2K(P_{F,G}, P_{H,G}) > 0$ *where* $K(P,Q) = E_P \log(dP/dQ)$ *is the Kullback-Leibler discrepancy between P and Q.*

(ii) Both $n^{-1}\,T_{n,1}$ and $n^{-1}\,T_{n,3}$ converge in probability to the positive constant

$$\frac{1}{\theta_0\,(1-\theta_0)} \parallel F - H \parallel^2_{L_2(G)} \equiv B_1(F, G, \theta_0)\,.$$

(iii) $n^{-1}\,T_{n,2}$ converges in probability to

$$\int_{[t_1\wedge t_0, t_1\vee t_0]} \frac{(F(t) - H(t))^2}{F(t)(1-F(t))}\,dG(t) \equiv B_2(F, G, \theta_0)\,.$$

(iv) $n^{-2/3} W_n \to_p (\theta_1 - \theta_0)^2/[\theta_1 (1-\theta_1) f(t_0)/2\, g(t_0)]^{2/3}$.

(v) The constants $2K(P_{F,G}, P_{H,F})$ *and* $B_i(F,G,\theta_0)$, $i = 1,2$, *satisfy the following inequalities:*
(a) $B_1(F,G,1/2) \le 2K(P_{F,G}, P_{H,F}) \le B_2(F,G,1/2)$.
(b) If $\theta_0 \le 1/2$ *and* $H(t) \ge F(t)$ *for* $t \in [t_0, t_1]$, *or if* $\theta_0 \ge 1/2$ *and* $H(t) \le F(t)$
for $t \in [t_0 \wedge t_1, t_0 \vee t_1]$, *then*

$$B_1(F,G,\theta_0) \le 2K(P_{F,G}, P_{H,F}) \le B_2(F,G,\theta_0)\,.$$

(c) If $\theta_0 \le 1/2$ *and* $H(t) \le F(t) \le 1/2$ *for* $t \in [t_0 \wedge t_1, t_0 \vee t_1]$, *or if* $\theta_0 > 1/2$ *and*
$1/2 \le F(t) \le H(t)$ *for* $t \in [t_0 \wedge t_1, t_0 \vee t_1]$, *then*

$$B_2(F,G,\theta_0) \le 2K(P_{F,G}, P_{H,F}) \le B_1(F,G,\theta_0)\,.$$

Note that (i) - (iv) of Theorem 5.1 imply that the Wald statistic diverges to infinity in probability at a slower rate ($n^{2/3}$) than the rate at which the likelihood ratio and score statistics converge to infinity. This leads us to conclude that the power of the Wald test is typically dominated by the power of the likelihood ratio and score tests, and this matches with our preliminary Monte Carlo studies. Parts (v)(a) and (b) of Theorem 5.1 lead us to suspect that the test based on the score statistic $T_{n,2}$ has higher power than the other score statistics and the likelihood ratio statistics when $\theta_0 = 1/2$ or when $1/2 \ge \theta_0 = H(t) > F(t)$ for $t \in [t_0, t_1)$ or when $1/2 \le \theta_0 = H(t) < F(t)$ for $t \in (t_1, t_0]$, and this is also borne out by the preliminary Monte Carlo work.

6 Monte Carlo study of the Power of the Tests

Simulation studies: We now study the rejection probabilities of the likelihood ratio test, the three (asymptotically equivalent) score statistics and the Wald statistic, for different values of the constants a, b and c, at different sample sizes n via Monte Carlo, and compare with the limiting power predicted by theory. Since our proposed tests are based on asymptotic critical points, there are obvious difficulties in interpretation of the rejection probabilities as "power" since the finite sample type I rejection probabilities differ from the nominal (asymptotic) values. Thus we have also carried out comparisons based on finite-sample Monte-Carlo critical values for a subset of the particular situations used in the Monte-Carlo study. It should be understood that these latter comparisons also suffer from the difficulty that such critical points are not available in practice. [Potential ways of solving the problem of accurate finite sample critical points might involve bootstrap methods or other resampling strategies, but we have not explored such methods here.]

One natural type of local alternatives $\{F_n\}$ to consider are of the form

$$F_n(t) = F(t) + n^{-1/3} B(n^{1/3}(t - t_0)) \tag{6.1}$$

where $B(z) = \eta f(t_0)(c - |z|) 1_{[-c,c]}(z)$ and t_0, c, and η are fixed positive numbers. The difficulty with this sequence $\{F_n\}$ for Monte-Carlo experiments is that F_n^{-1}, does not have a closed form, so that generating (pseudo-) random variables from the distribution function F_n is complicated. It is, however, relatively easy to simulate from the distributions which are piecewise linear on $[t_0 - cn^{-1/3}, t_0 + cn^{-1/3}]$ and agree with the distribution functions F_n at the points $t_0 - cn^{-1/3}$, t_0, and $t_0 + cn^{-1/3}$. Thus we simulated from the sequence of distribution functions G_n defined by

$$G_n(t) = \begin{cases} F(t)\,, & t \le t_0 - cn^{-1/3} \\ F(t_0 - cn^{-1/3}) & t_0 - cn^{-1/3} \le t \le t_0\,, \\ \quad + \frac{(t - t_0 + cn^{-1/3})(F_n(t_0) - F(t_0 - cn^{-1/3}))}{cn^{-1/3}}\,, & \\ F_n(t_0) + \frac{(t - t_0)(F(t_0 + cn^{-1/3}) - F_n(t_0))}{cn^{-1/3}}\,, & t_0 \le t \le t_0 + cn^{-1/3}\,, \\ F(t)\,, & t \ge t_0 + cn^{-1/3}\,, \end{cases}$$

with inverses given by

$$G_n^{-1}(u) = \begin{cases} F^{-1}(u), & u \le F(t_0 - cn^{-1/3})\,, \\ t_0 - cn^{-1/3} + \frac{cn^{-1/3}(u - F(t_0 - cn^{-1/3}))}{F_n(t_0) - F(t_0 - cn^{-1/3})}, & F(t_0 - cn^{-1/3}) \le u \le F_n(t_0) \\ t_0 + \frac{cn^{-1/3}(u - F_n(t_0))}{F(t_0 + cn^{-1/3}) - F_n(t_0)}, & F_n(t_0) \le u \le F(t_0 + cn^{-1/3})\,, \\ F^{-1}(u), & u \ge F(t_0 + cn^{-1/3})\,. \end{cases}$$

The distribution function G_n agrees with F (and F_n) on the complement of the interval $[t_0 - cn^{-1/3}, t_0 + cn^{-1/3}]$. Note that it is quite easy to generate observations from the sequence G_n, since there is an explicit expression for G_n^{-1} (so long as we have an explicit expression for F^{-1}). We have the following proposition.

Proposition: $C_n(z) \equiv n^{1/3}(G_n(t_0 + n^{-1/3}\, z) - F(t_0 + n^{-1/3}\, z))$, $-c \le z \le c$, converges uniformly to $B(z)$ on $[-c, c]$, where $B(z) = \eta f(t_0)(c - |z|) 1_{[-c,c]}(z)$. Consequently, the alternatives $\{G_n\}$ are contiguous.

The straightforward proof is omitted.

We proceeded to generate the likelihood ratio statistic for contiguous alternatives of the above kind, at different settings of the underlying parameters and for different values of n. We also generated all three score statistics and the Wald statistic. Recall that by contiguity, the three different versions of the score statistic are all asymptotically equivalent. The actual values, $f(t_0)$ and $g(t_0)$, of the densities (which were known) were used to compute the Wald statistic; note that in a real life situation these would need to be estimated. Thus, the Wald statistic that we computed is in some sense an "ideal" version. The asymptotic distribution of the likelihood ratio statistic under the sequence of contiguous alternatives F_n (and also G_n) is that of the random variable,

$$\mathbb{D}_\phi = \int \left((g_{1,1,\phi}(z))^2 - (g^0_{1,1,\phi}(z))^2 \right) \, dz \, ,$$

the asymptotic distribution of each of the score statistics is

$$\mathbb{T}_\phi = \int \left(g_{1,1,\phi}(z) - g^0_{1,1,\phi}(z) \right)^2 \, dz \, ,$$

and the asymptotic distribution of the Wald statistic is that of $g_{1,1,\phi}(0)^2$. Recall that $\phi(t) = (b/a)^{4/3}\, \Psi((b/a)^{-2/3}\, t)$ and Ψ is as defined in (4.1).

The parameter settings are as follows:

- $F \sim U(0,1)$, $G \sim U(0,1)$, $t_0 = 0.5$, so that $F(t_0) = 0.5$. The value of b/a in this situation is 1. The values of c chosen were 0, 0.5, 1 and 2. The value of η was taken to be 0.9.

The sample sizes chosen were $n = 100, 300, 500, 800, 1000, 2000, 4000, 8000$. For each value of n and each setting of the parameter c (there are four settings considered here), a sample of size 2000 (for $c > 0$)/ 5000 (for $c = 0$) was generated from the distributions of the likelihood ratio statistic, the three different versions of the score statistic and the Wald statistic. The rejection probability for each of these statistics was then computed at the nominal level $\alpha = 0.05$ by computing the proportion of values that exceeded the $1 - \alpha$'th quantile of the corresponding limit distribution under the null. Finally, the asymptotic rejection probability of each statistic was obtained under each parameter setting with $c \neq 0$ ($c = 0$ corresponds to the null hypothesis case) by generating a sample of size 2000 from (discrete approximations to) the limit distribution under that parameter setting. Thus, these samples were obtained by generating discrete approximations to two-sided Brownian motion on a grid over the interval $[-3, 3]$ with a grid size of .0002, adding on the function $t^2 + \phi(t)$ to the Brownian motion path and subsequently differentiating the unconstrained and constrained minorants. The results from these simulations are presented below; these allow us to compare the rejection probabilities of the three competing statistics under this particular sequence of contiguous alternatives.

Table 2: Rejection probabilities for nominal level 0.05 with $b/a = 1, c = 0$ (null hypothesis).

n	100	300	500	800	∞
$2\ \log(\lambda_n)$	0.0536	0.0540	0.0504	0.0546	0.05
$T_{n,1}$	0.0520	0.0566	0.0562	0.0580	0.05
$T_{n,2}$	0.0548	0.0590	0.0582	0.0594	0.05
$T_{n,3}$	0.0522	0.0566	0.0562	0.0580	0.05
W_n	0.0592	0.0622	0.0538	0.0564	0.05
n	1000	2000	4000	8000	∞
$2\ \log(\lambda_n)$	0.0536	0.0540	0.0504	0.0546	0.05
$T_{n,1}$	0.0520	0.0566	0.0562	0.0580	0.05
$T_{n,2}$	0.0548	0.0590	0.0582	0.0594	0.05
$T_{n,3}$	0.0522	0.0566	0.0562	0.0580	0.05
W_n	0.0592	0.0622	0.0538	0.0564	0.05

Table 3: Rejection probabilities for nominal level 0.05 with $b/a = 1, c = 0.5$.

n	1000	2000	4000	8000	∞
$2\ \log(\lambda_n)$	0.077	0.0705	0.0690	0.0725	0.0730
$T_{n,1}$	0.075	0.0750	0.0735	0.0725	0.0705
$T_{n,2}$	0.079	0.0800	0.0755	0.0730	0.0705
$T_{n,3}$	0.075	0.0755	0.0735	0.0725	0.0705
W_n	0.080	0.0700	0.0645	0.0725	0.0660
n	1000	2000	4000	8000	∞
$2\ \log(\lambda_n)$	0.077	0.0705	0.0690	0.0725	0.0730
$T_{n,1}$	0.075	0.0750	0.0735	0.0725	0.0705
$T_{n,2}$	0.079	0.0800	0.0755	0.0730	0.0705
$T_{n,3}$	0.075	0.0755	0.0735	0.0725	0.0705
W_n	0.080	0.0700	0.0645	0.0725	0.0660

Table 4: Rejection probabilities for nominal level 0.05 with $b/a = 1, c = 1.0$.

n	100	300	500	800	∞
$2\log(\lambda_n)$	0.2655	0.2465	0.2520	0.2600	0.2610
$T_{n,1}$	0.2655	0.2530	0.2505	0.2605	0.2565
$T_{n,2}$	0.3190	0.2725	0.2695	0.2720	0.2565
$T_{n,3}$	0.2670	0.2530	0.2525	0.2610	0.2565
W_n	0.2810	0.2530	0.2510	0.2500	0.2350
n	1000	2000	4000	8000	∞
$2\log(\lambda_n)$	0.2610	0.2535	0.2560	0.2535	0.2610
$T_{n,1}$	0.2750	0.2705	0.2485	0.2555	0.2565
$T_{n,2}$	0.2835	0.2760	0.2525	0.2580	0.2565
$T_{n,3}$	0.2750	0.2715	0.2490	0.2555	0.2565
W_n	0.2505	0.2375	0.2275	0.2275	0.2350

Table 5: Rejection probabilities for nominal level 0.05 with $b/a = 1.0, c = 2.0$.

n	100	300	500	800	∞
$2\log(\lambda_n)$	0.9130	0.8945	0.8870	0.8810	0.8685
$T_{n,1}$	0.9340	0.9165	0.9115	0.9035	0.8955
$T_{n,2}$	0.9525	0.9265	0.9200	0.9055	0.8955
$T_{n,3}$	0.9345	0.9165	0.9115	0.9035	0.8955
W_n	0.9365	0.9040	0.8955	0.8720	0.8650
n	1000	2000	4000	8000	∞
$2\log(\lambda_n)$	0.8755	0.8655	0.8625	0.8740	0.8685
$T_{n,1}$	0.9040	0.9030	0.8935	0.9095	0.8955
$T_{n,2}$	0.9100	0.9050	0.8950	0.9110	0.8955
$T_{n,3}$	0.9045	0.9030	0.8935	0.9095	0.8955
W_n	0.8775	0.8645	0.8630	0.8600	0.8650

Finally, we also carried out a set of experiments in which we first determined finite sample critical points for the various statistics, and then used these critical points to determine the corresponding rejection probabilities. We first did this for the same alternatives as used before, (with F uniform$(0,1)$, G uniform$(0,1)$, $\theta_0 = t_0 = 1/2$, 10000 Monte Carlo replicates), and then for a new situation with F uniform$(0,1)$, G uniform$(0,1)$, $\theta_0 = t_0 = 2/3$, 2000 Monte Carlo replicates),

Table 6 presents ideal 95% quantiles for the situation with F uniform$(0,1)$, G uniform$(0,1)$, $\theta_0 = t_0 = 1/2$:

Table 6: Finite sample 95% quantiles, null hypothesis

n	$2\log(\lambda_n)$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	W_n
2000	2.276517	1.607356	1.640066	1.608484	4.049419
4000	2.222080	1.607599	1.631386	1.609309	4.003441
∞	2.287	1.599	1.599	1.599	3.985

The resulting rejection probabilities (based on 10000 Monte Carlo replicates) for the contiguous alternatives with $c = 1$, $c = 2$, and $\eta = .9$ are given as follows:

Table 7: Rejection probabilities based on (ideal) finite sample critical points, $\theta_0 = t_0 = 1/2$, $c = 1$

n	$2\log(\lambda_n)$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	W_n
2000	0.2505	0.2511	0.2518	0.2514	0.2241
4000	0.2606	0.2552	0.2550	0.2552	0.2315
∞	0.2610	0.2565	0.2565	0.2565	0.2350

Table 8: Rejection probabilities based on (ideal) finite sample critical points, $\theta_0 = t_0 = 1/2$, $c = 2$

n	$2\log(\lambda_n)$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	W_n
2000	0.8774	0.9035	0.9043	0.9034	0.8674
4000	0.8815	0.9013	0.9018	0.9010	0.8652
∞	0.8685	0.8955	0.8955	0.8955	0.8650

Table 9 presents "ideal" 95% quantiles under the null hypothesis in the new situation with F uniform$(0,1)$, G uniform$(0,1)$, $\theta_0 = t_0 = 2/3$, 2000 Monte Carlo replicates).

Table 9: Finite sample 95% quantiles, null hypothesis

n	$2\log(\lambda_n)$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	W_n
2000	2.342	1.600	1.627	1.600	4.177
4000	2.233	1.614	1.615	1.618	4.199
∞	2.287	1.599	1.599	1.599	3.985

The resulting rejection probabilities (based on 2000 Monte Carlo replicates) for the contiguous alternative with $c = 1$ and $\eta = .9$ are given as follows:

Table 10: Rejection probabilities based on (ideal) finite sample critical points, $\theta_0 = t_0 = 2/3$

n	$2\log(\lambda_n)$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	W_n
2000	0.2865	0.2950	0.3385	0.2970	0.2835
4000	0.2845	0.2740	0.3040	0.2755	0.2615

Similarly, but now for the situation of $\theta_0 = t_0 = .8$, $\eta = .9$, and $c = 1$ (but now not showing the finite-sample critical points), we find the following rejection probabilities based on 2000 Monte Carlo replications:

Table 11: Rejection probabilities based on (ideal) finite sample critical points, $\theta_0 = t_0 = .8$

n	$2\log(\lambda_n)$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	W_n
2000	0.399	0.3945	0.4830	0.4005	0.4390
4000	0.396	0.3770	0.4855	0.4460	0.3965

7 Conclusions

We have introduced and studied five different test statistics for testing $H : F(t_0) = \theta_0$ versus $K : F(t_0) \neq \theta_0$ in the current status model: the Wald statistic, likelihood ratio statistic, and three variants of a score statistic. All five tests can be implemented by use of asymptotic critical values based on either theoretical critical values or estimated critical values based on simulations.

The asymptotic theory for the five statistics under the null hypothesis and local and fixed alternatives together with preliminary Monte Carlo studies given in section 7 (and some not shown here) lead to the following conclusions: (i) The score test $T_{n,2}$ and the Wald test based on the asymptotic critical points are both anti-conservative for moderate samples sizes. (ii) On the other hand, the likelihood ratio test and the score tests based on $T_{n,1}$ and $T_{n,3}$ and the asymptotic critical points have size quite close to the nominal one for moderate sample sizes. (iii) The rejection probabilities of the likelihood ratio test and the score tests dominate those of the Wald test. This is consistent with the fact that the Wald statistic grows at rate $n^{2/3}$ under fixed alternatives, while the likelihood ratio and score statistics grow at rate n under such alternatives. (iv) Among the three score statistics, $T_{n,2}$ seems to have slightly higher rejection probabilities under alternatives than the other two score statistics or the likelihood ratio statistic. This is consistent with the limit theory under fixed alternatives in view of (v)(b) of Theorem 5.1. Further study of the power behavior of these statistics is needed. Pointwise confidence intervals based on the likelihood ratio statistics are compared to confidence sets based on the Wald statistics and other competitors in Banerjee and Wellner (2005).

8 Proofs

To prepare for the proof of Proposition 2.1 and Theorems 3.1 and 4.1, we define the processes X_n and Y_n (as in Banerjee and Wellner (2001)) by

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

We will need the following result concerning the behavior of the processes X_n and Y_n under a sequence of contiguous alternatives.

Theorem 8.1 *Suppose that the distribution functions F, G satisfy assumption* **A***, and the sequence of distribution functions $\{F_n\}$ satisfies* **B(1)** *and* **B(2)***. Then the finite dimensional marginals of the process $(X_n(z), Y_n(z))$, considered as a process in the space $\mathcal{L}^p[-K,K] \times \mathcal{L}^p[-K,K]$ converge to the finite dimensional marginals of the process $(1/g(t_0))\,(g_{a,b,\Psi}(z), g^0_{a,b,\Psi}(z))$ under the sequence of (contiguous) alternatives $\{P^n_{F_n,G}\}$, where*

Furthermore it is also the case that under this sequence, for any $p \geq 1$,

$$(X_n, Y_n) \to_d (1/g(t_0))\,(g_{a,b,\Psi}, g^0_{a,b,\Psi})$$

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

For a sketch of the proof of Theorem 7.1 see BANERJEE AND WELLNER (2001). Theorem 8.1 will be used in characterizing the behavior of the three types of test statistics under alternatives $\{F_n\}$.

Proof of Proposition 2.1. First, note that $T_{n,2}$ can be written as

$$T_{n,2} = n\,\mathbb{P}_n\left(\frac{(\mathbb{F}_n(T) - \mathbb{F}_n^0(T))^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\right).$$

On the other hand, by straightforward algebra we find that

$$\begin{aligned} S_{n,2} + S_{n,1} &= n\,\mathbb{P}_n\left(\frac{\Delta - \mathbb{F}_n^0(T)}{\mathbb{F}_n^0(T)\,(1 - \mathbb{F}_n^0(T))}\,(\mathbb{F}_n - \mathbb{F}_n^0)(T)\right) \\ &\quad - n\,\mathbb{P}_n\left(\frac{\Delta - \mathbb{F}_n(T)}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\,(\mathbb{F}_n - \mathbb{F}_n^0)(T)\right) \\ &= n\,\mathbb{P}_n\left(\frac{(\mathbb{F}_n(T) - \mathbb{F}_n^0(T))^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\right) + n\,\mathbb{P}_n(r_n) = T_{n,2} + n^{1/3}\,\mathbb{P}_n(\tilde{r}_n)\,, \end{aligned}$$

where $r_n = r_n(\Delta, T)$ is the random function given by

$$r_n(\Delta, T) = \frac{(\Delta - \mathbb{F}_n^0(T))(\mathbb{F}_n - \mathbb{F}_n^0)^2(T)(1 - \mathbb{F}_n(T) - \mathbb{F}_n^0(T))}{\mathbb{F}_n(T)(1 - \mathbb{F}_n(T))\,\mathbb{F}_n^0(T)(1 - \mathbb{F}_n^0(T))}\,,$$

and $\tilde{r_n} = n^{2/3} r_n$. It can be shown, by writing $n^{1/3}\mathbb{P}_n\tilde{r}_n$ as $n^{1/3}(\mathbb{P}_n - P)\tilde{r}_n + n^{1/3}P\tilde{r}_n$, that $n^{1/3}(\mathbb{P}_n - P)\tilde{r}_n = O_p(n^{-1/6})$ whereas $n^{1/3}P\tilde{r}_n = O_p(n^{-1/3})$, showing that $n^{1/3}\mathbb{P}_n\tilde{r}_n = O_p(n^{-1/6})$. This yields the first equality of the proposition.

Now we prove the second equality. By standard preservation properties for Donsker classes of functions and using the facts that $\mathbb{F}_n$ and $\mathbb{F}_n^0$ are monotone and eventually bounded away from 0 and 1 on the difference set with arbitrarily high probability and that $n^{1/3}(\mathbb{F}_n - \mathbb{F}_n^0)$ is eventually bounded with arbitrarily high probability, it can be shown that $\tilde{r}_n$ eventually lies in a uniformly bounded Donsker class of functions with arbitrarily high probability, and hence it follows that $\sqrt{n}(\mathbb{P}_n - P)\,\tilde{r}_n$ is $O_p(1)$; consequently, $n^{1/3}\,(\mathbb{P}_n - P)\,\tilde{r}_n$ is $O_p(n^{-1/6})$.

Let D_n again denote the set where $\mathbb{F}_n$ and $\mathbb{F}_n^0$ differ, $\tilde{D}_n = n^{1/3}\,(D_n - t_0)$, and $t_n(z) = t_0 + z\,n^{-1/3}$, $X_n(z) = n^{1/3}\,(\mathbb{F}_n(t_n(z)) - F(t_0))$ and $Y_n(z) = n^{1/3}\,(\mathbb{F}_n^0(t_n(z)) - F(t_0))$. Thus $n^{1/3}\,P\,\tilde{r}_n$ can be rewritten as

$$n^{-1/3}\int_{\tilde{D}_n} \frac{(n^{1/3}(F(t_n(z)) - F(t_0)) - Y_n(z))\,(X_n(z) - Y_n(z))^2}{\mathbb{F}_n\,\mathbb{F}_n^0\,(1 - \mathbb{F}_n)\,(1 - \mathbb{F}_n^0)(t_n(z))}$$

$$
\begin{aligned}
& \qquad (1-\mathbb{F}_n(t_n(z)) - \mathbb{F}_n^0(t_n(z)))\, g(t_n(z))\, dz \\
= \;& n^{-1/3} \int_{\tilde{D}_n} \frac{(z\, f(t_0) + o(1) - Y_n(z))(X_n(z) - Y_n(z))^2}{\mathbb{F}_n\, \mathbb{F}_n^0\, (1-\mathbb{F}_n)\, (1-\mathbb{F}_n^0)(t_n(z))} \\
& \qquad (1-\mathbb{F}_n(t_n(z)) - \mathbb{F}_n^0(t_n(z)))\, g(t_n(z))\, dz
\end{aligned}
$$

where the $o(1)$ term converges to 0 uniformly in $z \in [-K, K]$ for any $K > 0$. By Proposition 8.1 and the fact that $\mathbb{F}_n(t_n(z))$ and $\mathbb{F}_n^0(t_n(z))$ are eventually bounded away from 0 and 1 with arbitrarily high probability for z in a compact set, it follows that the integral in the last expression of the above display is $O_p(1)$ and hence $n^{-1/3}$ times the integral is $O_p(n^{-1/3})$.

To prove the second claim of Proposition 2.1, we proceed as in the proof of Theorem 3.1 to show that

$$S_{n,1} = -\frac{g(t_0)}{\theta_0(1-\theta_0)} \int_{\tilde{D}_n} Y_n(z)\,(X_n(z) - Y_n(z))\, dz + o_p(1)\,,$$

and

$$S_{n,2} = \frac{g(t_0)}{\theta_0(1-\theta_0)} \int_{\tilde{D}_n} X_n(z)\,(X_n(z) - Y_n(z))\, dz + o_p(1)\,.$$

Using $a(a-b) + b(a-b) = a^2 - b^2$ and taking $a = X_n(z)$, $b = Y_n(z)$, it follows that

$$
\begin{aligned}
S_{n,2} - S_{n,1} &= \frac{g(t_0)}{\theta_0(1-\theta_0)} \int_{\tilde{D}_n} \{X_n^2(z) - Y_n^2(z)\}\, dz + o_p(1) \\
&= 2\log\lambda_n + o_p(1)
\end{aligned}
$$

by the proof of Theorem 2.5, Banerjee and Wellner (2001). □

Proof of Proposition 3.1. Since $S_{n,1} \le 0$ (as noted above (2.2)), the claim follows from Proposition 2.1 and Theorem 3.1. □

Before proving Theorems 3.1, 4.1, and 5.1, we state two useful lemmas.

Lemma 8.1 *Suppose that* $\{U_{n\epsilon}\}$, $\{V_n\}$, *and* $\{U_\epsilon\}$ *are three sets of random variables such that*
(i) $\lim_{\epsilon\to 0} \limsup_{n\to\infty} P[U_{n\epsilon} \neq V_n] = 0$;
(ii) $\lim_{\epsilon\to 0} P[U_\epsilon \neq V] = 0$;
(iii) for every $\epsilon > 0$, $U_{n\epsilon} \to_d U_\epsilon$ *as* $n \to \infty$.
Then $V_n \to_d V$ *as* $n \to \infty$.

Lemma 7.1 is Lemma 4.2 in Prakasa Rao (1969), page 30. The next lemma is a consequence of Brownian scaling arguments.

Lemma 8.2 *For any $M > 0$, the following distributional equality holds in the space $\mathcal{L}^2[-M, M] \times \mathcal{L}^2[-M, M]$:*

$$\left(g_{a,b,\Psi}(t), g^0_{a,b,\Psi}(t)\right) \stackrel{d}{=} \left(a\,(b/a)^{1/3} g_{1,1,\phi}\left((b/a)^{2/3} t\right), a\,(b/a)^{1/3} g^0_{1,1,\phi}\left((b/a)^{2/3} t\right)\right) .$$

where ϕ is as defined in (4.2).

Proof of Lemma 7.2. This is a straightforward extension of the equality in law under the the null hypothesis extablished in Banerjee and Wellner (2001), (6.8), page 1724, upon using (4.2) and the definitions. □

Proof of Theorem 3.1. The first assertion, $2 \log \lambda_n \to_d \mathbb{D}$, is proved in Banerjee and Wellner (2001), Theorem 2.5. Convergence in distribution of W_n to $4\mathbb{Z}^2$ follows immediately from (2.4) via continuous mapping and Slutsky's theorem.

It remains to deal with $T_{n,2}$. Denote the set where $\mathbb{F}_n$ and $\mathbb{F}^0_n$ differ by D_n and the transformed difference set corresponding to the local variable, $z = n^{1/3}\,(T - t_0)$, by $\tilde{D}_n$. We also denote $t_0 + n^{-1/3}\,z$ by $t_n(z)$. Now,

$$\begin{aligned} T_{n,2} &= \sum_{i=1}^{n} \frac{\left(\mathbb{F}_n(T_i) - \mathbb{F}^0_n(T_i)\right)^2}{\mathbb{F}_n(T_i)\,(1 - \mathbb{F}_n(T_i))} \\ &= n\,\mathbb{P}_n \left(\frac{\left(\mathbb{F}_n(T) - \mathbb{F}^0_n(T)\right)^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\,1_{D_n}(T)\right) \\ &= n\,(\mathbb{P}_n - P)\left(\frac{\left(\mathbb{F}_n(T) - \mathbb{F}^0_n(T)\right)^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\,1_{D_n}(T)\right) \\ &\qquad + nP\left(\frac{\left(\mathbb{F}_n(T) - \mathbb{F}^0_n(T)\right)^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\,1_{D_n}(T)\right) \\ &= T_{n,2,1} + T_{n,2,2}\,. \end{aligned}$$

But

$$T_{n,2,1} = n^{1/3}\,(\mathbb{P}_n - P)\left\{\frac{\left(n^{1/3}(\mathbb{F}_n(T) - \mathbb{F}^0_n(T))\right)^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\,1_{D_n}(T)\right\} ,$$

and this is $o_p(1)$ since the (random) function within the curly brackets can be shown to eventually lie in a uniformly bounded Donkser class of functions with arbitrarily high probability. Thus we only need to deal with $T_{n,2,2}$. Now,

$$T_{n,2,2} = n\,P\left(\frac{\left(\mathbb{F}_n(T) - \mathbb{F}^0_n(T)\right)^2}{\mathbb{F}_n(T)\,(1 - \mathbb{F}_n(T))}\,1_{D_n}(T)\right)$$

$$
\begin{aligned}
&= n \int_{D_n} \frac{\left(\mathbb{F}_n(t) - \mathbb{F}_n^0(t)\right)^2}{\mathbb{F}_n(t)\,(1-\mathbb{F}_n(t))}\, g(t)\, dt \\
&= n^{2/3} \int_{\tilde{D}_n} \frac{\left(\mathbb{F}_n(t_n(z)) - \mathbb{F}_n^0(t_n(z))\right)^2}{\mathbb{F}_n(t_n(z))\,(1-\mathbb{F}_n(t_n(z)))}\, g(t_n(z))\, dz \\
&= \int_{\tilde{D}_n} \frac{\left(X_n(z) - Y_n(z)\right)^2}{\mathbb{F}_n(t_n(z))\,(1-\mathbb{F}_n(t_n(z)))}\, g(t_n(z))\, dz \\
&= \frac{g(t_0)}{\theta_0\,(1-\theta_0)} \int_{\tilde{D}_n} \left(X_n(z) - Y_n(z)\right)^2\, dz + o_p(1)\,.
\end{aligned}
$$

The last step in the above string of equalities follows from the one above it using the continuity of g in a neighborhood of t_0, the almost sure uniform convergence of $\mathbb{F}_n$ to F in a neighborhood of t_0 and the continuity of F, the fact that $\tilde{D}_n$ is with arbitrarily high probability contained in a compact set around 0, eventually and the fact that the processes $X_n(z)$ and $Y_n(z)$ are bounded in probability on compact sets. Thus,

$$
T_{n,2} = \frac{g(t_0)}{\theta_0\,(1-\theta_0)} \int_{\tilde{D}_n} \left(X_n(z) - Y_n(z)\right)^2\, dz + o_p(1) \equiv T_n + o_p(1)\,. \tag{8.1}
$$

It can be shown by similar steps that the same representation holds for $T_{n,1}$ and $T_{n,3}$ under the null hypothesis, showing that all the three statistics are asymptotically equivalent to one another and to T_n. We will now deduce the limit distribution of T_n. To this end we will use Lemma 7.1. Now, for each $\epsilon > 0$ we can find an interval $[-K_\epsilon, K_\epsilon]$ such that

$$
\liminf_{n\to\infty} P(\tilde{D}_n \subset [-K_\epsilon, K_\epsilon]) > 1-\epsilon \quad \text{and} \quad P(D_{a,b} \subset [-K_\epsilon, K_\epsilon]) > 1-\epsilon\,.
$$

Here $D_{a,b}$ is the set on which the processes $g_{a,b}$ and $g_{a,b}^0$ differ. See the following Proposition 8.1 for proofs of these claims.

Now let

$$
\begin{aligned}
U_{n\epsilon} &= \frac{g(t_0)}{F(t_0)(1-F(t_0))} \int_{[-K_\epsilon, K_\epsilon]} \left(X_n(z) - Y_n(z)\right)^2\, dz\,, \\
U_\epsilon &= \int_{[-K_\epsilon, K_\epsilon]} \frac{1}{g(t_0)F(t_0)(1-F(t_0))} \left(g_{a,b}(z) - g_{a,b}^0(z)\right)^2\, dz\,, \\
V &= \int_{D_{a,b}} \frac{1}{g(t_0)F(t_0)(1-F(t_0))} (g_{a,b}(z) - g_{a,b}^0(z))^2\, dz\,.
\end{aligned}
$$

Set $V_n = T_n$. Since $[-K_\epsilon, K_\epsilon]$ contains $\tilde{D}_n$ with probability greater than $1-\epsilon$ eventually ($\tilde{D}_n$ is the left-closed, right-open interval over which the processes

X_n and Y_n differ) we have $P[U_{n\epsilon} \neq V_n] < \epsilon$ eventually. Similarly $P[U_\epsilon \neq V] < \epsilon$. Also $U_{n\epsilon} \to_d U_\epsilon$ as $n \to \infty$, for every fixed ϵ. This is so because by Theorem 8.1 with $\Psi = 0$,

$$(X_n(z), Y_n(z)) \to_d \frac{1}{g(t_0)} \left(g_{a,b}(z), g^0_{a,b}(z)\right)$$

as a process in $\mathcal{L}_2[-c, c] \times \mathcal{L}_2[-c, c]$ for every $c > 0$ (with a and b as defined in Theorem 8.1) and because

$$(f, g) \mapsto \int_{[-c,c]} (f(z) - g(z))^2 \, dz$$

is a continuous function from $\mathcal{L}_2[-c, c] \times \mathcal{L}_2[-c, c]$ to the reals. Thus all conditions of Lemma 8.1 are satisfied, leading to the conclusion that $V_n \to_d V$. We now establish that the limiting distribution is actually independent of the constants a and b, thereby showing universality. We shall show that

$$\begin{aligned} V &= \int_{D_{a,b}} \frac{1}{g(t_0)F(t_0)(1 - F(t_0))} \left(g_{a,b}(z) - g^0_{a,b}(z)\right)^2 dz \\ &\stackrel{d}{=} \int_{D_{1,1}} \left(g_{1,1}(z) - g^0_{1,1}(z)\right)^2 dz = \mathbb{T}\,. \end{aligned} \quad (8.2)$$

Now, by Lemma 8.2 with $\Psi = 0$ (and hence also $\phi = 0$) or by (6.9), Banerjee and Wellner (2001), page 1724, we have

$$\begin{aligned} &\left(g_{a,b}(t), g^0_{a,b}(t), D_{a,b}\right) \\ &\stackrel{d}{=} \left(a\,(b/a)^{1/3} g_{1,1}\left((b/a)^{2/3}t\right), a\,(b/a)^{1/3} g^0_{1,1}\left((b/a)^{2/3}t\right), (a/b)^{2/3} D_{1,1}\right) \end{aligned}$$

as processes in $\mathcal{L}_2[-c, c] \times \mathcal{L}_2[-c, c]$. Thus (8.2) follows by change of variables and straightforward calculation. □

Proposition 8.1 *For each $\epsilon > 0$ we can find an interval $[-K_\epsilon, K_\epsilon]$ such that*

$$\liminf_{n\to\infty} P(\tilde{D}_n \subset [-K_\epsilon, K_\epsilon]) > 1 - \epsilon \quad \textit{and} \quad P(D_{a,b} \subset [-K_\epsilon, K_\epsilon]) > 1 - \epsilon\,.$$

Proof. Let $\tilde{D}_n = [\tilde{A}_n, \tilde{B}_n)$. Note that $\tilde{D}_n \subset [-K, K]$ if and only if $-K \leq \tilde{A}_n \leq B_n \leq K$. Thus it suffices to show that

$$\lim_{K\to\infty} \limsup_{n\to\infty} P(\tilde{B}_n > K) = 0 \qquad \text{and} \qquad \lim_{K\to\infty} \limsup_{n\to\infty} P(\tilde{A}_n < -K) = 0\,.$$

We prove the first of these; the second is argued similarly. Note that $\tilde{B}_n > K$ implies that $t_0 + K n^{-1/3}$ is in the difference set D_n, and this implies that either $\mathbb{F}_n^0(t_0 + K n^{-1/3}) = \theta_0$ or $\mathbb{F}_n(t_0 + K n^{-1/3}) = \mathbb{F}_n(t_0)$ from the definitions of the constrained and unconstrained estimators. In terms of the processes X_n and Y_n this can be rewritten as

$$\{\tilde{B}_n > K\} \subset \{X_n(0) = X_n(K)\} \cup \{Y_n(K) = 0\}\,.$$

By Theorem 4.1 with $\phi \equiv 0$, and with $D_R(aW(h) + bh^2)$ denoting the right derivative of the greatest convex minorant of the process $aW(h) + bh^2$ on $[0, \infty)$,

$$\begin{aligned} P(Y_n(K) = 0) &\rightarrow P(g_{1,1}^0(K)/g(t_0) = 0) \\ &= P(D_R(aW(h) + bh^2)(K)/g(t_0) \le 0)\,. \end{aligned}$$

By choosing K large enough we can make the probability on the right side of the last display less than (say) $\epsilon/4$. Similarly, since $(X_n(0), X_n(K)) \rightarrow_d (g_{a,b}(0), g_{a,b}(K))/g(t_0)$,

$$\limsup_{n\to\infty} P(X_n(0) - X_n(K) = 0) \le P(g_{a,b}(0) - g_{a,b}(K) = 0)\,,$$

and the right side can be made arbitrarily small by choosing K large.

The argument for $D_{a,b}$ is similar (and is contained in the above proof for $\tilde{D}_n$). □

Proof of Theorem 4.1. The asymptotic behavior of the likelihood ratio statistic under contiguous alternatives is given in Theorem 2.8 of Banerjee and Wellner (2001).

The asymptotic distribution of the Wald statistic under contiguous alternatives follows from Theorem 4.1. By Theorem 4.1

$$n^{1/3}\left(\mathbb{F}_n(t_0 + z\,n^{-1/3}) - F(t_0)\right) \rightarrow_d g_{a,b,\Psi}(z)/g(t_0)\,,$$

under the sequence of contiguous alternatives $\{P^n_{F_n,G}\}$. Setting $z = 0$ yields

$$n^{1/3}\left(\mathbb{F}_n(t_0) - F(t_0)\right) \rightarrow_d g_{a,b,\Psi}(0)/g(t_0)\,.$$

By Lemma 8.2 we know that

$$\begin{aligned} \frac{1}{g(t_0)}\, g_{a,b,\Psi}(0) &\overset{d}{=} \frac{1}{g(t_0)}\, a\,(b/a)^{1/3}\, g_{1,1,\phi}(0) \\ &= \left(\frac{F(t_0)\,(1 - F(t_0))\,f(t_0)}{2\,g(t_0)}\right)^{1/3} g_{1,1,\phi}(0)\,, \end{aligned}$$

after plugging in the values of a and b. It follows that

$$\frac{n^{1/3}\,(\mathbb{F}_n(t_0) - F(t_0))}{(F(t_0)\,(1 - F(t_0))\,f(t_0)/2\,g(t_0))^{1/3}} \rightarrow_d g_{1,1,\phi}(0)\,.$$

Now, since $\mathbb{F}_n(t_0)\,(1 - \mathbb{F}_n(t_0))\,\hat{f}_n(t_0)/2\,\hat{g}_n(t_0)$ is a consistent estimator of $F(t_0)\,(1 - F(t_0))\,f(t_0)/2\,g(t_0)$ under the null hypothesis, it continues to stay consistent under any sequence of contiguous alternatives, and it follows from Slutsky's theorem that

$$\frac{n^{1/3}\,(\mathbb{F}_n(t_0) - F(t_0))}{(\mathbb{F}_n(t_0)(1 - \mathbb{F}_n(t_0))\,\hat{f}_n(t_0)/2\,\hat{g}_n(t_0))^{1/3}} \rightarrow_d g_{1,1,\phi}(0)\,.$$

The desired distributional convergence now follows as under the null by the continuous mapping theorem.

It remains to find the limiting distribution of the score statistic. Since $T_{n,1}$, $T_{n,2}$, and $T_{n,3}$ are all asymptotically equivalent to the sequence T_n defined in (8.1) in the proof of Theorem 3.1 under the null hypothesis, they are also asymptotically equivalent to T_n under a sequence of contiguous alternatives; it therefore suffices to find the asymptotic distribution of T_n under the given sequence. Recall, from the proof of Theorem 3.1, that

$$T_n = \frac{g(t_0)}{F(t_0)\,(1 - F(t_0))} \int_{\tilde{D}_n} (X_n(z) - Y_n(z))^2\, dz\,.$$

By Theorem 8.1 and Lemma 8.2 we have

$$\begin{aligned} &(X_n(z), Y_n(z)) \\ &\quad\rightarrow_d \frac{1}{g(t_0)} \left(g_{a,b,\Psi}(z), g^0_{a,b,\Psi}(z)\right) \qquad (8.3)\\ &\quad\stackrel{d}{=} \frac{1}{g(t_0)} \left(a\,(b/a)^{1/3} g_{1,1,\phi}\left((b/a)^{2/3} t\right), a\,(b/a)^{1/3} g^0_{1,1,\phi}\left((b/a)^{2/3} t\right)\right)\,. \end{aligned}$$

This convergence, coupled with the fact that $\tilde{D}_n$, the set on which X_n and Y_n differ, is contained in a compact set for sufficiently large n with arbitrarily high probability, and the fact that $D_{a,b,\Psi}$ is also contained in a compact set with arbitrarily high probability, yields

$$T_n \rightarrow_d \frac{1}{g(t_0)F(t_0)(1 - F(t_0))} \int_{D_{a,b,\Psi}} \left(g_{a,b,\Psi}(z) - g^0_{a,b,\Psi}(z)\right)^2 dz\,, \qquad (8.4)$$

through an application of Lemma 8.1 as in the proof of Theorem 3.1. That the right side in (8.4) has the same distribution as the random variable on the right side of (4.5) in the statement of the theorem follows from Lemma 7.2. □

Proof of Theorem 5.1. Part (i) was proved in Banerjee and Wellner (2001) where it was also shown that $\mathbb{F}_n^0(t) \to_p H(t)$ on (c, d). Parts (ii) and (iii) follow from this together with consistency of $\mathbb{F}_n$. The latter consistency also yields part (iv) via Slutsky's theorem. Part (v) follows from the following inequalities: for $0 < x \le y \le 1/2$, and for $1/2 \le y \le x < 1$,

$$\frac{(x-y)^2}{y(1-y)} \le 2\left\{x \log \frac{x}{y} + (1-x)\log\frac{1-x}{1-y}\right\} \le \frac{(x-y)^2}{x(1-x)} .$$

For $0 < y \le x \le 1/2$, and for $1/2 \le x \le y < 1$,

$$\frac{(x-y)^2}{x(1-x)} \le 2\left\{x \log \frac{x}{y} + (1-x)\log\frac{1-x}{1-y}\right\} \le \frac{(x-y)^2}{y(1-y)} .$$

□

References

Banerjee, M. (2000). *Likelihood Ratio Inference in Regular and Nonregular Problems.* Ph.D. dissertation, University of Washington. `http://www.stat.lsa.umich.edu/~ moulib/mythesis.ps`

Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699 - 1731.

Banerjee, M. and Wellner, J. A. (2003). Score statistics for current status data: comparisons with likelihood ratio and Wald statistics. Revision of *Technical Report* No. **376**, University of Michigan, Department of Statistics.

Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.*, to appear. `http://www.stat.lsa.umich.edu/~ moulib/CI-sendoff.pdf`

Dudley, R. M. (1989). *Real Analysis and Probability*. Chapman and Hall, New York.

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. Wadsworth, New York 529-555.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields* **81**, 79 - 109.

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

Groeneboom, P. and Wellner J. A. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics* **10**, 388-400.

Keiding, N., Begtrup, K., Scheike, T.H. and Hasibeder, G. (1996). Estimation from current status data in continuous time. *Lifetime Data Analysis* **2**, 119-129.

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

Wellner, J. A. (2003). Gaussian white noise models: some results for monotone functions. *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, IMS Lecture Notes-Monograph Series **43**, 87-104.