Likelihood Ratio Tests Under Local and Fixed Alternatives in Monotone Function Problems.

Moulinath Banerjee

University of Michigan

January 6, 2005

Running head: Likelihood ratios under alternatives.

Abstract

We focus on a class of nonstandard problems involving nonparametric estimation of a monotone function that is characterized by $n^{1/3}$ rate of convergence of the MLE, non-Gaussian limit distributions and the non–existence of $\sqrt{n}$–regular estimators. We have shown elsewhere that under a null hypothesis of the type $\psi(z_0) = \theta_0$ ($\psi$ being the monotone function of interest) in non-standard problems of the above kind, the likelihood ratio statistic has a “universal” limit distribution that is free of the underlying parameters in the model. In this paper, we illustrate its limiting behavior under local alternatives of the form $\psi_n(z)$, where $\psi_n(\cdot)$ and $\psi(\cdot)$ vary in $O(n^{-1/3})$ neighborhoods around $t_0$ and $\psi_n$ converges to $\psi$ at rate $n^{1/3}$ in an appropriate metric. Apart from local alternatives, we also consider the behavior of the likelihood ratio statistic under fixed alternatives and establish the convergence in probability of an appropriately scaled version of the same to a constant involving a Kullback-Leibler distance.

Key Words: asymptotic distribution, contiguity, fixed alternatives, greatest convex minorants, likelihood ratio statistic, local alternatives, monotone hazard model, monotone regression model, non-regular problems.
1 Introduction and Background

We consider the problem of estimating a monotone function at one or more points in a class of nonstandard problems, which can be generically described in the following way: Let $W_1, W_2, \ldots, W_n$ be i.i.d. observations from the density $f(w, \psi, \xi)$ (the corresponding distribution function is denoted by $P_{\psi, \xi}$). Here $f$ is a density with respect to some underlying measure, $\psi$ is a monotone function of interest and $\xi$ is a nuisance parameter. Denote by $\hat{\psi}_n$, the MLE of $\psi$ based on $W_1, W_2, \ldots, W_n$. The fundamental feature of this class of problems that sets it apart from the spectrum of regular parametric and semiparametric problems is the $n^{1/3}$ rate of convergence of $\hat{\psi}_n(t)$ to $\psi(t)$; in each case

$$n^{1/3}(\hat{\psi}_n(t) - \psi(t) \to_d C(\psi, \xi, t)Z,$$

where the random variable $Z$ is a symmetric (about 0) but non-Gaussian random variable and $C(\psi, \xi, t)$ is a constant depending upon the underlying parameters in the problem and the point of interest $t$. In fact, $Z = \arg \min h W(h) + h^2$, where $W(h)$ is standard two-sided Brownian motion on the line. Models of the above kind abound in the nonparametric literature. For example, the monotone function of interest can be a distribution function (survival analysis) which is the case, for example, with current status data studied in Groeneboom & Wellner (1992)), (or more generally Case $k$ interval censored data – Groeneboom (1996)), a monotone hazard function based on uncensored or right censored data, as in Prakasa Rao (1970) and Huang & Wellner (1995)), a monotone regression function as in Brunk (1970), the mean function of a counting process as Sun & Kalbfleisch (1995), Wellner & Zhang (2000), or a monotone density as in Prakasa Rao (1969), Huang & Zhang (1994) and Huang & Wellner (1995) .

The behavior of the likelihood ratio statistic (LRS) for testing a null hypothesis of the type $H_0 : \psi(z_0) = \theta_0$ (or more generally null hypotheses that constrain the monotone function of interest at finitely many points) in these models was first studied by Banerjee & Wellner (2001) in the context of the current status data model. Banerjee (2004) subsequently developed a unified theory of likelihood based estimation for monotone function models and showed that in a very
broad class of such problems, the LRS 2 \( \log \lambda_n \), where

\[
\lambda_n = \log \frac{\prod_{i=1}^{n} f(W_i, \hat{\psi}_n, \xi)}{\prod_{i=1}^{n} f(W_i, \psi_n^0, \xi)},
\]

( \( \hat{\psi}_n \) is the unconstrained MLE and \( \psi_n^0 \) is the MLE under \( H_0 \)) converges in distribution under \( H_0 \) to

\[
\mathbb{D} \equiv \int \left( (g_{1,1}(h))^2 - (g_{1,1}^0(h))^2 \right) dh,
\]

and the processes \( g_{1,1}(h) \) and \( g_{1,1}^0(h) \) are described as follows.

For positive constants \( a \) and \( b \) define the process \( X_{a,b}(h) := aW(h) + bh^2 \), where \( W(h) \) is standard two-sided Brownian motion starting from 0. Let \( G_{a,b}(h) \) denote the GCM (greatest convex minorant) of \( X_{a,b}(h) \) and \( g_{a,b}(h) \) denote the right derivative of \( G_{a,b} \). It can be shown that \( g_{a,b} \) is a piecewise constant increasing function with finitely many jumps in any compact interval. For \( h \leq 0 \), let \( G_{a,b,L}(h) \) denote the GCM of \( X_{a,b}(h) \) on the set \( h \leq 0 \) and \( g_{a,b,L}(h) \) denote its right-derivative process. For \( h > 0 \), let \( G_{a,b,R}(h) \) denote the GCM of \( X_{a,b}(h) \) on the set \( h > 0 \) and \( g_{a,b,R}(h) \) denote its right-derivative process. Define \( g_{a,b}^0(h) \) as \( g_{a,b,L}(h) \wedge 0 \) for \( h \leq 0 \) and as \( g_{a,b,R}(h) \vee 0 \) for \( h > 0 \). Then, \( g_{a,b}^0(h) \), like \( g_{a,b}(h) \), is a piecewise constant increasing function with finitely many jumps in any compact interval and differing (almost surely) from \( g_{a,b}(h) \) on a finite interval containing 0. In fact, with probability 1, \( g_{a,b}^0(h) \) is identically 0 in some (random) neighborhood of 0, whereas \( g_{a,b}(h) \) is almost surely non-zero in some (random) neighborhood of 0. Also, the interval \( D_{a,b} \) on which \( g_{a,b} \) and \( g_{a,b}^0 \) differ is \( O_p(1) \). For more detailed descriptions of the processes \( g_{a,b} \) and \( g_{a,b}^0 \), see Groeneboom (1989), Banerjee (2000), Banerjee & Wellner (2001) and Wellner (2003). Thus, \( g_{1,1} \) and \( g_{1,1}^0 \) are the unconstrained and constrained versions of the slope processes associated with the canonical process \( X_{1,1}(z) \).

Our goal in this paper is to study the behavior of \( 2 \log \lambda_n \) under alternative hypotheses, both local and fixed, in monotone function models. The results will be established in the general framework developed by Banerjee (2004) which we present below.
The Unified Framework: Let \( \{p(x, \theta) : \theta \in \Theta \} \) with \( \Theta \) being an open subset of \( \mathbb{R} \), be a one-parameter family of probability densities with respect to a dominating measure \( \mu \) and satisfying the assumptions A.1 – A.7 given below. Let \( \psi \) be an increasing or decreasing continuous function defined on an interval \( \tilde{I} \) and taking values in \( \Theta \). Consider i.i.d. data \( \{(X_i, Z_i)\}_{i=1}^n \) where \( Z_i \sim p_Z \), \( p_Z \) being a Lebesgue density defined on \( \tilde{I} \) and \( X_i \mid Z_i = z \sim p(x, \psi(z)) \). Here \( p_Z \) plays the role of the nuisance parameter \( \xi \) referred to at the beginning of this section while \( \psi \) is the parameter of interest. In what follows, it will be seen that the computation of the maximum likelihood estimates of \( \psi \) and the likelihood ratio statistic for hypothesis tests involving \( \psi \) do not require estimation of the nuisance parameter \( \xi \).

Assume that for a fixed interior point of \( I \), say \( z_0 \),

(a) \( p_Z \) is positive and continuous in a neighborhood of \( z_0 \),

(b) \( \psi \) is continuously differentiable in a neighborhood of \( z_0 \) with \( |\psi'(z_0)| > 0 \).

Denote by \( \hat{\psi}_n \) the unconstrained MLE of \( \psi \) and by \( \hat{\psi}_n^0 \) the MLE of \( \psi \) under the constraint imposed by the pointwise null hypothesis \( H_0 : \psi(z_0) = \theta_0 \). Consider the likelihood ratio statistic (LRS) for testing the hypothesis \( H_0 : \psi(z_0) = \theta_0 \), where \( \theta_0 \) is an interior point of \( \Theta \). Denoting the LRS by \( 2 \log \lambda_n \), we have

\[
2 \log \lambda_n = 2 \log \frac{\Pi_{i=1}^n p(X_i, \hat{\psi}_n(Z_i))}{\Pi_{i=1}^n p(X_i, \hat{\psi}_n^0(Z_i))}.
\] (1)

Assumptions: We state below our assumptions about the parametric model \( p(x, \theta) \).

(A.1) The set \( \mathcal{X}_\theta : \{x : p(x, \theta) > 0\} \) does not depend on \( \theta \) and is denoted by \( \mathcal{X} \).

(A.2) \( l(x, \theta) = \log p(x, \theta) \) is at least three times differentiable with respect to \( \theta \) and is strictly concave in \( \theta \) for every fixed \( x \). We write: \( \dot{l}(x, \theta) = \partial l(x, \theta)/\partial \theta \) and \( \ddot{l}(x, \theta) = \partial^2 l(x, \theta)/\partial \theta^2 \).

(A.3) If \( T \) is any statistic such that \( E_\theta(|T|) < \infty \), then:

\[
\frac{\partial}{\partial \theta} \int_{\mathcal{X}} T(x) p(x, \theta) \, dx = \int_{\mathcal{X}} T(x) \frac{\partial}{\partial \theta} p(x, \theta) \, dx
\]
and
\[ \frac{\partial^2}{\partial \theta^2} \int_{\mathcal{X}} T(x) p(x, \theta) \, dx = \int_{\mathcal{X}} T(x) \frac{\partial^2}{\partial \theta^2} p(x, \theta) \, dx. \]

Under these assumptions, \( I(\theta) \equiv E_{\theta}(\dot{l}(X, \theta)^2) = -E_{\theta}(\ddot{l}(X, \theta)) \).

(A.4) \( I(\theta) \) is finite and continuous at \( \theta_0 \).

(A.5) There exists a neighborhood \( \mathcal{N} \) of \( \theta_0 \) such that for all \( x \), \( \sup_{\theta \in \mathcal{N}} |l'''(x, \theta)| \leq B(x) \), with \( B \) satisfying \( \sup_{\theta \in \mathcal{N}} E_{\theta}(B(X)) < \infty \).

(A.6) The functions:
\[ f_1(\theta_1, \theta_2) = E_{\theta_1}(\dot{l}(X, \theta_2)^2) \quad \text{and} \quad f_2(\theta_1, \theta_2) = E_{\theta_1}(\ddot{l}(X, \theta_2)) \]
are continuous in a neighborhood of \((\theta_0, \theta_0)\). Also, the function \( f_3(\theta) = E_{\theta}(\ddot{l}(X, \theta)^2) \) is uniformly bounded in a neighborhood of \( \theta_0 \).

(A.7) Let \( H(\theta, M) \) be defined as:
\[ H(\theta, M) = E_{\theta} \left[ \left( |\dot{l}(X, \theta)|^2 + \ddot{l}(X, \theta)^2 \right) \left( 1 \{ |\dot{l}(X, \theta)| > M \} + 1 \{ |\ddot{l}(X, \theta)| > M \} \right) \right]. \]

Then, \( \lim_{M \to \infty} \sup_{\theta \in \mathcal{N}} H(\theta, M) = 0 \).

Note, in particular, that assumption (A.7) is easily satisfied if \( \dot{l}(x, \theta) \) and \( \ddot{l}(x, \theta) \) are uniformly bounded for \( x \in \mathcal{X} \) and \( \theta \in \mathcal{N} \). It also holds fairly easily for one–parameter exponential families.

Finally, we assume that with probability increasing to 1 as \( n \to \infty \), the MLE's \( \hat{\psi}_n \) and \( \hat{\psi}^0_n \) exist.

Theorem 2.2 of Banerjee (2004) shows that under Assumptions (a), (b) and (A.1)–(A.7), the likelihood ratio statistic in (1) converges in distribution under \( H_0 \) to \( \mathbb{D} \).

The above conditionally parametric formulation captures many well studied models in the literature. For example, the standard monotone regression model where \( X_i = \psi(Z_i) + \epsilon_i \),
\( \{\epsilon_i, Z_i\}_{i=1}^n \) are i.i.d. random variables, the \( \epsilon_i \)'s have a common distribution with mean 0 and a finite variance, the \( Z_i \)'s have a common Lebesgue density \( p_Z \) and \( \psi \) is a monotone function. If the common error density is \( N(0, \sigma^2) \), conditions (A.1)–(A.7) can be easily verified and the conclusion of Theorem 2.2 in Banerjee (2004) is valid. The same conclusion also holds for more general error distributions. Other examples include the current status model discussed in Banerjee & Wellner (2001) and discrete response regression models like binary regression models and Poisson regression models under monotonicity constraints on the corresponding regression functions. For a more detailed discussion, see Section 3 and the beginning of Section 2 in Banerjee (2004).

The rest of the paper is organized as follows. In Section 2 we study the behavior of \( 2 \log \lambda_n \) under a sequence of local alternatives. Thus, at stage \( n \), \( (X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n) \) are generated from the density \( g_n(x, z) = p(x, \psi_n(z)) p_Z(z) \) where \( \psi_n \) is a sequence of monotone functions converging to \( \psi_0 \) (\( \psi_0 \) being a monotone function satisfying \( H_0 : \psi(z_0) = \theta_0 \)) in such a way that the sequence of probability measures \( \{P_{\psi_n}^n\} \) (the joint distribution of \( (X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n) \) under \( \psi_n \)) is contiguous to \( \{P_{\psi_0}^n\} \) (the joint distribution of \( (X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n) \) under \( \psi_0 \)). Contiguity is deduced in each case via a LAN (local asymptotic normality) expansion of the local log–likelihood ratio and is then invoked to deduce the asymptotic distribution of \( 2 \log \lambda_n \) under \( \{P_{\psi_n}^n\} \). The sequence of monotone functions \( \{\psi_n\} \) used to define the local alternatives are of the form: \( \psi_n(z) = \psi_0(z) + n^{-1/3} A_n(z) \), where \( A_n(z) = B_n(n^{1/3}(z - z_0)) \). Here \( \{B_n(z)\} \) is a sequence of continuously differentiable functions vanishing outside of \( (-c, c) \) and converging uniformly to a continuously differentiable function \( B(z) \). A sufficient condition for \( \psi_n \) to be monotone increasing (decreasing) is that \( B_n \) is increasing (decreasing). Indeed, the optimal rate of convergence of local alternatives in monotone function problems is \( n^{1/3} \), matching the rate of convergence of the MLE. If we consider alternatives converging at \( \sqrt{n} \) rate in these problems, it can be shown that the asymptotic distribution of the likelihood ratio statistic under these alternatives is the same as that under the null. In other words, \( \sqrt{n} \) alternatives converge too quickly in these models for the likelihood ratio to detect deviations from the null hypothesis. On the other hand, under local alternatives converging at rate slower than \( n^{1/3} \), the power of the LRS will converge to 1.
Section 3 deals with the behavior of the LRS under a fixed alternative hypothesis.

We end this section with some terminology that will be required later on. First define \( \mathcal{L} \) to be the space of locally square integrable real–valued functions on \( \mathbb{R} \) equipped with the topology of \( L_2 \) convergence on compact sets. Thus \( \mathcal{L} \) comprises all functions \( \phi \) that are square integrable on every compact set and \( \phi_n \) is said to converge to \( \phi \) if \( \int_{[-K,K]} (\phi_n(t) - \phi(t))^2 \, dt \to 0 \) for every \( K \).

The space \( \mathcal{L} \times \mathcal{L} \) denotes the cartesian product of two copies of \( \mathcal{L} \) with the usual product topology. Also define \( B_{\text{loc}}(\mathbb{R}) \) to be the set of all real–valued functions defined on \( \mathbb{R} \) that are bounded on every compact set, equipped with the topology of uniform convergence on compacta. Thus \( h_n \) converges to \( h \) in \( B_{\text{loc}}(\mathbb{R}) \) if \( h_n \) and \( h \) are bounded on every compact interval \( [-K,K] \) \((K > 0)\) and \( \sup_{x \in [-K,K]} |h_n(x) - h(x)| \to 0 \) for every \( K > 0 \).

2 Behavior of likelihood ratio statistics under local alternatives

In what follows we deal (without loss of generality) with monotone increasing functions in the set–up of the previous section.

The likelihood function corresponding to \( P_{\psi_0}^n \), the joint distribution of \((X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n)\) with respect to an appropriate dominating measure, is given by:

\[
L_n(\psi_0) = \prod_{i=1}^n p(X_i, \psi_0(Z_i)) p_Z(Z_i),
\]

whereas the likelihood function corresponding to \( P_{\psi_n}^n \) is

\[
L_n(\psi_n) = \prod_{i=1}^n p(X_i, \psi_n(Z_i)) p_Z(Z_i).
\]

We have the following theorem.

**Theorem 1** *The sequences of probability measures \( \{P_{\psi_n}^n\} \) and \( \{P_{\psi_0}^n\} \) are mutually contiguous.*
Proof: In the following derivation all expectations are computed with respect to the distribution under \( \psi_0 \) which satisfies the null hypothesis. We have,

\[
\log L_n(\psi_n) - \log L_n(\psi_0) = \log \Pi_{i=1}^n \frac{p(X_i, \psi_n(Z_i))}{p(X_i, \psi_0(Z_i))} \\
= \sum_{i=1}^n l(X_i, \psi_n(Z_i)) - \sum_{i=1}^n l(X_i, \psi_0(Z_i)) \\
= \sum_{i=1}^n \hat{l}(X_i, \psi_0(Z_i)) (\psi_n(Z_i) - \psi_0(Z_i)) \\
+ \sum_{i=1}^n \frac{1}{2} \hat{l}(X_i, \psi_0(Z_i)) (\psi_n(Z_i) - \psi_0(Z_i))^2 \\
+ \sum_{i=1}^n \frac{1}{6} \hat{l}''(X_i, \psi_n^*(Z_i)) (\psi_n(Z_i) - \psi_0(Z_i))^3 \\
\equiv I_n + II_n + III_n,
\]

where \( \psi_n^*(Z_i) \) is a point intermediate between \( \psi_n(Z_i) \) and \( \psi_0(Z_i) \). Now, recall that,

\[
\psi_n(Z_i) - \psi_0(Z_i) = n^{-1/3} A_n(Z_i) \equiv n^{-1/3} B_n(n^{1/3}(Z_i - z_0))
\]

which implies that, \( I_n = \sum_{i=1}^n \hat{l}(X_i, \psi_0(Z_i)) n^{-1/3} A_n(Z_i) = n \mathbb{P}_n \left[ \hat{l}(X, \psi_0(Z)) n^{-1/3} A_n(Z) \right] = n \left( \mathbb{P}_n - P \right) \left[ \hat{l}(X, \psi_0(Z)) n^{-1/3} A_n(Z) \right] \), since \( P \left[ \hat{l}(X, \psi_0(Z)) n^{-1/3} A_n(Z) \right] = 0 \) by virtue of the fact that \( E(\hat{l}(X, \psi_0(Z)) \mid Z = z) = E(\hat{l}(X, \psi_0(z)) \mid Z = z) = 0 \). Hence, \( I_n = \sqrt{n} (\mathbb{P}_n - P) s_n \), where \( s_n = n^{1/6} \hat{l}(X, \psi_0(Z)) A_n(Z) \).

Using the Lindeberg CLT for triangular arrays of random variables (the details will be provided later in a more general context), it follows that \( \sqrt{n}(\mathbb{P}_n - P) s_n \to_d Z_\sigma \sim N(0, \sigma^2) \) where \( \sigma^2 = \lim_{n \to \infty} E(s_n^2) \). We compute \( \sigma^2 \) explicitly below:

\[
\sigma^2 = \lim_{n \to \infty} E(s_n^2) \\
= \lim_{n \to \infty} E \left[ n^{1/3} \hat{l}^2(X, \psi_0(Z)) A_n^2(Z) \right] \\
= \lim_{n \to \infty} n^{1/3} \int_{z_0 - c_n^{-1/3}}^{z_0 + c_n^{-1/3}} \hat{l}(\psi_0(z)) A_n^2(z) p_Z(z) \, dz \\
= \lim_{n \to \infty} \int_{-\infty}^{\infty} \hat{l}(\psi_0(z_0 + h n^{-1/3})) A_n^2(z_0 + h n^{-1/3}) p_Z(z_0 + h n^{-1/3}) \, dh
\]
\[
\begin{align*}
= \lim_{n \to \infty} \int_{-c}^{c} I(\psi_0(z_0 + h n^{-1/3})) B_n^2(h) p_Z(z_0 + h n^{-1/3}) \, dh \\
= I(\psi_0(z_0)) p_Z(z_0) \int_{-c}^{c} B^2(h) \, dh \\
= \frac{1}{a^2} \int_{-c}^{c} B^2(h) \, dh ,
\end{align*}
\]

with \(a^2 = (I(\psi_0(z_0)) p_Z(z_0))^{-1}\). Next, we consider the asymptotic behavior of II\(n\). We have

\[
II_n = \frac{1}{2} \sum_{i=1}^{n} \ddot{I}(X_i, \psi_0(Z_i)) \left( \psi_n(Z_i) - \psi_0(Z_i) \right)^2 = \frac{n}{2} \mathbb{P}_n \left[ \ddot{I}(X, \psi_0(Z)) \left( \psi_n(Z) - \psi_0(Z) \right) \right] \equiv \frac{1}{2} \mathbb{P}_n \xi_n(X, Z)
\]

where \(\xi_n(x, z) = n^{1/3} \ddot{I}(x, \psi_0(z)) A_n^2(z)\). Now,

\[
E(\mathbb{P}_n \xi_n(X, Z)) = n^{1/3} E \left[ \ddot{I}(X, \psi_0(Z)) A_n^2(Z) \right] = -n^{1/3} \int_{z_0-c n^{-1/3}}^{z_0+c n^{-1/3}} I(\psi_0(z)) A_n^2(z) p_Z(z) \, dz = - \int_{-c}^{c} I(\psi_0(z_0 + h n^{-1/3})) B_n^2(h) p_Z(z_0 + h n^{-1/3}) \, dh = -I(\psi_0(z_0)) p_Z(z_0) \int_{-c}^{c} B^2(h) \, dh + o(1) \equiv -\sigma^2 + o(1) .
\]

On the other hand,

\[
\text{Var}(\mathbb{P}_n \xi_n(X, Z)) = \frac{\text{Var}(\xi_n(X, Z))}{n} \leq \frac{1}{n} E \left[ n^{2/3} \ddot{I}(X, \psi(Z))^2 A_n^4(Z) \right] = n^{-2/3} E \left[ n^{1/3} \ddot{I}(X, \psi(Z))^2 A_n^4(Z) \right] = n^{-2/3} \int_{-c}^{c} f_3(\psi_0(z_0 + h n^{-1/3})) B_n^4(h) \, dh ,
\]

where \(f_3\) is defined as in (A.6). It follows from (A.6) and the uniform convergence of \(B_n\) to \(B\) that the integral in the last step of the above display is \(O(1)\), showing that \(\text{Var}(\mathbb{P}_n \xi_n(X, Z)) = O(n^{-2/3})\) and hence \(o(1)\). By Chebyshev’s inequality, it follows that \(\mathbb{P}_n(\xi_n(X, Z)) \to_p -\sigma^2\), whence \(II_n \to_p -\sigma^2/2\).

Finally, consider the term III\(n\). Using Assumption A.5 and the facts that (a) \(\psi_n^*(Z_i)\) lies between \(\psi_n(Z_i)\) and \(\psi_0(Z_i)\) and (b) \(\sup_{z \in [z_0-c n^{-1/3}, z_0+c n^{-1/3}]} | \psi_n(z) - \psi_0(z) | \to 0\), it follows easily that III\(n\) is eventually bounded in absolute value by \(\mathbb{P}_n (B(X) | A_n(Z) |^3) / 6\), the mean and
variance of which both converge to 0 with increasing \( n \). Hence \( III_n \) converges in probability to 0. Using 2 in conjunction with Slutsky’s theorem leads to the conclusion that

\[
\log L_n(\psi_n) - \log L_n(\psi_0) \rightarrow_d Z_\sigma - \sigma^2/2 \equiv W \sim N(-\sigma^2/2, \sigma^2),
\]

under the sequence of probability measures \( \{P_n^{\psi_0}\} \). Consequently,

\[
\frac{d P_n^{\psi_n}}{d P_n^{\psi_0}} = \frac{L_n(\psi_n)}{L_n(\psi_0)} \rightarrow_d e^W,
\]

with \( E(e^W) = 1 \). Thus, by a direct consequence of Le Cam’s first lemma (see Example 6.5 on page 89 of Van der Vaart (1998)), it follows that the sequence of probability measures \( \{P_n^{\psi_n}\} \) and \( \{P_n^{\psi_0}\} \) are mutually contiguous. \( \square \)

As in Banerjee (2004), we denote the unconstrained MLE of \( \psi \) by \( \hat{\psi}_n \) and the constrained MLE of \( \psi \) by \( \hat{\psi}_n^0 \). We now briefly review how the unconstrained and constrained MLE’s of \( \psi \) are characterized. First, we introduce some notation. For points \( \{(x_0, y_0), (x_1, y_1), \ldots, (x_k, y_k)\} \) where \( x_0 = y_0 = 0 \) and \( x_0 < x_1 < \ldots < x_k \), consider the left-continuous function \( P(x) \) such that \( P(x_i) = y_i \) and such that \( P(x) \) is constant on \( (x_{i-1}, x_i) \). We will denote the vector of slopes (left–derivatives) of the GCM of \( P(x) \) computed at the points \( (x_1, x_2, \ldots, x_k) \) by \( \text{slogcm} \{(x_i, y_i)\} \).

**Characterizing \( \hat{\psi}_n \):** In what follows, we define:

\[
\phi(x, \theta) \equiv -l(x, \theta) \equiv -\log p(x, \theta), \quad \phi'(x, \theta) = -\dot{l}(x, \theta) \quad \text{and} \quad \phi''(x, \theta) = -\ddot{l}(x, \theta).
\]

The log–likelihood function for the data is given by:

\[
l_n((X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n), \psi) = \sum_{i=1}^{n} l(X_i, \psi(Z_i)).
\]

The goal is to maximize this expression over all increasing functions \( \psi \). Let \( Z_{(1)} < Z_{(2)} < \ldots < Z_{(n)} \) denote the ordered values of \( Z \) and \( X_{(i)} \) denote the observed value of \( X \) corresponding to \( Z_{(i)} \). Since \( \psi \) is increasing \( \psi(Z_{(1)}) \leq \psi(Z_{(2)}) \ldots \leq \psi(Z_{(n)}) \). Finding \( \hat{\psi}_n \) therefore reduces to minimizing

\[
\tilde{\psi}(u_1, u_2, \ldots, u_n) = \sum_{i=1}^{n} \phi(X_{(i)}, u_i),
\]
over all \( u_1 \leq u_2 \leq \ldots \leq u_n \). Once we obtain the (unique) minimizer \( \hat{u} \equiv (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n) \), the MLE \( \hat{\psi}_n \) is given by: 
\[
\hat{\psi}_n(Z_{(i)}) = \hat{u}_i \text{ for } i = 1, 2, \ldots, n.
\]
Let \( \nabla_i \hat{\psi}(\hat{u}) \) denote the \( i \)’th partial derivative of \( \hat{\psi} \), evaluated at the vector \( \hat{u} \) and let \( d_i \) denote the \( ii \)’th second partial derivative of \( \hat{\psi} \), evaluated at the point \( \hat{u} \).

The solution \( \hat{u} \) can be characterized elegantly by using the Kuhn–Tucker theorem from convex optimization theory. This characterization can then be exploited to show that \( \hat{u} \) satisfies the solution to a least squares regression problem under monotonicity constraints (for the details, see Banerjee (2004)). Thus, \( \hat{u} \) minimizes
\[
A(u_1, u_2, \ldots, u_n) = \sum_{i=1}^{n} \left[ u_i - \left( \hat{u}_i - \nabla_i \hat{\psi}(\hat{u}) d_i^{-1} \right) \right]^2 \ d_i
\]
subject to the constraints that \( u_1 \leq u_2 \leq \ldots \leq u_n \) and hence furnishes the isotonic regression of the function
\[
g(i) = \hat{u}_i - \nabla_i \hat{\psi}(\hat{u}) d_i^{-1}
\]
on the ordered set \( \{1, 2, \ldots, n\} \) with weight function \( d_i \). It is well known that the solution is
\[
(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n) = \text{slogcm} \left\{ \sum_{j=1}^{i} d_j, \sum_{j=1}^{i} g(j) d_j \right\}_{i=0}^{n}
\]
See, for example Theorem 1.2.1 of Robertson, Wright & Dykstra (1988). In terms of the function \( \phi \) the solution can be written as:
\[
(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n) \equiv \left[ \text{slogcm} \left\{ \sum_{j=1}^{i} \phi''(X_{(i)}, \hat{u}_j), \sum_{j=1}^{i} \hat{u}_j \phi''(X_{(i)}, \hat{u}_j) - \phi'(X_{(i)}, \hat{u}_j) \right\}_{i=0}^{n} \right]. \quad (3)
\]
Note that the above is an implicitly defined equation for the \( \hat{u}_i \)’s and solving it requires iterative techniques like the modified ICM algorithm due to Jongbloed (1998). For a more detailed description of these issues, see Banerjee (2004), Section 2.

Recall that \( \hat{\psi}_n(Z_{(i)}) = \hat{u}_i \); for a \( z \) that lies strictly between \( Z_{(i)} \) and \( Z_{(i+1)} \), we set \( \hat{\psi}_n(z) = \hat{\psi}_n(Z_{(i)}) \). The MLE \( \hat{\psi}_n \) thus defined is a piecewise constant right–continuous function.
Characterizing $\hat{\psi}_n^0$: Let $m$ be the number of $Z_i$'s that are less than or equal to $z_0$. Finding $\hat{\psi}_n^0$ amounts to minimizing $\psi(u) = \sum_{i=1}^n \phi(X(i), u_i)$ over all $u_1 \leq u_2 \ldots \leq u_m \leq \theta_0 \leq u_{m+1} \leq \ldots \leq u_n$. This can be reduced to solving two separate optimization problems. These are: (1) Minimize $\sum_{i=1}^n \phi'(X(i), u_i)$ over $u_1 \leq u_2 \leq \ldots \leq u_m \leq \theta_0$ and (2) Minimize $\sum_{i=m+1}^n \phi'(X(i), u_i)$ over $\theta_0 \leq u_{m+1} \leq u_{m+2} \leq \ldots \leq u_n$. As in the case of the unconstrained solution, the constrained solution has a slope of greatest convex minorant characterization. Denoting the constrained solution vector by $\hat{u}^0 \equiv (\hat{u}_1^0, \hat{u}_2^0, \ldots, \hat{u}_n^0)$, it can be shown that $\hat{u}^0$ minimizes

$$A(u_1, u_2, \ldots, u_n) = \sum_{i=1}^n \left[ u_i - \left( \hat{u}_i^0 - \nabla_i \hat{\psi}(\hat{u}^0) d_i \right) \right]^2 d_i$$

subject to the constraints that $u_1 \leq u_2 \leq \ldots \leq u_m \leq \theta_0 \leq u_{m+1} \leq \ldots \leq u_n$, where $d_i = \nabla_i \hat{\psi}(\hat{u}^0)$. It is not difficult to see that

$$(\hat{u}_1^0, \hat{u}_2^0, \ldots, \hat{u}_m^0) \equiv \left[ \operatorname{sgcm} \left\{ \sum_{j=1}^i \phi''(X(i), \hat{u}_1^0), \sum_{j=1}^i \phi''(X(i), \hat{u}_1^0) - \phi'(X(i), \hat{u}_1^0) \right\} \right] \land \theta_0,$$  

and

$$(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \ldots, \hat{u}_n^0) \equiv \left[ \operatorname{sgcm} \left\{ \sum_{j=m+1}^i \phi''(X(i), \hat{u}_1^0), \sum_{j=1}^i \phi''(X(i), \hat{u}_1^0) - \phi'(X(i), \hat{u}_1^0) \right\} \right] \lor \theta_0.$$  

The constrained MLE $\hat{\psi}_n^0$ is the piecewise constant right–continuous function satisfying $\hat{\psi}_n^0(Z(i)) = \hat{u}_i^0$ for $i = 1, 2, \ldots, n$, $\hat{\psi}_n^0(z_0) = \theta_0$ and having no jump points outside the set $\{Z(i)\}_{i=1}^n \cup \{z_0\}$.

Define local processes $X_n$ and $Y_n$ corresponding to the MLE's $\hat{\psi}_n$ and $\hat{\psi}_n^0$ respectively, as follows:

$$X_n(h) = n^{1/3} \left( \hat{\psi}_n(z_0 + h n^{-1/3}) - \psi_0(z_0) \right)$$  

and  

$$Y_n(h) = n^{1/3} \left( \hat{\psi}_n^0(z_0 + h n^{-1/3}) - \psi_0(z_0) \right).$$
We will study the limiting behavior of the above processes under the contiguous sequence \( \{ P_n^\psi \} \).

We first introduce the following quantities. Define the drift function \( D(h) \) as follows:

\[
D(h) = \int_0^{h \wedge c} B(u) \, du, \quad h > 0
\]

and

\[
D(h) = \int_{-c \vee h}^0 B(u) \, du, \quad h \leq 0.
\]

Let,

\[
a = (I(\psi_0(z_0)) p_z(z_0))^{-1/2} \quad \text{and} \quad b = \frac{\psi'_0(z_0)}{2}.
\]

Define the process \( X_{a,b,D}(h) \equiv a W(h) + b h^2 + D(h) \). Let \( g_{a,b,D}(h) \) be the slope (right–derivative) process of the GCM of \( X_{a,b,D}(h) \). Let \( g^0_{a,b,D}(h) \) be the (piecewise constant) process, which for \( h \leq 0 \) is the minimum of the slope (right–derivative) of the GCM of \( X_{a,b,D}(h) \) on the set \( \{ h \leq 0 \} \) and \( 0 \), and for \( h > 0 \) is the maximum of the slope (right–derivative) of the GCM of \( X_{a,b,D}(h) \) on the set \( \{ h > 0 \} \) and \( 0 \).

Theorem 2 (to be presented below) deals with the limit distribution of the processes \( (X_n(h), Y_n(h)) \) and Theorem 3 (also presented below) which uses Theorem 2 provides the limit distribution of the likelihood ratio statistic \( 2 \log \lambda_n \) for testing \( H_0 : \psi(z_0) = \theta_0 \) under the local alternatives \( \{ P_n^\psi \} \).

**Theorem 2** Under the sequence of local alternatives \( \{ P_n^\psi \} \), the processes \( (X_n(h), Y_n(h)) \) converge finite-dimensionally and also as elements of the space \( L \times L \) to the processes \( (g_{a,b,D}(z), g^0_{a,b,D}(z)) \).

**Proof of Theorem 2:** For a monotone function \( \Lambda \) taking values in \( \Theta \) define the following processes:

\[
W_n,\Lambda(r) = P_n \left[ \phi'(X, \Lambda(Z)) 1(Z \leq r) \right],
\]

\[
G_n,\Lambda(r) = P_n \left[ \phi''(X, \Lambda(Z)) 1(Z \leq r) \right],
\]

and

\[
B_n,\Lambda(r) = \int_0^r \Lambda(z) \, dG_{n,\Lambda}(z) - W_{n,\Lambda}(r).
\]

The unconstrained MLE \( \hat{\psi}_n \) is given by:

\[
\{ \hat{\psi}_n(Z(i)) \}_{i=1}^n = \text{slogcm} \{ G_{n,\hat{\psi}_n}(Z(i)), B_{n,\hat{\psi}_n}(Z(i)) \}_{i=0}^n;
\]
this is a direct consequence of (3). Consequently,

\[
\{ \hat{\psi}_n(Z_{(i)}) - \psi_0(z_0) \}_{i=1}^n = \text{slogcm} \{ G_{n,\hat{\psi}_n}(Z_{(i)}), B_{n,\hat{\psi}_n}(Z_{(i)}) - \psi_0(z_0) G_{n,\hat{\psi}_n}(Z_{(i)}) \}_{i=1}^n.
\]

Now, the functions \( B_{n,\hat{\psi}_n} \) and \( G_{n,\hat{\psi}_n} \) are piecewise constant right–continuous functions with possible jumps only at the \( Z_{(i)} \)'s. Consider the function \( G_{n,\hat{\psi}_n}^{-1} \) (recall that for a right–continuous increasing function \( H \) we define \( H^{-1}(t) = \inf \{ s : H(s) \geq t \} \)). Note that \( G_{n,\hat{\psi}_n}^{-1}(0) = -\infty \). Defining \( Z(0) = -\infty \), so that \( G_{n,\hat{\psi}_n}(Z(0)) = 0 \), we have:

\[
G_{n,\hat{\psi}_n}^{-1}(t) = Z(i) \text{ for } t \in (G_{n,\hat{\psi}_n}(Z(i-1)), G_{n,\hat{\psi}_n}(Z(i)]).
\]

Thus, the function \( G_{n,\hat{\psi}_n}^{-1} \) is a piecewise constant \textit{left–continuous} function; consequently, so is the function \( (B_{n,\hat{\psi}_n} - G_{n,\hat{\psi}_n}) \cdot G_{n,\hat{\psi}_n}^{-1} \), with jumps at the points \( \{ G_{n,\hat{\psi}_n}(Z(i)) \}_{i=1}^n \). Denote by \( \text{slogcm} f \) the slope (left–derivative) of the GCM of a function \( f \) on its domain. From the characterization of \( \hat{\psi}_n \), we have:

\[
\hat{\psi}_n(z) - \psi_0(z_0) = \text{slogcm} \left( (B_{n,\hat{\psi}_n} - \psi_0(z_0) G_{n,\hat{\psi}_n}) \cdot G_{n,\hat{\psi}_n}^{-1} \right) \left( G_{n,\hat{\psi}_n}(z) \right).
\]

Let \( h \equiv n^{1/3}(z - z_0) \) be the local variable and define the normalized processes:

\[
\hat{B}_{n,\Lambda}(h) = n^{2/3} \frac{1}{I(\psi_0(z_0)) p_Z(z_0)} \left[(B_{n,\Lambda}(z_0 + h n^{-1/3}) - B_n(z_0)) - \psi_0(z_0) (G_{n,\Lambda}(z_0 + h n^{-1/3}) - G_{n,\Lambda}(z_0)) \right]
\]

and

\[
\hat{G}_{n,\Lambda}(h) = n^{1/3} \frac{1}{I(\psi_0(z_0)) p_Z(z_0)} (G_{n,\Lambda}(z_0 + h n^{-1/3}) - G_{n,\Lambda}(z_0)).
\]

In terms of the local variable and the normalized processes, it is not difficult to see that:

\[
n^{1/3} (\hat{\psi}_n(z_0 + h n^{-1/3}) - \psi_0(z_0)) = \text{slogcm} \left( \hat{B}_{n,\hat{\psi}_n} \cdot \hat{G}_{n,\hat{\psi}_n}^{-1} \right) \left( \hat{G}_{n,\hat{\psi}_n}(h) \right).
\]

For a function \( g \) defined on \( \mathbb{R} \) let \( \text{slogcm}^0 g \) denote (i) for \( h < 0 \), the minimum of the slope (left–derivative) of the GCM of the restriction of \( g \) to \( \mathbb{R}^- \) and 0, and (ii) for \( h > 0 \), the maximum of the slope (left–derivative) of the GCM of the restriction of \( g \) to \( \mathbb{R}^+ \) and 0. From the characterization of \( \hat{\psi}_n \) (refer to (4) and (5)) and the definitions of the normalized processes it follows that:

\[
n^{1/3} (\hat{\psi}_n^0(z_0 + h n^{-1/3}) - \psi_0(z_0)) = \text{slogcm}^0 \left( \hat{B}_{n,\hat{\psi}_n^0} \cdot \hat{G}_{n,\hat{\psi}_n^0}^{-1} \right) \left( \hat{G}_{n,\hat{\psi}_n^0}(h) \right).
\]
Thus,

\[(X_n(h), Y_n(h)) = \left\{ \text{slogcm} \left( \tilde{B}_{n\psi_0} \cdot \tilde{G}_{n\psi_0}^{-1} \right) \left( \tilde{G}_{n\psi_0}^{-1}(h) \right), \text{slogcm}^0 \left( \tilde{B}_{n\psi_0} \cdot \tilde{G}_{n\psi_0}^{-1} \right) \left( \tilde{G}_{n\psi_0}^{-1}(h) \right) \right\}. \]

By Lemma 2.4 of Banerjee (2004), the processes \( \tilde{B}_{n\psi_0}(h) - \tilde{B}_{n\psi_0}(h) \) and \( \tilde{B}_{n\psi_0}(h) - \tilde{B}_{n\psi_0}(h) \) converge in probability to 0 uniformly on every compact set under \( \{P^n_\psi\} \). By contiguity, the same convergences in probability continue to hold under \( \{P^n_\psi\} \). Furthermore, by Lemma 1 below, the process \( \tilde{B}_{n\psi_0}(h) \) converges to the process \( X_{a,b,D}(h) \) in \( B_{loc}(\mathbb{R}) \), under \( \{P^n_\psi\} \). It follows that the processes \( (\tilde{B}_{n\psi_0}(h), \tilde{B}_{n\psi_0}(h)) \rightarrow_{d} (X_{a,b,D}(h), X_{a,b,D}(h)) \) in the space \( B_{loc}(\mathbb{R}) \times B_{loc}(\mathbb{R}) \), equipped with the product topology. Furthermore, by Lemma 2.5 of Banerjee (2004), the processes \( (\tilde{G}_{n\psi_0}(h), \tilde{G}_{n\psi_0}(h)) \rightarrow_{p} (h, h) \) in the space \( B_{loc}(\mathbb{R}) \times B_{loc}(\mathbb{R}) \), under \( \{P^n_\psi\} \). Again, by contiguity, the same convergences in probability continue to hold under \( \{P^n_\psi\} \). The proof is now completed by invoking standard continuous mapping arguments for slopes of greatest convex minorant estimators (see, for example, Prakasa Rao (1969) or Wright (1981)); thus, the limit distributions of \( X_n \) and \( Y_n \) are obtained by replacing the processes on the right side of (6) by their limits. It follows that for any \((h_1, h_2, \ldots, h_k)\),

\[\{X_n(h_i), Y_n(h_i)\}_{i=1}^{k} \rightarrow_{d} \{\text{slogcm} X_{a,b,D}(h_i), \text{slogcm}^0 X_{a,b,D}(h_i)\}_{i=1}^{k} \equiv_{d} \{g_{a,b,D}(h_i), g_{a,b,D}^0(h_i)\}_{i=1}^{k}, \]

where \( \equiv_{d} \) denotes equality in distribution. The above finite dimensional convergence, coupled with the monotonicity of the functions involved, allows us to conclude that convergence happens in the space \( L \times L \). The strengthening of finite dimensional convergence to convergence in the \( L_2 \) metric is deduced from the monotonicity of the processes \( X_n \) and \( Y_n \), as in Corollary 2 of Theorem 3 in Huang and Zhang (1994). If \( \phi_n, \phi \) are monotone functions such that

\[\phi_n(t) \mid (t_1, t_2, \ldots, t_k) \rightarrow \phi(t) \mid (t_1, t_2, \ldots, t_k)\]

for every \( t_1 < t_2 < \ldots < t_k \), then \( \phi_n \) converges to \( \phi \) in the \( L_2 \) sense, on every compact set. \( \Box \)

We will now deduce the asymptotic distribution of the LRS, \( 2 \log \lambda_n \), for testing \( \psi(z_0) = \theta_0 \) under the sequence of local alternatives \( \{P^n_\psi\} \).
Theorem 3 Let 
\[ \phi(t) = (b/a)^{4/3} \mathcal{D}_0((b/a)^{-2/3} t) \text{ and } \mathcal{D}_0(t) = b^{-1} \mathcal{D}(t). \]

Let \( X_{1,1,\phi}(t) = W(t) + t^2 + \phi(t) \) and let \( g_{1,1,\phi} \) and \( g_{1,1,\phi}^0 \) be functionals of the process \( X_{1,1,\phi}(t) \) that are extracted in the exact same way that \( g_{a,b,\mathcal{D}} \) and \( g_{a,b,\mathcal{D}}^0 \) are extracted from the process \( X_{a,b,\mathcal{D}} \).

Under the sequence of local alternatives, \( \{P^n_{\psi_n}\} \), the LRS, \( 2 \log \lambda_n \), for testing \( H_0 : \psi(z_0) = \theta_0 \) converges in distribution to,
\[ \int \left( (g_{1,1,\phi}(h))^2 - (g_{1,1,\phi}^0(h))^2 \right) dh. \]

Proof of Theorem 3: From Theorem 2.2 of Banerjee (2004), we know that
\[ 2 \log \lambda_n = \frac{1}{a^2} \int (X_n^2(h) - Y_n^2(h)) \, dh + o_p(1) \]
under the sequence \( \{P^n_{\psi_0}\} \). By contiguity, it follows that under the sequence \( \{P^n_{\psi_n}\} \), the above representation continues to hold. It therefore suffices to find the asymptotic distribution of
\[ \xi_n = \frac{1}{a^2} \int (X_n^2(h) - Y_n^2(h)) \, dh. \]

Let \( \mathcal{D}_n \) denote the set on which \( X_n \) and \( Y_n \) differ. Let \( D_{a,b,\mathcal{D}} \) denote the set on which the processes \( g_{a,b,\mathcal{D}} \) and \( g_{a,b,\mathcal{D}}^0 \) differ. We will now deduce the limit distribution of \( \xi_n \). For each \( \epsilon > 0 \) we can find a compact set \( M_\epsilon \) of the form \([-K_\epsilon, K_\epsilon]\) such that eventually,
\[ P \left[ \mathcal{D}_n \subset [-K_\epsilon, K_\epsilon] \right] > 1 - \epsilon \text{ and } P \left[ D_{a,b,\mathcal{D}} \subset [-K_\epsilon, K_\epsilon] \right] > 1 - \epsilon. \]

For a proof of this result when \( \mathcal{D} = 0 \) (the situation under the null hypothesis) see Lemma 2.6 of Banerjee (2004). Minor extensions allow one to incorporate a general shift function \( \mathcal{D} \). Let
\[ X_{ne} = a^{-2} \int_{M_\epsilon} (X_n^2(h) - Y_n^2(h)) \, dh, \quad W_\epsilon = \int_{M_\epsilon} a^{-2} \left( (g_{a,b,\mathcal{D}}(h))^2 - (g_{a,b,\mathcal{D}}^0(h))^2 \right) \, dh. \]

Since \( M_\epsilon \) contains \( \mathcal{D}_n \) with probability greater than \( 1 - \epsilon \) eventually, we have \( P [X_{ne} \neq \xi_n] < \epsilon \) eventually. Similarly \( P [W_\epsilon \neq \xi] < \epsilon \), where
\[ \xi = \int a^{-2} \left( (g_{a,b,\mathcal{D}}(h))^2 - (g_{a,b,\mathcal{D}}^0(h))^2 \right) \, dz. \]
Also $X_{ne} \rightarrow_d W_{\epsilon}$ as $n \rightarrow \infty$, for every fixed $\epsilon$. This is so because by Theorem 2,

$$(X_n(h), Y_n(h)) \rightarrow_d (g_{a,b,D}(z), g_{a,b,D}^0(z))$$

as a process in $\mathcal{L} \times \mathcal{L}$ and because for any $K > 0$

$$(f, g) \mapsto \int_{[-K,K]} (f^2(h) - g^2(h)) \, dh$$

is a continuous function from $\mathcal{L} \times \mathcal{L}$ to the reals. Thus all conditions of Lemma 2 appearing below are satisfied, leading to the conclusion that $\xi_n \rightarrow_d \xi$. The fact that

$$\xi \equiv_d \int ((g_{1,1,\phi}(h))^2 - (g_{1,1,\phi}^0(h))^2) \, dh$$

where $\phi$ is defined in the statement of this theorem follows on using Lemma 2 in conjunction with Lemma 3 presented below. \(\square\)

### 2.1 Technical Lemmas

In this subsection, we collect the lemmas used in the proofs of Theorems 2 and 3 along with proofs.

The first lemma in this subsection is used in the proof of Theorem 2 and gives the limit distribution of the process $\tilde{B}_{n,\psi_0}(h)$ defined in the proof of that theorem.

**Lemma 1** Under the sequence of contiguous alternatives $\{P_{\psi_n}\}$, the process $\tilde{B}_{n,\psi_0}$ converges to the process $X_{a,b,D}$ in the space $B_{loc}(\mathbb{R})$.

**Proof:** We can write,

$$\tilde{B}_{n,\psi_0}(h) = \sqrt{n} \left( P_n - P_0 \right) f_{n,h} + \sqrt{n} P_0 f_{n,h},$$

where $P_0$ denotes the probability measure under $\psi_0$ and

$$f_{n}(h) = n^{1/6} \frac{[\psi_0(Z) - \psi_0(z_0)] \phi''(X, \psi_0(Z)) - \phi'(X, \psi_0(Z))}{I(\psi_0(z_0)) p_Z(z_0)} \left( 1 \{ Z \leq z_0 + h n^{-1/3} \} - 1 \{ Z \leq z_0 \} \right).$$

Now, $\sqrt{n} P_0 f_{n,h}$ is a deterministic sequence and converges uniformly on every compact interval to $b z^2$ (see the proof of Lemma 2.3 of Banerjee (2004) for the details). We will be concerned with
studying the behavior of $\tilde{M}_n(h) = \sqrt{n} \left( P_n - P_0 \right) f_{n,h}$. To establish the limit distribution of the process $\tilde{M}_n(h)$ under $\{P_{\psi_0}^n\}$, we will show that the finite dimensional marginals of $\tilde{M}_n(h)$ converge to the corresponding finite dimensional marginals of $aW(h) + D(h)$, $h \in [-K, K]$ for every $K > 0$ and then verify uniform asymptotic equicontinuity of the process $\tilde{M}_n(h)$ under $\{P_{\psi_0}^n\}$.

So, consider $\tilde{h} = (h_1, h_2, \ldots, h_k)$, where each $h_i \in [-K, K]$. We will find the joint distribution of $(\tilde{M}_n(h_1), \tilde{M}_n(h_2), \ldots, \tilde{M}_n(h_k))$ under the sequence $\{P_{\psi_0}^n\}$. To this end, we first obtain the joint distribution of $\{\tilde{M}_n(h_i)\}_{i=1,2,\ldots,k}$ and the local log-likelihood ratio $\log L_n(\psi_n) - \log L_n(\psi_0)$ under the sequence $\{P_{\psi_0}^n\}$. From Theorem 1 we obtain,

$$
\begin{pmatrix}
\{\tilde{M}_n(h_i)\}_{i=1,2,\ldots,k} \\
\log L_n(\psi_n) - \log L_n(\psi_0)
\end{pmatrix}
= \begin{pmatrix}
\{\sqrt{n}(P_n - P_0)f_{n,h_i}\}_{i=1,2,\ldots,k} \\
\sqrt{n}(P_n - P_0)s_n
\end{pmatrix} + \begin{pmatrix}
0_{k \times 1} \\
-\sigma^2/2
\end{pmatrix} + o_p(1),
$$

under $P_{\psi_0}^n$; here

$$s_n = -n^{1/6} \phi'(X, \psi_0(Z)) A_n(Z) 1\{Z \in [z_0 - c n^{-1/3}, z_0 + c n^{-1/3}]\}.$$

Denote the (column) vector $\{\tilde{M}_n(h_i)\}_{i=1,2,\ldots,k}$ by $T_{n,k}$. An application of the Lindeberg CLT for triangular arrays, in conjunction with the Cramer-Wold device (for details, see the technical complements to the proof at the end of this subsection) allows us to conclude that under $P_{\psi_0}^n$,

$$
\begin{pmatrix}
T_{n,k} \\
\log L_n(\psi_n) - \log L_n(\psi_0)
\end{pmatrix} \xrightarrow{d} N \left( \begin{pmatrix}
0_{k \times 1} \\
-\sigma^2/2
\end{pmatrix}, \begin{pmatrix}
\Sigma(\tilde{h}) & D(\tilde{h})^T \\
D(\tilde{h}) & \sigma^2
\end{pmatrix} \right),
$$

where $\Sigma(\tilde{h})_{ij} = \lim_{n \to \infty} P_0(f_{n,h_i} f_{n,h_j})$ and is 0 if $h_i$ and $h_j$ have opposite signs and is precisely $a^2(\| h_i \wedge h_j \|)$ if they have the same sign – this is the finite dimensional covariance matrix corresponding to the process $aW(h)$. For the computational details, see the proof of Lemma 2.3 of Banerjee (2004). On the other hand, $D(\tilde{h}) = (D(h_1), D(h_2), \ldots, D(h_k))^T$ is the asymptotic covariance between $s_n$ and $T_{n,k}$; thus $D(h_i) = \lim_{n \to \infty} P_0(f_{n,h_i} s_n)$. Let us compute the form of $D(\tilde{h})$. Consider $h > 0$. Then $E(f_{n,h} s_n) = I_n + II_n$ (here $E$ denotes expectation under $\psi_0$) where $I_n$ is given by

$$
-n^{1/3} \frac{E[(\psi_0(Z) - \psi_0(z_0)) \phi'(X, \psi_0(Z)) \phi''(X, \psi_0(Z)) B_n(n^{1/3}(Z - z_0)) 1\{Z \in [z_0, z_0 + (c \wedge h) n^{-1/3}]\}}{I(\psi_0(z_0)) p_Z(z_0)}.
$$

18
On changing to the local variable $u \equiv n^{1/3}(z - z_0)$, so that $z \equiv z_n(u) = z_0 + u n^{-1/3}$, the above expression becomes

$$- \int_0^{h \wedge c} (\psi_0(z_n(u)) - \psi_0(z_0)) E_{\psi_0(z_n(u))}[\phi'(X, \psi_0(z_n(u))) \phi''(X, \psi_0(z_n(u)))] p_Z(z_n(u)) B_n(u) \, du$$

$$I(\psi_0(z_0)) p_Z(z_0))$$

and this is clearly $o(1)$. On the other hand,

$$II_n = n^{1/3} \int_0^{h \wedge c} E[\phi'(X, \psi_0(Z)) \phi''(X, \psi_0(Z))] B_n(n^{1/3}(Z - z_0)) 1 \{ Z \in [z_0, z_0 + (c \wedge h) n^{-1/3}] \}$$

$$= \frac{1}{I(\psi_0(z_0)) p_Z(z_0))} \int_0^{h \wedge c} E_{\psi_0(z_0 + u n^{-1/3})}[\phi'(X, \psi_0(z_0 + u n^{-1/3}))^2] B_n(u) p_Z(z_0 + u n^{-1/3}) \, du$$

$$= \frac{1}{I(\psi_0(z_0)) p_Z(z_0))} \int_0^{h \wedge c} I(\psi_0(z_0 + u n^{-1/3})) B_n(u) p_z(z_0 + u n^{-1/3}) \, du$$

$$= o(1) + \int_0^{h \wedge c} B(u) \, du.$$

Thus, $E(f_n, s_n) = o(1) + \int_0^{h \wedge c} B(u) \, du$ when $h > 0$. Similarly, when $h < 0$ it can be shown that

$$E(f_n, s_n) = o(1) + \int_{-c \wedge h}^0 B(u) \, du.$$

Thus $D(h) \equiv (D(h_1), D(h_2), \ldots, D(h_k))$. Now, by LeCam’s third lemma (see Example 6.7, Page 90 of Van der Vaart (1998)), it follows immediately that

$$T_{n,k} \rightarrow_d N((D(h_1), D(h_2), \ldots, D(h_k))^T, \Sigma(h)),$$

under $\{P_{\psi_n}^n\}$. Thus, finite dimensionally, $\widehat{M}_n(h)$ converges to the process $a W(h) + D(h)$. The finite dimensional convergence can be extended to convergence in the space $l^\infty[-K, K]$, provided we can verify uniform asymptotic equicontinuity of the process $\widehat{M}_n(h)$ under $\{P_{\psi_n}^n\}$. This amounts to showing that for any sequence $\{\delta_n\}$ decreasing to 0,

$$\left( P_{\psi_n}^n * \left( \sup_{|s - t| < \delta_n, |s|, |t| \leq K} \left| \widehat{M}_n(s) - \widehat{M}_n(t) \right| > \epsilon \right) \right) \rightarrow 0,$$

for any preassigned $\epsilon > 0$. In the above display, the superscript $*$ is used to designate outer probability; this is necessary since the supremum in the above display is not necessarily measurable with respect to the underlying sigma field. The fact that the above display holds follows on noting that $\{P_{\psi_n}^n\}$ and $\{P_{\psi_0}^n\}$ are mutually contiguous and that the process $\widehat{M}_n(h)$ converges in $l^\infty[-K, K]$.
to a tight limit process under \( \{P^n_{\psi_0}\} \), which entails that under \( \{P^n_{\psi_0}\} \) it is uniformly asymptotically equicontinuous, i.e.

\[
(P^n_{\psi_0})^* \left( \sup_{|s-t|<\delta_n,|s|,|t|\leq K} \left| \tilde{M}_n(s) - \tilde{M}_n(t) \right| > \epsilon \right) \to 0.
\]

This finishes the proof. \( \square \)

The next lemma is a crucial tool in the proof of Theorem 3 and is adapted from Prakasa Rao (1969).

**Lemma 2** Suppose that \( \{X_{n\epsilon}\}_{n\in\mathbb{N}, \epsilon > 0}, \{\xi_n\}_{n\in\mathbb{N}} \) and \( \{W_\epsilon\}_{\epsilon > 0} \) are three sets of random variables and \( \xi \) is another random variable such that

(i) \( \lim_{\epsilon \to 0} \limsup_{n \to \infty} P[X_{n\epsilon} \neq \xi_n] = 0 \),

(ii) \( \lim_{\epsilon \to 0} P[W_\epsilon \neq \xi] = 0 \) and (iii) For every \( \epsilon > 0 \), \( X_{n\epsilon} \to_d W_\epsilon \) as \( n \to \infty \). Then \( \xi_n \to_d \xi \) as \( n \to \infty \).

For a proof of this lemma, see Prakasa Rao (1969).

The final lemma in this subsection is a Brownian scaling result and is used in the last step of the proof of Theorem 3 to express the limit distribution in an appropriate form.

**Lemma 3** For any \( M > 0 \), the following distributional equality holds in the space \( \mathcal{L} \times \mathcal{L} \):

\[
(g_{a,b,D}(t), g^0_{a,b,D}(t)) \equiv_d \left( a (b/a)^{1/3} g_{1,1,\phi} ((b/a)^{2/3} t), a (b/a)^{1/3} g^0_{1,1,\phi} ((b/a)^{2/3} t) \right),
\]

with \( \phi(t) \) as defined in Theorem 3.

**Proof of lemma:** We establish the distributional equality

\[
g_{a,b,D}(t) \equiv_d a (b/a)^{1/3} g_{1,1,\phi} ((b/a)^{2/3} t).
\]

The joint distributional equality is established similarly. We will use the following fact from Banerjee (2000): \( X_{a,b}(t) \equiv_d a (a/b)^{1/3} X_{1,1}((a/b)^{2/3} t) \). We also recall from the definition of \( \phi(t) \) in Theorem 3 that \( \phi((b/a)^{2/3} t) = (b/a)^{4/3} D_0(t) \). Now, \( a (a/b)^{1/3} X_{1,1,\phi}(t) = a (a/b)^{1/3} (X_{1,1}(t) + \phi(t)) \). Thus

\[
a (a/b)^{1/3} X_{1,1,\phi}((b/a)^{2/3} t) = a (a/b)^{1/3} X_{1,1}((b/a)^{2/3} t) + a (a/b)^{1/3} \phi((b/a)^{2/3} t)
\]

20
\[
\equiv_d X_{a,b}(t) + a (a/b)^{1/3} (b/a)^{4/3} D_0(t) \\
= X_{a,b}(t) + b D_0(t) \\
= X_{a,b}(t) + D(t).
\]

It follows that \( a (a/b)^{1/3} X_{1,1,\phi}((b/a)^{2/3} t) \equiv_d X_{a,b}(t) + D(t) \). Forming greatest convex minorants and differentiating with respect to \( t \) readily yields:

\[
g_{a,b,D}(t) = d a \left( a/b \right)^{1/3} g_{1,1,\phi} \left( (b/a)^{2/3} t \right).
\]

This finishes the proof. \( \square \)

We end with

**Technical complements to the proof of Lemma 1:** Here we provide the technical details behind the arguments involved in showing that under the sequence of probability measures \( \{P^\psi_n\} \),

\[
\left( \begin{array}{c}
T_{n,k} \\
\log L_n(\psi_n) - \log L_n(\psi_0)
\end{array} \right) \rightarrow_d N \left( \left( \begin{array}{c}
0_{k \times 1} \\
-\sigma^2/2
\end{array} \right), \left( \begin{array}{cc}
\Sigma(\tilde{h}) & D(\tilde{h})^T \\
D(\tilde{h}) & \sigma^2
\end{array} \right) \right),
\]

We use Lindeberg’s CLT along with the Cramer-Wold device. We first state Lindeberg’s theorem.

**Lindeberg’s CLT:** Suppose that for each \( n \), \( \{X_{n,1}, X_{n,2}, \ldots X_{n,r_n}\} \) are independent. Also suppose that \( E(X_{n,k}) = 0 \) and let \( \sigma^2_{nk} = E[X_{n,k}]^2 < \infty \). Define \( S_n \equiv X_{n,1} + X_{n,2} + \ldots + X_{n,r_n} \) and \( s_n^2 \equiv \sum_{k=1}^{r_n} \sigma^2_{nk} \). Suppose that the following condition (the Lindeberg condition) holds:

For every pre-assigned positive \( \epsilon \),

\[
\lim_{n \to \infty} \sum_{j=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{n,j}| \geq \epsilon s_n} X_{n,j}^2 dP = 0.
\]

Then

\[
\frac{S_n}{s_n} \rightarrow_d N(0,1).
\]

Consequently, if \( s_n \rightarrow \tau > 0 \), then \( S_n \rightarrow_d N(0, \tau^2) \). Recalling that \( T_{n,k} = \{\sqrt{n}(\mathbb{P}_n - P_0) f_{n,h_i}\}_{i=1,2,\ldots,k} \) and that

\[
\log L_n(\psi_n) - \log L_n(\psi_0) = \sqrt{n}(\mathbb{P}_n - P_0) s_n - \frac{1}{2} \sigma^2 + o_p(1),
\]

\[21\]
under \( \{P_n^{\psi_0}\} \), it suffices to show that
\[
\left( \sqrt{n}(P_n - P_0) f_{n,h_i}(X, Z) \right)_{i=1,2,...,k} \xrightarrow{d} N \left( \left( \begin{array}{c} 0_{k \times 1} \\ \Sigma(\tilde{h}) \\ D(\tilde{h})^T \end{array} \right), \left( \begin{array}{ccc} D(\tilde{h}) & 0 \\ 0 & \sigma^2 \end{array} \right) \right).
\]

We denote the dispersion matrix on the right side of the above display by \( V \). By the Cramer-Wold theorem it suffices to show that for any \( k + 1 \) dimensional vector \( (a_1, a_2, \ldots, a_k, b) \), it is the case that
\[
\sqrt{n}(P_n - P_0) \left[ \sum_{i=1}^{k} a_i f_{n,h_i}(X, Z) + bs_n(X, Z) \right] \xrightarrow{d} N(0, (a^T, b) V (a^T, b)^T).
\]

The left hand side of the above display is \( S_n \) in Lindeberg’s CLT, for
\[
X_{n,j} = \sum_{i=1}^{k} a_i f_{n,h_i}(X_j, Z_j) + bs_n(X_j, Z_j) - P_0 \left( \sum_{i=1}^{k} a_i f_{n,h_i}(X_j, Z_j) + bs_n(X_j, Z_j) \right) / \sqrt{n},
\]
for \( j = 1, 2, \ldots, n \). Clearly the \( X_{n,j} \)’s are i.i.d. for each \( n \). Write \( \sqrt{n} X_{n,j} = U_{n,j} - P_0(U_{n,j}) \). Then
\[
E S_n^2 = E [X_{n,1}]^2 + E [X_{n,2}]^2 + \ldots + E [X_{n,n}]^2
= n E [X_{n,1}]^2
= \text{Var}(U_{n,1})
= (a^T, b) \text{Cov} \left( \{f_{n,h_i}(\Delta_1, T_1)\}_{i=1,...,k}, s_n(\Delta_1, T_1) \right)^T (a^T, b)^T
= (a^T, b) V (a^T, b)^T + o(1).
\]

Thus \( s_n = \sqrt{(a^T, b) V (a^T, b)^T + o(1) = \tau + o(1) \). To show that \( S_n \to N(0, \tau^2) \) we just need to verify the Lindeberg condition. At this point, we note that, for a suitably large constant \( G \), we can obtain a function
\[
F_n = G n^{1/6} \left( | \phi'(X, \psi_0(Z)) | + | \phi''(X, \psi_0(Z)) | \right) 1 (Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}])
\]
(without loss of generality \( K \) can be taken to be larger than \( c \)) which dominates each \( f_{n,h_i} \) and \( s_n \) and satisfies
\[
\int_{|F_n| > \eta \sqrt{n}} F_n^2 dP \to 0 \quad \forall \eta > 0. \quad (7)
\]

To show this, note that the expression in the above display is bounded (up to a constant) by
\[
n^{1/3} E \left( (\phi''(X, \psi_0(Z))^2 + \phi'(X, \psi_0(Z))^2) 1 \{ | \phi'(X, \psi_0(Z)) | + | \phi''(X, \psi_0(Z)) | > \tilde{\eta} \} \right) > \tilde{\eta} n^{1/3}
\]
\[ Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \] , for some positive constant \( \tilde{\eta} \). This, in turn, is bounded by

\[ n^{1/3} \int_{z_0 - K n^{-1/3}}^{z_0 + K n^{-1/3}} E_\psi(z) \left[ (\phi''(X, \psi_0(z))^2 + \phi'(X, \psi_0(z))^2) \right. \left. (1 \{|\phi'(X, \psi_0(z))| > \tilde{\eta} n^{1/3}/2\} \right. \\
+ \left. 1 \{ |\phi''(X, \psi_0(z))| > \tilde{\eta} n^{1/3}/2\} \right) p_Z(z) \, dz . \]

Using Assumption (A.7) we can ensure that for any pre-assigned \( \epsilon > 0 \), for all sufficiently large \( n \), the above integral is bounded by \( 2 \epsilon p_Z(z_0) \). Since \( \epsilon > 0 \) can be made arbitrarily small it follows that the above integral converges to 0.

We can also assume without loss of generality that \( U_{n,j} \leq F_n \) (this is achieved at the cost of increasing the constant \( G \), depending upon the \( a_i \)'s and \( b \)). Now, using the fact that \( s_n \to \tau \), the Lindeberg condition boils down to showing that for every \( \epsilon > 0 \), \( \sum_{j=1}^{n} \int_{|X_{n,j}| \geq \epsilon} X_{n,j}^2 \, dP \to 0 \). Since the \( X_{n,j} \)'s are i.i.d. this retranslates as \( n \int_{|X_{n,1}| \geq \epsilon} X_{n,1}^2 \, dP \to 0 \). Using that \( \sqrt{n} X_{n,1} = U_{n,1} - P U_{n,1} \) this can be rewritten as

\[ \int_{|U_{n,1} - P U_{n,1}| \geq \sqrt{\epsilon} \sqrt{n}} [U_{n,1} - P U_{n,1}]^2 \, dP \to 0 . \] (8)

Now, since \( P U_{n,1} \to 0 \), it follows that for all sufficiently large \( n \),

\[ \{|U_{n,1} - P U_{n,1}| \geq \sqrt{n} \epsilon\} \subset \{|U_{n,1}| \geq \sqrt{n} \epsilon/2\} . \]

Denote the set on the right side of the above display by \( D_\epsilon \). Also, \( [U_{n,1} - P U_{n,1}]^2 \leq 2 U_{n,1}^2 + 2 (P U_{n,1})^2 \). In light of these observations (8) follows if we can show that

\[ \int_{D_\epsilon} U_{n,1}^2 \, dP + \int_{D_\epsilon} (P U_{n,1})^2 \, dP \to 0 . \]

The second term goes to 0, since \( P U_{n,1} \) does. The first term is dominated by \( \int_{F_n \geq \sqrt{n} \epsilon/2} F_n^2 \, dP \) which goes to 0, in light of (7). \( \Box \)
3 Behavior of likelihood ratio tests under fixed alternatives

Likelihood ratio statistics in nonregular monotone function problems behave in a way similar to those in regular parametric models, under a fixed alternative hypothesis. The LRS scaled by the sample size converges, under a fixed alternative hypothesis, to the Kullback–Leibler distance between the true underlying measure in the alternative and the measure in the null hypothesis that is closest to the true measure. We state this result in a more formal manner below.

In what follows we work under the setup described in Section 1. As before, our null hypothesis $H_0$ stipulates $\psi(z_0) = \theta_0$. We consider the alternative hypothesis $H_1: \psi(z_0) = \theta_1$ with $\theta_1 \neq \theta_0$. There are two possibilities – (a) $\theta_1 < \theta_0$ and (b) $\theta_1 > \theta_0$. We deal with case (a). Case (b) can be handled similarly.

**Theorem 4** Consider testing $H_0: \psi(z_0) = \theta_0$ versus $\psi(z_0) = \theta_1$ with $\theta_1 < \theta_0$. Suppose that $\psi_0$, the true monotone (increasing) function satisfies $\psi_0(z_0) = \theta_1 < \theta_0$. Also suppose that there exists an interior point in the support of $Z$, say $z_1 > z_0$, such that $\psi_0(z_1) = \theta_0$ and that $\psi_0$ is continuously differentiable in a neighborhood of $z_1$ with $\psi_0'(z_1) > 0$. Define the function $\tilde{\psi}_0$ as follows:

\[
\tilde{\psi}_0(z) = \psi_0(z) \lor \theta_0, \text{ for } z \geq z_0,
\]

\[
\tilde{\psi}_0(z) = \psi_0(z) \land \theta_0 \text{ for } z < z_0.
\]

Then,

\[
2 \frac{\log \lambda_n}{n} \to_p 2 K(P_{\psi_0}, P_{\tilde{\psi}_0}) = 2 \inf \{ K(P_{\psi_0}, P_{\psi}) : \psi \in R(z_0, \theta_0) \},
\]

where $K(P, Q) = E_P \log (dP/dQ)$ is the Kullback–Leibler discrepancy between $P$ and $Q$, $P_\psi$ denotes the joint distribution of $(X, Z)$ under the monotone increasing function $\psi$ and $R(z_0, \theta_0)$ is the set of all monotone increasing functions defined on the support of $Z$ that assume the value $\theta_0$ at the point $z_0$. 

24
Since $2K(P_{\psi_0}, P_{\tilde{\psi}_0}) > 0$ it follows that the likelihood ratio test is consistent under the fixed alternative.

The proof of this theorem is somewhat long and technical and is skipped. For a proof of this proposition in the special case of the current status model, see Chapter 3 of Banerjee (2000). Here, we only show that in the class of functions $\mathcal{R}(z_0, \theta_0)$, $\tilde{\psi}_0$ is closest to $\psi_0$ in Kullback–Leibler distance.

For any function $\tilde{\psi} \in \mathcal{R}(z_0, \theta_0)$, we have

\[ K(P_{\psi_0}, P_{\tilde{\psi}}) = \int \tilde{K}(\psi_0(z), \tilde{\psi}(z)) p_Z(z) \, dz, \quad (9) \]

where $\tilde{K}(\theta_1, \theta_2) = E_{P_{\theta_1}}( \log dP_{\theta_1}/dP_{\theta_2} )$ and for $\theta \in \Theta$, $P_\theta$ denotes the probability measure described by the probability density $p(\cdot, \theta)$. Note that $K(\theta_1, \theta_2)$ is always non–negative. Hence,

\[ K(P_{\psi_0}, P_{\tilde{\psi}}) \geq \int_{[z_0, z_1]} \tilde{K}(\psi_0(z), \tilde{\psi}(z)) p_Z(z) \, dz \]

\[ = \int_{[z_0, z_1]} \tilde{K}(\psi_0(z), \tilde{\psi}_0(z)) p_Z(z) \, dz \]

\[ + \int_{[z_0, z_1]} \left[ \int \left( l(x, \tilde{\psi}_0(z)) - l(x, \tilde{\psi}(z)) \right) p(x, \psi_0(z)) \, d\mu(x) \right] p_Z(z) \, dz \]

\[ = K(P_{\psi_0}, P_{\tilde{\psi}_0}) + \int_{[z_0, z_1]} \left[ \int \left( l(x, \tilde{\psi}_0(z)) - l(x, \tilde{\psi}(z)) \right) p(x, \psi_0(z)) \, d\mu(x) \right] p_Z(z) \, dz. \]

It suffices to show that the second term in the display immediately above is non–negative. To this end, it suffices to show that for each fixed $z$ in $[z_0, z_1]$,

\[ \int \left( l(x, \tilde{\psi}(z)) - l(x, \tilde{\psi}_0(z)) \right) p(x, \psi_0(z)) \, d\mu(x) \geq 0; \]

equivalently,

\[ \int \left( l(x, \tilde{\psi}(z)) - l(x, \tilde{\psi}_0(z)) \right) p(x, \psi_0(z)) \, d\mu(x) \leq 0. \]

Note that, for $z$ in $[z_0, z_1]$, $\tilde{\psi}(z) \geq \psi(z_0) = \theta_0 = \tilde{\psi}_0(z)$. Hence,

\[ l(x, \tilde{\psi}(z)) - l(x, \tilde{\psi}_0(z)) = (\tilde{\psi}(z) - \tilde{\psi}_0(z)) \hat{l}(x, \psi^*(z)), \]

for some point $\psi^*(z)$ in $[\tilde{\psi}_0(z), \tilde{\psi}(z)]$, that possibly depends on $x$. Since $\psi^*(z) \geq \theta_0 \geq \psi_0(z)$ and $\hat{l}(x, \theta)$ is decreasing in $\theta$ it follows that $\hat{l}(x, \psi^*(z)) \leq \hat{l}(x, \psi_0(z))$. It is readily concluded that,

\[ l(x, \tilde{\psi}(z)) - l(x, \tilde{\psi}_0(z)) \leq (\tilde{\psi}(z) - \tilde{\psi}_0(z)) \hat{l}(x, \psi_0(z)). \]
Consequently,
\[
\int \left( l(x, \psi(z)) - l(x, \tilde{\psi}_0(z)) \right) p(x, \psi_0(z)) \, d\mu(x) \leq \int (\psi(z) - \tilde{\psi}_0(z)) \dot{l}(x, \psi_0(z)) p(x, \psi_0(z)) \, d\mu(x) = 0.
\]
This finishes the proof.

4 Comments

The ideas in this paper can be extended in a straightforward way to the problem of constraining a monotone function at finitely many points. Under a null hypothesis of the form \( \psi(z_i) = \theta_i \) where \( z_1 < z_2 < \ldots < z_k \) and \( \theta_1 < \theta_2 < \ldots < \theta_k \), it can be shown that the LRS converges in distribution to the \( k \)-fold convolution of \( \mathbb{D} \). The right kind of local alternatives \( \psi_n \) to consider in this situation are ones that differ from \( \psi \) (which lies in the null) in shrinking \( O(n^{-1/3}) \) neighborhoods of \( z_i \) for each \( i \) and converge to \( \psi \) with increasing \( n \). The likelihood ratio statistic under such local alternatives will converge to the sum of \( k \) independent random variables \( V_1 + V_2 + \ldots + V_k \), where \( V_i \) has the form
\[
\int \left\{ (g_{1,1,\phi_i}(h))^2 - (g_{1,1,\phi_i}^0(h))^2 \right\} \, dh
\]
and the function \( \phi_i \) depends on the underlying parameters of the problem and the point \( z_i \).

While we have constructed contiguous alternatives by using the functions \( B_n \) that vanish outside of a compact set and converge uniformly to some \( B \), it is not yet clear to what extent the restrictions on the \( B_n \)'s imposed here are necessary for contiguity. In particular, is it possible to construct contiguous alternatives using square integrable functions on the line, i.e. functions \( B \) that do not vanish outside a compact set but satisfy \( \int_{-\infty}^{\infty} B^2(z) \, dz < \infty \)? This question is suggested by the fact that in each of these nonregular problems the integral of \( B^2 \) determines the variance in the LAN expansion (up to constants). More generally, there seems to be some potential for research in the direction of obtaining the weakest possible conditions for contiguity in these nonstandard problems.

Acknowledgments: The author acknowledges the support of National Science Foundation
grant DMS-0306235.

References


---

**MOULINATH BANERJEE**

**UNIVERSITY OF MICHIGAN**

**DEPARTMENT OF STATISTICS**

**439, WEST HALL, 550, EAST UNIVERSITY**

**ANN ARBOR, MI, 48105**

U.S.A.

*e-mail: moulib@umich.edu*