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

MOULINATH BANERJEE

Department of Statistics, University of Michigan

ABSTRACT. We focus on a class of non-standard problems involving non-parametric estimation of a monotone function that is characterized by $n^{1/3}$ rate of convergence of the maximum likelihood estimator, 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 $\psi(z_0) = \theta_0$ (ψ 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 behaviour under local alternatives of the form $\psi_n(z)$, where $\psi_n(\cdot)$ and $\psi(\cdot)$ vary in $O(n^{-1/3})$ neighbourhoods around z_0 and ψ_n converges to ψ at rate $n^{1/3}$ in an appropriate metric. Apart from local alternatives, we also consider the behaviour 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 non-standard 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, ψ is a monotone function of interest and ξ is a nuisance parameter. Denote by $\hat{\psi}_n$, the maximum likelihood estimator (MLE) of ψ 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

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

where the random variable $\mathbb{Z}$ 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, $\mathbb{Z} = \text{argmin}_h W(h) + h^2$, where $W(h)$ is standard two-sided Brownian motion on the line. Models of the above kind abound in the non-parametric literature. For example, the monotone function of interest can be a distribution function (survival analysis) which is the case with current status data studied in Groeneboom & Wellner (1992), or more generally Case k interval-censored data (see, e.g. 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 in Sun & Kalbfleisch (1995), Wellner & Zhang (2000), or a monotone density as in Prakasa Rao (1969), Huang & Zhang (1994) and Huang & Wellner (1995).

The behaviour 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

$$\lambda_n = \log \frac{\Pi_{i=1}^n f(W_i, \hat{\psi}_n, \xi)}{\Pi_{i=1}^n f(W_i, \hat{\psi}_n^0, \xi)},$$

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

$$\mathbb{D} \equiv \int \left((g_{1,1}(h))^2 - (g_{1,1}^0(h))^2 \right) \, \mathrm{d}h,$$

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 greatest convex minorant (GCM) 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) neighbourhood of 0, whereas $g_{a,b}(h)$ is almost surely non-zero in some (random) neighbourhood 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 behaviour 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 Θ being an open subset of $\mathbb{R}$, be a one-parameter family of probability densities with respect to a dominating measure μ and satisfying the assumptions A1–A7 given below. Let ψ be an increasing or decreasing continuous function defined on an interval $\tilde{I}$ and taking values in Θ. 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 ξ referred to at the beginning of this section while ψ is the parameter of interest. In what follows, it will be seen that the computation of the MLEs of ψ and the LRS for hypothesis tests involving ψ do not require estimation of the nuisance parameter ξ.

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

(a) p_Z is positive and continuous in a neighbourhood of z_0,
(b) ψ is continuously differentiable in a neighbourhood of z_0 with $|\psi'(z_0)| > 0$.

Denote by $\hat{\psi}_n$ the unconstrained MLE of ψ and by $\hat{\psi}_n^0$ the MLE of ψ under the constraint imposed by the pointwise null hypothesis $H_0 : \psi(z_0) = \theta_0$. Consider the LRS for testing the hypothesis $H_0 : \psi(z_0) = \theta_0$, where θ_0 is an interior point of Θ. Denoting the LRS by $2 \log \lambda_n$, we have

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

Assumptions

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

(A1) The set $\mathcal{X}_\theta : \{x : p(x, \theta) > 0\}$ does not depend on θ and is denoted by $\mathcal{X}$.

(A2) $l(x, \theta) = \log p(x, \theta)$ is at least three times differentiable with respect to θ and is strictly concave in θ for every fixed x. We write: $\dot{l}(x, \theta) = \partial l(x, \theta)/\partial \theta$ and $\ddot{l}(x, \theta) = \partial^2 l(x, \theta)/\partial \theta^2$.

(A3) If T is any statistic such that $E_\theta(|T|) < \infty$, then:

$$\frac{\partial}{\partial \theta} \int_\mathcal{X} T(x) p(x, \theta) \, \mathrm{d}x = \int_\mathcal{X} T(x) \frac{\partial}{\partial \theta} p(x, \theta) \, \mathrm{d}x$$

and

$$\frac{\partial^2}{\partial \theta^2} \int_\mathcal{X} T(x) p(x, \theta) \, \mathrm{d}x = \int_\mathcal{X} T(x) \frac{\partial^2}{\partial \theta^2} p(x, \theta) \, \mathrm{d}x.$$

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

(A4) $I(\theta)$ is finite and continuous at θ_0.

(A5) There exists a neighbourhood $\mathcal{N}$ of θ_0 such that for all x, $\sup_{\theta \in \mathcal{N}} |l'''(x, \theta)| \leq B(x)$, with B satisfying $\sup_{\theta \in \mathcal{N}} E_\theta(B(X)) < \infty$.

(A6) The functions:

$$f_1(\theta_1, \theta_2) = E_{\theta_1}(\dot{l}(X, \theta_2)^2) \text{ and } f_2(\theta_1, \theta_2) = E_{\theta_1}(\ddot{l}(X, \theta_2))$$

are continuous in a neighbourhood of (θ_0, θ_0). Also, the function $f_3(\theta) = E_\theta(\ddot{l}(X, \theta)^2)$ is uniformly bounded in a neighbourhood of θ_0.

(A7) Let $H(\theta, M)$ be defined as:

$$H(\theta, M) = E_\theta[(\dot{l}(X, \theta)^2 + \ddot{l}(X, \theta)^2)(1\{| \dot{l}(X, \theta) | > M\} + 1\{| \ddot{l}(X, \theta) | > M\})].$$

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

Note, in particular, that assumption (A7) 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 MLEs $\hat{\psi}_n$ and $\hat{\psi}_n^0$ exist.

Theorem 2.2 of Banerjee (2004) shows that under assumptions (a), (b) and (A1)–(A7), the LRS 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 ϵ_is have a common distribution with mean 0 and a finite variance, the Z_is have a common Lebesgue density p_Z and ψ is a monotone function. If the common error density is $N(0, \sigma^2)$, conditions (A1)–(A7) can be easily verified and the conclusion of theorem 2.2 of 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 behaviour 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 ψ_n is a sequence of monotone functions converging to ψ_0 (ψ_0 being a monotone function satisfying $H_0 : \psi(z_0) = \theta_0$) in such a way that the sequence of probability measures $\{P^n_{\psi_n}\}$ [the joint distribution of $(X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n)$ under ψ_n] is contiguous to $\{P^n_{\psi_0}\}$ [the joint distribution of $(X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n)$ under ψ_0]. Contiguity is deduced in each case via a local asymptotic normality (LAN) expansion of the local log-likelihood ratio and is then invoked to deduce the asymptotic distribution of $2 \log \lambda_n$ under $\{P^n_{\psi_n}\}$. 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 ψ_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 LRS 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 behaviour 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 ϕ that are square integrable on every compact set and ϕ_n is said to converge to ϕ 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. Behaviour of LRS 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 $P^n_{\psi_0}$, 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:

$$L_n(\psi_0) = \Pi^n_{i=1} \, p(X_i, \psi_0(Z_i)) \, p_Z(Z_i),$$

whereas the likelihood function corresponding to $P^n_{\psi_n}$ is

$$L_n(\psi_n) = \Pi^n_{i=1} \, p(X_i, \psi_n(Z_i)) p_Z(Z_i).$$

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 ψ_0 which satisfies the null hypothesis. We have,

$$
\begin{aligned}
\log L_n(\psi_n) - \log L_n(\psi_0) &= \log \Pi_{i=1}^n \frac{p(X_i, \psi_n(Z_i))}{p(X_i, \psi_0(Z_i))} \\
&= \sum_{i=1}^n l(X_i, \psi_n(Z_i)) - \sum_{i=1}^n l(X_i, \psi_0(Z_i)) \\
&= \sum_{i=1}^n \dot{l}(X_i, \psi_0(Z_i))(\psi_n(Z_i) - \psi_0(Z_i)) \\
&\quad + \sum_{i=1}^n \frac{1}{2} \ddot{l}(X_i, \psi_0(Z_i))(\psi_n(Z_i) - \psi_0(Z_i))^2 \\
&\quad + \sum_{i=1}^n \frac{1}{6} l'''(X_i, \psi_n^\star(Z_i))(\psi_n(Z_i) - \psi_0(Z_i))^3 \\
&\equiv I_n + II_n + III_n,
\end{aligned}
\tag{2}
$$

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

Using Lindeberg's central limit theorem (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 σ^2 explicitly below:

$$
\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} E(s_n^2) \\
&= \lim_{n\to\infty} E[n^{1/3} \dot{l}^2(X, \psi_0(Z)) A_n^2(Z)] \\
&= \lim_{n\to\infty} n^{1/3} \int_{z_0 - cn^{-1/3}}^{z_0 + cn^{-1/3}} I(\psi_0(z)) A_n^2(z) p_Z(z) \, dz \\
&= \lim_{n\to\infty} \int_{-c}^{c} I(\psi_0(z_0 + hn^{-1/3})) A_n^2(z_0 + hn^{-1/3}) p_Z(z_0 + hn^{-1/3}) \, dh \\
&= \lim_{n\to\infty} \int_{-c}^{c} I(\psi_0(z_0 + hn^{-1/3})) B_n^2(h) p_Z(z_0 + hn^{-1/3}) \, dh \\
&= I(\psi_0(z_0)) p_Z(z_0) \int_{-c}^{c} B^2(h) \, dh \\
&\equiv \frac{1}{a^2} \int_{-c}^{c} B^2(h) \, dh,
\end{aligned}
$$

with $a^2 = (I(\psi_0(z_0)) p_Z(z_0))^{-1}$. Next, we consider the asymptotic behaviour of II_n. We have $II_n = \frac{1}{2} \sum_{i=1}^n \ddot{l}(X_i, \psi_0(Z_i))(\psi_n(Z_i) - \psi_0(Z_i))^2 = \frac{n}{2} \mathbb{P}_n[\ddot{l}(X, \psi_0(Z))(\psi_n(Z) - \psi_0(Z))^2] \equiv \frac{1}{2} \mathbb{P}_n \xi_n(X, Z)$ where $\xi_n(x, z) = n^{1/3} \ddot{l}(x, \psi_0(z)) A_n^2(z)$. Now,

$$E(\mathbb{P}_n \xi_n(X, Z)) = n^{1/3} E[\ddot{l}(X, \psi_0(Z)) A_n^2(Z)]$$

$$= -n^{1/3} \int_{z_0 - cn^{-1/3}}^{z_0 + cn^{-1/3}} I(\psi_0(z)) A_n^2(z) p_Z(z) \, \mathrm{d}z$$

$$= -\int_{-c}^{c} I(\psi_0(z_0 + hn^{-1/3})) B_n^2(h) p_Z(z_0 + hn^{-1/3}) \, \mathrm{d}h$$

$$= -I(\psi_0(z_0)) p_Z(z_0) \int_{-c}^{c} B^2(h) \, \mathrm{d}h + o(1) \equiv -\sigma^2 + o(1).$$

On the other hand,

$$\mathrm{Var}(\mathbb{P}_n \xi_n(X, Z)) = \frac{\mathrm{Var}(\xi_n(X, Z))}{n}$$

$$\leq \frac{1}{n} E[n^{2/3} \ddot{l}(X, \psi(Z))^2 A_n^4(Z)]$$

$$= n^{-2/3} E[n^{1/3} \ddot{l}(X, \psi(Z))^2 A_n^4(Z)]$$

$$= n^{-2/3} \int_{-c}^{c} f_3(\psi_0(z_0 + hn^{-1/3})) B_n^4(h) \, \mathrm{d}h,$$

where f_3 is defined as in (A6). It follows from (A6) 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 $\mathrm{Var}(\mathbb{P}_n \xi_n(X, Z)) = O(n^{-2/3})$ 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 (A5) and the facts that (a) $\psi_n^\star(Z_i)$ lies between $\psi_n(Z_i)$ and $\psi_0(Z_i)$ and (b) $\sup_{z \in [z_0 - cn^{-1/3}, \, z_0 + cn^{-1/3}]} |\psi_n(z) - \psi_0(z)| \to 0$, 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

$$\log L_n(\psi_n) - \log L_n(\psi_0) \to_d Z_\sigma - \frac{\sigma^2}{2} \equiv W \sim N\left(\frac{-\sigma^2}{2}, \sigma^2\right),$$

under the sequence of probability measures $\{P_{\psi_0}^n\}$. Consequently,

$$\frac{\mathrm{d}P_{\psi_n}^n}{\mathrm{d}P_{\psi_0}^n} \equiv \frac{L_n(\psi_n)}{L_n(\psi_0)} \to_d \mathrm{e}^W,$$

with $E(\mathrm{e}^W) = 1$. 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 $\{P_{\psi_n}^n\}$ and $\{P_{\psi_0}^n\}$ are mutually contiguous.

As in Banerjee (2004), we denote the unconstrained MLE of ψ by $\hat{\psi}_n$ and the constrained MLE of ψ by $\hat{\psi}_n^0$. We now briefly review how the unconstrained and constrained MLEs of ψ 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 $x_0 < x_1 < \cdots < x_k$, 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 slogcm $\{(x_i, y_i)\}_{i=0}^{k}$.

Characterizing $\hat{\psi}_n$

In what follows, we define:

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

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 ψ. Let $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$ denote the ordered values of Z and $X_{(i)}$ denote the observed value of X corresponding to $Z_{(i)}$. As ψ is increasing $\psi(Z_{(1)}) \leq \psi(Z_{(2)}) \cdots \leq \psi(Z_{(n)})$. 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 $u_1 \leq u_2 \leq \cdots \leq u_n$. 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: $\hat{\psi}_n(Z_{(i)}) = \hat{u}_i$ for $i = 1, 2, \ldots, n$. Let $\nabla_i \tilde{\psi}(\hat{u})$ denote the ith partial derivative of $\tilde{\psi}$, evaluated at the vector $\hat{u}$ and let d_i denote the iith second partial derivative of $\tilde{\psi}$, 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

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

subject to the constraints that $u_1 \leq u_2 \leq \cdots \leq u_n$ and hence furnishes the isotonic regression of the function

$$g(i) = \hat{u}_i - \nabla_i \tilde{\psi}(\hat{u}) d_i^{-1}$$

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, e.g. theorem 1.2.1 of Robertson *et al.* (1988). In terms of the function ϕ the solution can be written as:

$$(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n) \equiv \left[\text{slogcm} \left\{ \sum_{j=1}^{i} \phi''(X_{(j)}, \hat{u}_j), \sum_{j=1}^{i} \hat{u}_j \phi''(X_{(j)}, \hat{u}_j) - \phi'(X_{(j)}, \hat{u}_j) \right\}_{i=0}^{n} \right]. \tag{3}$$

Note that the above is an implicitly defined equation for the $\hat{u}_i$s and solving it requires iterative techniques like the modified iterative convex minorant 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_is 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 $u_1 \leq u_2 \leq \cdots \leq u_m \leq \theta_0 \leq u_{m+1} \leq \cdots \leq u_n$. This can be

reduced to solving two separate optimization problems. These are: (1) minimize $\sum_{i=1}^{m} \phi'(X_{(i)}, u_i)$ over $u_1 \le u_2 \le \cdots \le u_m \le \theta_0$ and (2) minimize $\sum_{i=m+1}^{n} \phi'(X_{(i)}, u_i)$ over $\theta_0 \le u_{m+1} \le u_{m+2} \le \cdots \le u_n$. As in the case of the unconstrained solution, the constrained solution has a slope of GCM 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

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

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

$$(\hat{u}_1^0, \hat{u}_2^0, \ldots, \hat{u}_m^0) \equiv \left[\mathrm{slogcm} \left\{ \sum_{j=1}^{i} \phi''(X_{(j)}, \hat{u}_j^0), \sum_{j=1}^{i} \hat{u}_j^0 \phi''(X_{(j)}, \hat{u}_j^0) - \phi'(X_{(j)}, \hat{u}_j^0) \right\}_{i=0}^{m} \right] \wedge \theta_0,$$

(4)

and

$$(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \ldots, \hat{u}_n^0) \equiv \left[\mathrm{slogcm} \left\{ \sum_{j=m+1}^{i} \phi''(X_{(i)}, \hat{u}_i^0), \sum_{j=m+1}^{i} \hat{u}_i^0 \phi''(X_{(i)}, \hat{u}_i^0) - \phi'(X_{(i)}, \hat{u}_i^0) \right\}_{i=m}^{n} \right] \vee \theta_0.$$

(5)

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 MLEs $\hat{\psi}_n$ and $\hat{\psi}_n^0$, respectively, as follows:

$$X_n(h) = n^{1/3}(\hat{\psi}_n(z_0 + hn^{-1/3}) - \psi_0(z_0)) \text{ and } Y_n(h) = n^{1/3}(\hat{\psi}_n^0(z_0 + hn^{-1/3}) - \psi_0(z_0)).$$

We will study the limiting behaviour of the above processes under the contiguous sequence $\{P_{\psi_n}^n\}$. We first introduce the following quantities. Define the drift function $\mathcal{D}(h)$ as follows:

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

and

$$\mathcal{D}(h) = \int_{-c \vee h}^{0} B(u) \, du, \quad h \le 0.$$

Let,

$$a = (I(\psi_0(z_0)) p_Z(z_0))^{-1/2} \text{ and } b = \frac{\psi_0'(z_0)}{2}.$$

Define the process $X_{a,b,\mathcal{D}}(h) \equiv aW(h) + bh^2 + \mathcal{D}(h)$. Let $g_{a,b,\mathcal{D}}(h)$ be the slope (right-derivative) process of the GCM of $X_{a,b,\mathcal{D}}(h)$. Let $g_{a,b,\mathcal{D}}^0(h)$ be the (piecewise constant) process, which for $h \le 0$ is the minimum of the slope (right-derivative) of the GCM of $X_{a,b,\mathcal{D}}(h)$ on the set $\{h \le 0\}$ and 0, and for $h > 0$ is the maximum of the slope (right-derivative) of the GCM of $X_{a,b,\mathcal{D}}(h)$ on the set $\{h > 0\}$ and 0.

Theorem 2 deals with the limit distribution of the processes $(X_n(h), Y_n(h))$ and theorem 3 which uses theorem 2 provides the limit distribution of the LRS $2 \log \lambda_n$ for testing $H_0 : \psi(z_0) = \theta_0$ under the local alternatives $\{P_{\psi_n}^n\}$.

Theorem 2

Under the sequence of local alternatives $\{P^n_{\psi_n}\}$, the processes $(X_n(h), Y_n(h))$ converge finite-dimensionally and also as elements of the space $\mathcal{L} \times \mathcal{L}$ to the processes $(g_{a,b,\mathcal{D}}(z), g^0_{a,b,\mathcal{D}}(z))$.

Proof of theorem 2. For a monotone function Λ taking values in Θ define the following processes:

$$W_{n,\Lambda}(r) = \mathbb{P}_n[\phi'(X, \Lambda(Z))1(Z \le r)],$$

$$G_{n,\Lambda}(r) = \mathbb{P}_n[\phi''(X, \Lambda(Z))1(Z \le r)],$$

and

$$B_{n,\Lambda}(r) = \int_0^r \Lambda(z)\, \mathrm{d}G_{n,\Lambda}(z) - W_{n,\Lambda}(r).$$

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

$$\{\hat{\psi}_n(Z_{(i)})\}^n_{i=1} = \mathrm{slogcm}\ \{G_{n,\hat{\psi}_n}(Z_{(i)}), B_{n,\hat{\psi}_n}(Z_{(i)})\}^n_{i=0};$$

this is a direct consequence of (3). Consequently,

$$\{\hat{\psi}_n(Z_{(i)}) - \psi_0(z_0)\}^n_{i=1} = \mathrm{slogcm}\{G_{n,\hat{\psi}_n}(Z_{(i)}), B_{n,\hat{\psi}_n}(Z_{(i)}) - \psi_0(z_0)G_{n,\hat{\psi}_n}(Z_{(i)})\}^n_{i=0}.$$

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 $G^{-1}_{n,\hat{\psi}_n}$ (recall that for a right-continuous increasing function H we define $H^{-1}(t) = \inf\{s : H(s) \ge t\}$). Note that $G^{-1}_{n,\hat{\psi}_n}(0) = -\infty$. Defining $Z_{(0)} = -\infty$, so that $G_{n,\hat{\psi}_n}(Z_{(0)}) = 0$, we have:

$$G^{-1}_{n,\hat{\psi}_n}(t) = Z_{(i)} \text{ for } t \in (G_{n,\hat{\psi}_n}(Z_{(i-1)}), G_{n,\hat{\psi}_n}(Z_{(i)})].$$

Thus, the function $G^{-1}_{n,\hat{\psi}_n}$ is a piecewise constant *left-continuous* function; consequently, so is the function $(B_{n,\hat{\psi}_n} - G_{n,\hat{\psi}_n}) \cdot G^{-1}_{n,\hat{\psi}_n}$, with jumps at the points $\{G_{n,\hat{\psi}_n}(Z_{(i)})\}^n_{i=1}$. Denote by 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:

$$\hat{\psi}_n(z) - \psi_0(z_0) = \mathrm{slogcm}\left((B_{n,\hat{\psi}_n} - \psi_0(z_0)G_{n,\hat{\psi}_n}) \cdot G^{-1}_{n,\hat{\psi}_n}\right)\left(G_{n,\hat{\psi}_n}(z)\right).$$

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

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

and

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

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

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

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 $\hat{\psi}_n^0$ [refer to (4) and (5)] and the definitions of the normalized processes it follows that:

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

Thus,

$$(X_n(h), Y_n(h)) = \left\{\text{slogcm}\left(\tilde{B}_{n,\hat{\psi}_n} \cdot \tilde{G}_{n,\hat{\psi}_n}^{-1}\right)\left(\tilde{G}_{n,\hat{\psi}_n}(h)\right), \text{slogcm}^0\left(\tilde{B}_{n,\hat{\psi}_n^0} \cdot \tilde{G}_{n,\hat{\psi}_n^0}^{-1}\right)\left(\tilde{G}_{n,\hat{\psi}_n^0}(h)\right)\right\}.$$

(6)

By lemma 2.4 of Banerjee (2004), the processes $\tilde{B}_{n,\hat{\psi}_n^0}(h) - \tilde{B}_{n,\psi_0}(h)$ and $\tilde{B}_{n,\hat{\psi}_n}(h) - \tilde{B}_{n,\psi_0}(h)$ converge in probability to 0 uniformly on every compact set under $\{P_{\psi_0}^n\}$. By contiguity, the same convergences in probability continue to hold under $\{P_{\psi_n}^n\}$. Furthermore, by lemma 1 below, the process $\tilde{B}_{n,\psi_0}(h)$ converges to the process $X_{a,b,\mathcal{D}}(h)$ in $B_{\text{loc}}(\mathbb{R})$, under $\{P_{\psi_n}^n\}$. It follows that the processes $(\tilde{B}_{n,\hat{\psi}_n^0}(h), \tilde{B}_{n,\hat{\psi}_n}(h)) \to_d (X_{a,b,\mathcal{D}}(h), X_{a,b,\mathcal{D}}(h))$ in the space $B_{\text{loc}}(\mathbb{R}) \times B_{\text{loc}}(\mathbb{R})$, equipped with the product topology. Furthermore, by lemma 2.5 of Banerjee (2004), the processes $(\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{G}_{n,\hat{\psi}_n^0}(h)) \to_p (h, h)$ in the space $B_{\text{loc}}(\mathbb{R}) \times B_{\text{loc}}(\mathbb{R})$, under $\{P_{\psi_0}^n\}$. Again, by contiguity, the same convergences in probability continue to hold under $\{P_{\psi_n}^n\}$. The proof is now completed by invoking standard continuous mapping arguments for slopes of GCM estimators (see, e.g. Prakasa Rao, 1969 and 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 $(h_1 h_2, \ldots, h_k)$,

$$\{X_n(h_i), Y_n(h_i)\}_{i=1}^k \to_d \{\text{slogcm}\, X_{a,b,\mathcal{D}}(h_i), \text{slogcm}^0 X_{a,b,\mathcal{D}}(h_i)\}_{i=1}^k \equiv_d \{g_{a,b,\mathcal{D}}(h_i), g_{a,b,\mathcal{D}}^0(h_i)\}_{i=1}^k,$$

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 $\mathcal{L} \times \mathcal{L}$. 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 & Zhang (1994). If ϕ_n, ϕ are monotone functions such that

$$\phi_n(t) \mid (t_1, t_2, \ldots, t_k) \to \phi(t) \mid (t_1, t_2, \ldots, t_k)$$

for every $t_1 < t_2 < \cdots < t_k$, then ϕ_n converges to ϕ in the L_2 sense, on every compact set.

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 $\{P_{\psi_n}^n\}$.

Theorem 3
Let

$$\phi(t) = \left(\frac{b}{a}\right)^{4/3} \mathcal{D}_0\left(\left(\frac{b}{a}\right)^{-2/3} t\right) \quad and \quad \mathcal{D}_0(t) = b^{-1}\mathcal{D}(t).$$

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, $\{P_{\psi_n}^n\}$, the LRS, $2 \log \lambda_n$, for testing $H_0: \psi(z_0) = \theta_0$ converges in distribution to,

$$\int \left((g_{1,1,\phi}(h))^2 - (g_{1,1,\phi}^0(h))^2 \right) \, \mathrm{d}h.$$

Proof of theorem 3. From theorem 2.2 of Banerjee (2004), we know that

$$2\log \lambda_n = \frac{1}{a^2} \int (X_n^2(h) - Y_n^2(h)) \, \mathrm{d}h + o_p(1)$$

under the sequence $\{P_{\psi_0}^n\}$. By contiguity, it follows that under the sequence $\{P_{\psi_n}^n\}$, the above representation continues to hold. It therefore suffices to find the asymptotic distribution of

$$\xi_n = \frac{1}{a^2} \int (X_n^2(h) - Y_n^2(h)) \, \mathrm{d}h.$$

Let $\tilde{D}_n$ 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 ξ_n. For each $\epsilon > 0$ we can find a compact set M_ϵ of the form $[-K_\epsilon, K_\epsilon]$ such that eventually,

$$P\left[\tilde{D}_n \subset [-K_\epsilon, K_\epsilon] \right] > 1 - \epsilon \quad \text{and} \quad P\left[D_{a,b,\mathcal{D}} \subset [-K_\epsilon, K_\epsilon] \right] > 1 - \epsilon.$$

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

$$X_{n\epsilon} = a^{-2} \int_{M_\epsilon} (X_n^2(h) - Y_n^2(h)) \, \mathrm{d}h, \quad W_\epsilon = \int_{M_\epsilon} a^{-2} \left((g_{a,b,\mathcal{D}}(h))^2 - (g_{a,b,\mathcal{D}}^0(h))^2 \right) \, \mathrm{d}h.$$

As M_ϵ contains $\tilde{D}_n$ with probability greater than $1 - \epsilon$ eventually, we have $P[X_{n\epsilon} \neq \xi_n] < \epsilon$ eventually. Similarly $P[W_\epsilon \neq \xi] < \epsilon$, where

$$\xi = \int a^{-2} ((g_{a,b,\mathcal{D}}(h))^2 - (g_{a,b,\mathcal{D}}^0(h))^2) \, \mathrm{d}z.$$

Also $X_{n\epsilon} \to_d W_\epsilon$ as $n \to \infty$, for every fixed ϵ. This is so because by theorem 2,

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

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

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

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 $\xi_n \to_d \xi$. The fact that

$$\xi \equiv_d \int \left((g_{1,1,\phi}(h))^2 - (g_{1,1,\phi}^0(h))^2 \right) \, \mathrm{d}h$$

where ϕ is defined in the statement of this theorem follows on using lemma 2 in conjunction with lemma 3 presented below.

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 section 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 $\{P^n_{\psi_n}\}$, the process $\tilde{B}_{n,\psi_0}$ converges to the process $X_{a,b,\mathcal{D}}$ in the space $B_{\mathrm{loc}}(\mathbb{R})$.

Proof. We can write,

$$\tilde{B}_{n,\psi_0}(h) = \sqrt{n}(\mathbb{P}_n - P_0)f_{n,h} + \sqrt{n}P_0 f_{n,h},$$

where P_0 denotes the probability measure under ψ_0 and

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

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

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 $(\tilde{\mathbb{M}}_n(h_1), \tilde{\mathbb{M}}_n(h_2), \ldots, \tilde{\mathbb{M}}_n(h_k))$ under the sequence $\{P^n_{\psi_n}\}$. To this end, we first obtain the joint distribution of $\{\tilde{\mathbb{M}}_n(h_i)\}_{i=1,2,\ldots,k}$ and the local log-likelihood ratio $\log L_n(\psi_n) - \log L_n(\psi_0)$ under the sequence $\{P^n_{\psi_0}\}$. From theorem 1 we obtain,

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

under $P^n_{\psi_0}$; here

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

Denote the (column) vector $\{\tilde{\mathbb{M}}_n(h_i)\}_{i=1,2,\ldots,k}$ by $T_{n,k}$. An application of Lindeberg's 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 section) allows us to conclude that under $P^n_{\psi_0}$,

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

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, $D(\tilde{h}) = (D(h_1), D(h_2), \ldots, D(h_k))^{\mathrm{T}}$ 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 ψ_0) where I_n is given by

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

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

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

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

$$
\begin{aligned}
II_n &= n^{1/3} \frac{1}{I(\psi_0(z_0)) p_Z(z_0)} E[\phi'(X, \psi_0(Z))^2 B_n(n^{1/3}(Z - z_0)) 1[Z \in [z_0, z_0 + (c \wedge h) n^{-1/3}] \\
&= \frac{1}{I(\psi_0(z_0)) p_Z(z_0)} \int_0^{h \wedge c} E_{\psi_0(z_0 + un^{-1/3})}[\phi'(X, \psi_0(z_0 + un^{-1/3}))^2] B_n(u) p_Z(z_0 + un^{-1/3}) \, du \\
&= \frac{1}{I(\psi_0(z_0)) p_Z(z_0)} \int_0^{h \wedge c} I(\psi_0(z_0 + un^{-1/3})) B_n(u) p_z(z_0 + un^{-1/3}) \, du \\
&= o(1) + \int_0^{h \wedge c} B(u) \, du.
\end{aligned}
$$

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

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

Thus $D(\tilde{h}) \equiv (\mathcal{D}(h_1), \mathcal{D}(h_2), \ldots, \mathcal{D}(h_k))$. Now, by LeCam's third lemma (see example 6.7, page 90 of Van der Vaart, 1998), it follows immediately that

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

under $\{P^n_{\psi_n}\}$. Thus, finite dimensionally, $\tilde{\mathbb{M}}_n(h)$ converges to the process $aW(h) + \mathcal{D}(h)$. 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 $\tilde{\mathbb{M}}_n(h)$ under $\{P^n_{\psi_n}\}$. This amounts to showing that for any sequence $\{\delta_n\}$ decreasing to 0,

$$(P^n_{\psi_n})^* \left(\sup_{|s-t| < \delta_n; |s|, |t| \leq K} |\tilde{\mathbb{M}}_n(s) - \tilde{\mathbb{M}}_n(t)| > \epsilon \right) \to 0,$$

for any preassigned $\epsilon > 0$. In the above display, the superscript * is used to designate *outer probability*; this is necessary as 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 $\{P^n_{\psi_n}\}$ and $\{P^n_{\psi_0}\}$ are mutually contiguous and that the process $\tilde{\mathbb{M}}_n(h)$ 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.

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

This finishes the proof.

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

Lemma 2
Suppose that $\{X_{n\epsilon}\}_{n \in \mathcal{N}, \epsilon > 0}$, $\{\xi_n\}_{n \in \mathcal{N}}$ and $\{W_\epsilon\}_{\epsilon > 0}$ are three sets of random variables and ξ 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 section 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 $\mathcal{L} \times \mathcal{L}$:

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

with $\phi(t)$ as defined in theorem 3.

Proof of lemma 3. We establish the distributional equality

$$g_{a,b,\mathcal{D}}(t) \equiv_d a\left(\frac{b}{a}\right)^{1/3} g_{1,1,\phi}\left(\left(\frac{b}{a}\right)^{2/3}t\right).$$

The joint distributional equality is established similarly. We will use the following fact from Banerjee (2000): $X_{a,b}(t) \equiv_d a(a/b)^{1/3}X_{1,1}((b/a)^{2/3}t)$. We also recall from the definition of $\phi(t)$ in theorem 3 that $\phi((b/a)^{2/3}t) = (b/a)^{4/3}\mathcal{D}_0(t)$. Now, $a(a/b)^{1/3}X_{1,1,\phi}(t) = a(a/b)^{1/3}(X_{1,1}(t) + \phi(t))$. Thus

$$a\left(\frac{a}{b}\right)^{1/3} X_{1,1,\phi}\left(\left(\frac{b}{a}\right)^{2/3}t\right) = a\left(\frac{a}{b}\right)^{1/3} X_{1,1}\left(\left(\frac{b}{a}\right)^{2/3}t\right) + a\left(\frac{a}{b}\right)^{1/3}\phi\left(\left(\frac{b}{a}\right)^{2/3}t\right)$$

$$\equiv_d X_{a,b}(t) + a\left(\frac{a}{b}\right)^{1/3}\left(\frac{b}{a}\right)^{4/3}\mathcal{D}_0(t)$$

$$= X_{a,b}(t) + b\mathcal{D}_0(t)$$

$$= X_{a,b}(t) + \mathcal{D}(t).$$

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

$$g_{a,b,\mathcal{D}}(t) \equiv_d a\left(\frac{b}{a}\right)^{1/3} g_{1,1,\phi}\left(\left(\frac{b}{a}\right)^{2/3}t\right).$$

This finishes the proof.

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 $\{P^n_{\psi_0}\}$,

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

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 $S_n \equiv X_{n,1} + X_{n,2} + \cdots + X_{n,r_n}$ and $v^2_n \equiv \sum_{k=1}^{r_n} \sigma^2_{nk}$. Suppose that the following condition (the Lindeberg condition) holds:

For every pre-assigned positive ϵ,

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

Then

$$\frac{S_n}{v_n} \to_d N(0,1).$$

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

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

under $\{P_{\psi_0}^n\}$, it suffices to show that

$$\begin{pmatrix} \{\sqrt{n}(\mathbb{P}_n - P_0)f_{n,h_i}(X,Z)\}_{i=1,2,\ldots,k} \\ \sqrt{n}(\mathbb{P}_n - P_0)s_n(X,Z) \end{pmatrix} \to_d N\left(\begin{pmatrix} 0_{k\times 1} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma(\tilde{h}) & D(\tilde{h})^T \\ D(\tilde{h}) & \sigma^2 \end{pmatrix}\right).$$

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

$$\sqrt{n}(\mathbb{P}_n - P_0)\left[\sum_{i=1}^{k} a_i f_{n,h_i}(X,Z) + bs_n(X,Z)\right] \to_d N(0, (a^T, b)V(a^T, b)^T).$$

The left-hand side of the above display is S_n in Lindeberg's CLT, for

$$X_{n,j} \equiv \left[\sum_{i=1}^{k} a_i f_{n,h_i}(X_j,Z_j) + bs_n(X_j,Z_j) - P_0\left(\sum_{i=1}^{k} a_i f_{n,h_i}(X_j,Z_j) + bs_n(X_j,Z_j)\right)\right]n^{1/2},$$

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

$$\begin{aligned} ES_n^2 &= E[X_{n,1}]^2 + E[X_{n,2}]^2 + \cdots + E[X_{n,n}]^2 \\ &= nE[X_{n,1}]^2 \\ &= \mathrm{Var}(U_{n,1}) \\ &= (a^T, b)\mathrm{Cov}\left(\{f_{n,h_i}(\Delta_1, T_1)\}_{i=1,\ldots,k}, s_n(\Delta_1, T_1)\right)^T (a^T, b)^T \\ &= (a^T, b)V(a^T, b)^T + o(1). \end{aligned}$$

Thus $v_n = \sqrt{(a^T, b)V(a^T, b)^T} + o(1) = \tau + o(1)$. 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

$$F_n = Gn^{1/6}(|\phi'(X, \psi_0(Z))| + |\phi''(X, \psi_0(Z))|)1(Z \in [z_0 - Kn^{-1/3}, z_0 + Kn^{-1/3}])$$

(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

$$\int_{|F_n|>\eta\sqrt{n}} F_n^2 \, dP \to 0 \quad \forall \eta > 0. \tag{7}$$

To show this, note that the expression in the above display is bounded (up to a constant) by

$$n^{1/3} E\big[(\phi''(X,\psi_0(Z))^2 + \phi'(X,\psi_0(Z))^2) 1\{(|\ \phi'(X,\psi_0(Z))\ | + |\ \phi''(X,\psi_0(Z))\ |) > \tilde{\eta} n^{1/3}\}$$
$$\times\ 1\{Z \in [z_0 - Kn^{-1/3}, z_0 + Kn^{-1/3}]\}\big],$$

for some positive constant $\tilde{\eta}$. This, in turn, is bounded by

$$n^{1/3} \int_{z_0 - Kn^{-1/3}}^{z_0 + Kn^{-1/3}} E_{\psi(z)}\big[(\phi''(X,\psi_0(z))^2 + \phi'(X,\psi_0(z))^2)(1\{|\ \phi'(X,\psi_0(z))\ | > \tilde{\eta} n^{1/3}/2\}$$
$$+ 1\{|\ \phi''(X,\psi_0(z))\ | > \tilde{\eta} n^{1/3}/2\})\big] p_Z(z)\ \mathrm{d}z.$$

Using assumption (A7) 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_is and b). Now, using the fact that $v_n \to \tau$, the Lindeberg condition boils down to showing that for every $\epsilon > 0$, $\sum_{j=1}^{n} \int_{|X_{n,j}| \geq \epsilon} X_{n,j}^2\ \mathrm{d}P \to 0$. Since the $X_{n,j}$s are i.i.d. this retranslates as $n \int_{|X_{n,1}| \geq \epsilon} X_{n,1}^2\ \mathrm{d}P \to 0$. Using that $\sqrt{n} X_{n,1} = U_{n,1} - PU_{n,1}$ this can be rewritten as

$$\int_{|U_{n,1} - PU_{n,1}| \geq \sqrt{n}\epsilon} [U_{n,1} - PU_{n,1}]^2\ \mathrm{d}P \to 0. \tag{8}$$

Now, since $PU_{n,1} \to 0$, it follows that for all sufficiently large n,

$$\{|U_{n,1} - PU_{n,1}| \geq \sqrt{n}\epsilon\} \subset \{|U_{n,1}| \geq \sqrt{n}\epsilon/2\}.$$

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

$$\int_{D_\epsilon} U_{n,1}^2\ \mathrm{d}P + \int_{D_\epsilon} (PU_{n,1})^2\ \mathrm{d}P \to 0.$$

The second term goes to 0, since $PU_{n,1}$ does. The first term is dominated by $\int_{F_n \geq \sqrt{n}\epsilon/2} F_n^2\ \mathrm{d}P$ which goes to 0, in light of (7).

3. Behaviour of likelihood ratio tests under fixed alternatives

Likelihood ratio statistics in non-regular 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 ψ_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 ψ_0 is continuously differentiable in a neighbourhood of z_1 with $\psi_0'(z_1) > 0$. Define the function $\tilde{\psi}_0$ as follows:

$$\tilde{\psi}_0(z) = \psi_0(z) \vee \theta_0, \quad \textit{for } z \geq z_0,$$

$$\tilde{\psi}_0(z) = \psi_0(z) \wedge \theta_0 \quad \textit{for } z < z_0.$$

Then,

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

where $K(P, Q) = E_P \log (dP/dQ)$ is the Kullback–Leibler discrepancy between P and Q, P_ψ denotes the joint distribution of (X, Z) under the monotone increasing function ψ and $\mathcal{R}(z_0, \theta_0)$ is the set of all monotone increasing functions defined on the support of Z that assume the value θ_0 at the point z_0. 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 ψ_0 in Kullback–Leibler distance.

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

$$K(P_{\psi_0}, P_{\underline{\psi}}) = \int \tilde{K}(\psi_0(z), \underline{\psi}(z)) p_Z(z) \, \mathrm{d}z, \tag{9}$$

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

$$
\begin{aligned}
K(P_{\psi_0}, P_{\underline{\psi}}) &\geq \int_{[z_0, z_1]} \tilde{K}(\psi_0(z), \underline{\psi}(z)) p_Z(z) \, \mathrm{d}z \\
&= \int_{[z_0, z_1]} \tilde{K}(\psi_0(z), \tilde{\psi}_0(z)) p_Z(z) \, \mathrm{d}z \\
&\quad + \int_{[z_0, z_1]} \left[\int \left(l(x, \tilde{\psi}_0(z)) - l(x, \underline{\psi}(z)) \right) p(x, \psi_0(z)) \, \mathrm{d}\mu(x) \right] p_Z(z) \, \mathrm{d}z \\
&= K(P_{\psi_0}, P_{\tilde{\psi}_0}) + \int_{[z_0, z_1]} \left[\int \left(l(x, \tilde{\psi}_0(z)) - l(x, \underline{\psi}(z)) \right) p(x, \psi_0(z)) \, \mathrm{d}\mu(x) \right] p_Z(z) \, \mathrm{d}z.
\end{aligned}
$$

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 $[z_0, z_1]$,

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

equivalently,

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

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

$$l(x, \underline{\psi}(z)) - l(x, \tilde{\psi}_0(z)) = (\underline{\psi}(z) - \tilde{\psi}_0(z))\dot{l}(x, \psi^\star(z)),$$

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

$$l(x, \underline{\psi}(z)) - l(x, \tilde{\psi}_0(z)) \leq (\underline{\psi}(z) - \tilde{\psi}_0(z)) \dot{l}(x, \psi_0(z)).$$

Consequently,

$$\int \left(l(x, \underline{\psi}(z)) - l(x, \tilde{\psi}_0(z)) \right) p(x, \psi_0(z)) \, \mathrm{d}\mu(x)$$

$$\leq \int (\underline{\psi}(z) - \tilde{\psi}_0(z)) \dot{l}(x, \psi_0(z)) p(x, \psi_0(z)) \, \mathrm{d}\mu(x) = 0.$$

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 $z_1 < z_2 < \cdots < z_k$ and $\theta_1 < \theta_2 < \cdots < \theta_k$, it can be shown that the LRS converges in distribution to the k-fold convolution of $\mathbb{D}$. The right kind of local alternatives ψ_n to consider in this situation are ones that differ from ψ (which lies in the null) in shrinking $O(n^{-1/3})$ neighbourhoods of z_i for each i and converge to ψ with increasing n. The LRS under such local alternatives will converge to the sum of k independent random variables $V_1 + V_2 + \cdots + V_k$, where V_i has the form

$$\int \{(g_{1,1,\phi_i}(h))^2 - (g_{1,1,\phi_i}^0(h))^2\} \, \mathrm{d}h$$

and the function ϕ_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_ns 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 $\int_{-\infty}^{\infty} B^2(z) \, \mathrm{d}z < \infty$? This question is suggested by the fact that in each of these non-regular 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 non-standard problems.

Acknowledgements

References

Banerjee, M. (2000). *Likelihood ratio inference in regular and nonregular problems*. PhD dissertation, University of Washington, Washington.

Banerjee, M. (2004). Likelihood based inference for monotone functions: towards a general theory. *Technical Report 414*, Department of Statistics, University of Michigan. Available at http://www.stat.lsa.umich.edu/~moulib/wilks3.pdf.

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

Brunk, H. D. (1970). Estimation of isotonic regression. In *Nonparametric techniques in statistical inference* (ed. M. L. Puri), 177–195. Cambridge University Press, Cambridge.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Relat. Fields* **81**, 79–109.

Groeneboom, P. (1996). Lectures on Inverse Problems (École d'Été de Probabilités de Saint-Flour XXIV, 1994). In *Lecture Notes in Mathematics*, Vol. 1648 (ed. P. Bernard), 67–164. Springer-Verlag, New York.

Groeneboom, P. & Wellner, J. A. (1992). *Information bounds and nonparametric likelihood estimation*. Birkhäuser, Basel.

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

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

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7**, 310–321.

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

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

Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). *Order restricted statistical inference*. Wiley, New York.

Sun, J. & Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5**, 279–290.

Van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

Wellner, J. A. (2003). Gaussian white noise models: some results for monotone functions. In *Crossing boundaries: statistical essays in honor of Jack Hall* (eds J. E. Kolassa & D. Oakes), 87–104. IMS Lecture Notes-Monograph Series, Vol. 43.

Wellner, J. & Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.

Wright, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9**, 443–448.

Moulinath Banerjee, Department of Statistics, 439, West Hall, 550, East University, University of Michigan, Ann Arbor, MI 48105, USA.
E-mail: moulib@umich.edu