

# Likelihood Based Inference For Monotone Functions: Towards a General Theory

Moulinath Banerjee <sup>1</sup>

*University of Michigan*

*January 6, 2005*

## Abstract

The behavior of maximum likelihood estimates (MLE's) and the likelihood ratio statistic in a family of problems involving pointwise nonparametric estimation of a monotone function is studied. This class of problems differs radically from the usual parametric or semiparametric situations in that the MLE of the monotone function at a point converges to the truth at rate  $n^{1/3}$  (slower than the usual  $\sqrt{n}$  rate) with a non-Gaussian limit distribution. A unified framework for likelihood based estimation of monotone functions is developed and very general limit theorems describing the behavior of the MLE's and the likelihood ratio statistic are established. In particular, the likelihood ratio statistic is found to be asymptotically pivotal with a limit distribution that is no longer  $\chi^2$  but can be explicitly characterized in terms of a functional of Brownian motion. Special instances of the general theorems and potential extensions are discussed.

## 1 Introduction

A very common problem in nonparametric statistics is the need to estimate a function, like a density, a distribution, a hazard or a regression. Background knowledge about the statistical problem can provide information about certain aspects of the function of interest, which, if incorporated in the analysis, enables one to draw meaningful conclusions from the data. Often, this manifests itself in the nature of shape-restrictions (on the function) Monotonicity, in particular, is a shape-restriction that shows up very naturally in different areas of application like reliability, renewal theory, epidemiology and biomedical studies. Depending on the underlying problem, the monotone function of interest could be a distribution function or a cumulative hazard function (survival analysis), the mean function of a counting processes (demography, reliability, clinical trials), a monotone regression function (dose-response modeling, modeling disease incidence as a

---

<sup>1</sup>Research supported in part by National Science Foundation grant DMS-0306235

*Key words and phrases.* greatest convex minorant, ICM, likelihood ratio statistic, monotone function, self-consistent characterization, switching relationship, two-sided Brownian motion, universal limit, Wilks' phenomenon

function of distance from a toxic source), a monotone density (inference in renewal theory and other applications) or a monotone hazard rate (reliability). Unimodal functions are piecewise monotone; thus the problem of estimating a unimodal function can often be boiled down to estimating a monotone function.

Consequently, monotone functions have been fairly well-studied in the literature and several authors have addressed the problem of maximum likelihood estimation under monotonicity constraints. We point out some of the well-known ones. One of the earliest results of this type goes back to Prakasa Rao (1969) who derived the asymptotic distribution of the Grenander estimator (the MLE of a decreasing density); Brunk (1970) explored the limit distribution of the MLE of a monotone regression function, Groeneboom and Wellner (1992) studied the limit distribution of the MLE of the survival time distribution with current status data, Huang and Zhang (1994) and Huang and Wellner (1995) obtained the asymptotics for the MLE of a monotone density and a monotone hazard respectively with right censored data (the asymptotics for a monotone hazard under no censoring had been earlier addressed by Prakasa Rao (1970)) while Wellner and Zhang (2000) deduced the large sample theory for a pseudo-likelihood estimator for the mean function of a counting process. A common feature of these monotone function problems that sets them apart from the spectrum of regular parametric and semiparametric problems is the slower rate of convergence ( $n^{1/3}$ ) of the maximum likelihood estimates of the value of the monotone function at a fixed point (recall that the usual rate of convergence in regular parametric/semiparametric problems is  $\sqrt{n}$ ). What happens in each case is the following: If  $\hat{\psi}_n$  is the MLE of the monotone function  $\psi$ , then provided that  $\psi'(z)$  does not vanish,

$$n^{1/3}(\hat{\psi}_n(z) - \psi(z)) \rightarrow_d C(z)\mathbb{Z}, \quad (1.1)$$

where the random variable  $\mathbb{Z}$  is a symmetric (about 0) but *non-Gaussian* random variable and  $C(z)$  is a constant depending upon the underlying parameters in the problem and the point of interest  $z$ . In fact,  $\mathbb{Z} = \operatorname{argmin}_h W(h) + h^2$ , where  $W(h)$  is standard two-sided Brownian motion on the line. The distribution of  $\mathbb{Z}$  was analytically characterized by Groeneboom (1989) and more recently its distribution and functionals thereof have been computed by Groeneboom and Wellner (2001).

Note that the above result is reminiscent of the asymptotic normality of MLE's in regular parametric (and semiparametric) settings. Classical parametric theory provides very general theorems on the asymptotic normality of the MLE. If  $X_1, X_2, \dots, X_n$  are i.i.d. observations from a regular parametric model  $p(x, \theta)$ , and  $\hat{\theta}_n$  denotes the MLE of  $\theta$  based on this data, then  $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d (I(\theta))^{-1/2} Z$  where  $Z$  follows  $N(0, 1)$  and  $I(\theta)$  is the Fisher information. In monotone function estimate  $\mathbb{Z}$  plays the role of  $Z$ . A question that then naturally arises is whether one can find a unified framework in which pointwise MLE's of monotone functions exhibit an  $n^{1/3}$  rate of convergence to  $C\mathbb{Z}$  (i.e. (1.1) holds) and if possible identify the constant  $C$  in terms of some concrete aspect of the framework. A result of the type (1.1) permits the construction of confidence intervals for  $\psi(z)$  using the quantiles of  $\mathbb{Z}$  which are well-tabulated. The constant  $C(z)$  however needs to be estimated and often this involves nuisance parameters. One way to go about

estimating  $C(z)$  is to use a model-specific approach; in other words quantify  $C(z)$  explicitly as a function of the underlying model parameters and then search for ways to estimate these. A more attractive way to achieve this goal is through the use of subsampling techniques as developed in Politis, Romano and Wolf (1999); because of the slower convergence rate of estimators and lack of asymptotic normality the usual Efron-type bootstrap will not work in this situation. While the asymptotics for the various monotone function models considered above do exhibit fundamental commonalities, it does not seem that any sort of unification of these methods has been attempted in the literature.

A unified approach apart from being aesthetically attractive is also desirable from a practical standpoint as it precludes the need to work out different models on a case by case basis, as and when they arise. It can also provide a deeper understanding of the key features of these models. Our goal in this paper is to provide a unified perspective on pointwise estimation of monotone functions from the angle of likelihood based inference, with emphasis on likelihood ratios. We will be interested in testing a null hypothesis of the form  $\psi(z_0) = \theta_0$  and more generally null hypotheses where the monotone function is constrained at multiple points, using the likelihood ratio statistic. The motivation behind studying likelihood ratios comes from the experience with regular parametric and semiparametric models, where, under modest regularity conditions, likelihood ratio statistics for testing that the parameter of interest belongs to “nice” subsets of the parameter space, converge to a  $\chi^2$  distribution. This is a powerful result, since it enables one to construct confidence regions for the parameter by mere knowledge of the quantiles of the  $\chi^2$  distribution (through an inversion of the acceptance region of the likelihood ratio test), thereby obviating the need to estimate the information  $I(\theta)$  from the data. As in classical parametric statistics, is there some universal distribution that describes the limiting behavior of the likelihood ratio statistic for monotone function models? The existence of such a universal limit would facilitate the construction of confidence sets for the monotone function immensely as it would preclude the need to estimate the constant  $C(z)$  that appears in the limit distribution of the MLE. Furthermore, it is well known that likelihood ratio based confidence sets are more data-driven than MLE based ones and possess better finite sample properties in many different settings, so that one may expect to reap the same benefits for this class of problems. Note however that we do not expect a  $\chi^2$  distribution for the limiting likelihood ratio statistic for monotone functions; the  $\chi^2$  limit in regular settings is intimately connected with the  $\sqrt{n}$  rate of convergence of MLE’s to a Gaussian limit.

The hope that such a universal limit may exist is bolstered by the work of Banerjee and Wellner (2001), who studied the limiting behavior of the likelihood ratio statistic for testing the value of the distribution function of the survival time at a fixed point in the current status model. They found that in the limit, the likelihood ratio statistic behaves like  $\mathbb{D}$ , which is a well-defined functional of  $W(t) + t^2$  (and is described below). Here  $W(t)$  is standard two-sided Brownian motion starting from 0. Based on fundamental similarities among different monotone function models, Banerjee and Wellner (2001) conjectured that  $\mathbb{D}$  could be expected to arise as the limit distribution of likelihood ratios in several different models where (1.1) holds. Thus the relationship

of  $\mathbb{D}$  to  $\mathbb{Z}$  in the context of monotone function estimation would be analogous to that of the  $\chi_1^2$  distribution to  $N(0, 1)$  in the context of likelihood based inference in parametric models. Indeed  $\mathbb{D}$  would be a “non-regular” version of the  $\chi_1^2$  distribution. We will show that in a certain sense, this is indeed the case.

We are now in a position to describe the agenda for this paper. In Section 2, we develop a unified framework for estimating monotone functions in which the pointwise limit distribution of the MLE of the function is characterized (up to constants) by  $\mathbb{Z}$  and the pointwise likelihood ratio statistic converges to  $\mathbb{D}$ . We state and prove the main theorems describing the limit distributions of the MLE’s and the likelihood ratio statistic. As will be seen, our results apply in great generality and show that to a large extent, the similarities in the asymptotic behavior of the MLE in various different monotone function models are really instances of a very general phenomenon. In particular, the emergence of a fixed limit law for the pointwise likelihood ratio statistic has some of the flavor of Wilks’ classical result (Wilks (1938)) on the limiting  $\chi^2$  distribution of likelihood ratios in standard parametric models. Section 3 discusses applications of the main theorems and Section 4 contains concluding remarks. Section 5 contains the proofs of some of the lemmas used to establish the main results in Section 2, an alternative proof of Theorem 2.1 and is followed by references.

## 2 The Unified Framework

We now introduce our basic framework for investigating the asymptotic theory for monotone functions.

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  $\mu$ . 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  $X_i | Z_i = z \sim p(x, \psi(z))$ . Interest focuses on estimating the function  $\psi$ . We call these models *monotone response models*; given  $Z$ , which one may think of as a covariate, the parameter governing the distribution of the response  $X$  is monotone in the covariate. Models of this kind are frequently encountered in statistics. We provide below some motivating examples of the above scenario that have been fairly well-studied in the literature.

(a) **Monotone Regression Model:** Consider the model

$$X_i = \psi(Z_i) + \epsilon_i,$$

where  $\{(\epsilon_i, Z_i)\}_{i=1}^n$  are i.i.d. random variables,  $\epsilon_i$  is independent of  $Z_i$ , each  $\epsilon_i$  has mean 0 and variance  $\sigma^2$ , each  $Z_i$  has a Lebesgue density  $p_Z(\cdot)$  and  $\psi$  is a monotone function. The above model and its variants have been fairly well-studied in the literature on isotonic regression (see, for example, Brunk (1970), Wright(1981), Mukherjee (1991), Mammen (1991), Huang (2002)). Now suppose that the  $\epsilon_i$ ’s are Gaussian. We are then in the above framework:

$Z \sim p_Z(\cdot)$  and  $X | Z = z \sim N(\psi(z), \sigma^2)$ . We want to estimate  $\psi$  and test  $\psi(z_0) = \theta_0$  for an interior point  $z_0$  in the domain of  $\psi$ .

- (b) **Binary Choice Model:** Here we have a dichotomous response variable  $\Delta = 1$  or  $0$  and a continuous covariate  $Z$  with a Lebesgue density  $p_Z(\cdot)$  such that  $P(\Delta = 1 | Z) \equiv \psi(Z)$  is a smooth function of  $Z$ . Thus, conditional on  $Z$ ,  $X$  has a Bernoulli distribution with parameter  $\psi(Z)$ . Models of this kind have been quite broadly studied in econometrics and statistics (see, for example, Dunson (2004), Newton, Czado and Chappell (1996), Salanti and Ulm (2003)). In a biomedical context one could think of  $\Delta$  as representing the indicator of a disease/infection and  $Z$  the level of exposure to a toxin, or the measured level of a biomarker that is predictive of the disease/infection (see, for example, Ghosh, Banerjee and Biswas (2004)). In such cases it is often natural to impose a monotonicity assumption on  $\psi$ . As in (a), we want to make inference on  $\psi$ .
- (c) **Current Status Model:** The Current Status Model is used extensively in biomedical studies and epidemiological contexts and has received much attention among biostatisticians and statisticians (see, for example, Sun and Kalbfleisch (1993), Sun (1999), Shiboski (1998), Huang (1996), Banerjee and Wellner (2001)). Consider  $n$  individuals who are checked for infection at independent random times  $T_1, T_2, \dots, T_n$ ; we set  $\Delta_i = 1$  if individual  $i$  is infected by time  $T_i$  and  $0$  otherwise. We can think of  $\Delta_i$  as  $1\{X_i \leq T_i\}$  where  $X_i$  is the (random) time to infection (measured from some baseline period). The  $X_i$ 's are assumed to be independent and also independent of the  $T_i$ 's and are unknown. We are interested in making inference on the increasing function  $F$ , which is the distribution of the survival time for an individual. We note that  $\{\Delta_i, T_i\}_{i=1}^n$  is an i.i.d. sample from the distribution of  $(\Delta, T)$  where  $T \sim g(\cdot)$  for some Lebesgue density  $g$  and  $\Delta | T \sim \text{Bernoulli}(F(T))$ . This is precisely the model considered in (b)
- (d) **Poisson Regression Model:** Suppose that  $Z \sim p_Z(\cdot)$  and  $X | Z = z \sim \text{Poisson}(\psi(z))$  where  $\psi$  is a monotone function. We have  $n$  i.i.d. observations from this model. Here one can think of  $Z$  as the distribution of a region from a point source (for example, a nuclear processing plant) and  $X$  the number of cases of disease incidence at distance  $Z$ . Given  $Z = z$ , the number of cases of disease incidence  $X$  at distance  $z$  from the source is assumed to be a follow a Poisson distribution with mean  $\psi(z)$  where  $\psi$  can be expected to be monotonically decreasing in  $z$ . Variants of this model have received considerable attention in epidemiological contexts (Stone (1988), Diggle, Morris and Morton–Jones (1999), Morton–Jones, Diggle and Elliott (1999)).

We now return to the general monotone response model. Let  $z_0$  be an interior point of  $\tilde{I}$  at which one seeks to estimate  $\psi$ . Assume that

- (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 for testing the hypothesis  $H_0 : \psi(z_0) = \theta_0$ , where  $\theta_0$  is an interior point of  $\Theta$ . Denoting the likelihood ratio statistic by  $2 \log \lambda_n$ , we have

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

**Further Assumptions:** We now state 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  $x$ .
- (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  $x$ ,

$$\sup_{\theta \in \mathcal{N}} |l'''(x, \theta)| \leq B(x)$$

such that  $\sup_{\theta \in \mathcal{N}} E_\theta(B(X)) < \infty$ .

- (A.6) The functions:

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

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,

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

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 will also be seen to hold fairly easily for one-parameter exponential families.

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

We are interested in the describing the asymptotic behavior of the MLE's of  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$  in local neighborhoods of  $z_0$  and that of the likelihood ratio statistic  $2 \log \lambda_n$ . In order to do so, we first need to introduce the basic spaces and processes (and relevant functionals of the processes) that will figure in the asymptotic theory.

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

$$\int_{[-K, K]} (\phi_n(t) - \phi(t))^2 dt \rightarrow 0$$

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  $B_{loc}(\mathbb{R})$  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  $B_{loc}(\mathbb{R})$  if  $h_n$  and  $h$  are bounded on every compact interval  $[-K, K]$  ( $K > 0$ ) and  $\sup_{x \in [-K, K]} |h_n(x) - h(x)| \rightarrow 0$  for every  $K > 0$ .

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 and Wellner (2001) and Wellner (2001). Thus,  $g_{1,1}$  and  $g_{1,1}^0$  are the unconstrained and constrained versions of the slope processes associated with the canonical process  $X_{11,1}(h)$ . Finally define,

$$\mathbb{D} := \int ((g_{1,1}(z))^2 - (g_{1,1}^0(z))^2) dz.$$

Recall that  $\mathbb{Z}$  is the (almost surely unique) minimizer of  $W(h) + h^2$  over the line. We illustrate below an intimate connection between  $\mathbb{Z}$  and  $g_{1,1}(0)$ , the slope of the convex minorant of  $W(h) + h^2$  at the point 0. From the work of Groeneboom (1989) it follows that with probability 1

$$\operatorname{argmin}_h a W(h) + b h^2 - c h \geq \xi \Leftrightarrow g_{a,b}(\xi) \leq c.$$

This is the so-called ‘‘switching relationship’’ and will be found to play a crucial role later. Using the above display, we get:

$$P(g_{1,1}(0) \leq c) = P(\operatorname{argmin}_h W(h) + h^2 - c h \geq 0).$$

But

$$\operatorname{argmin}_h W(h) + h^2 - c h \equiv_d c/2 + \operatorname{argmin}_h (W(h) + h^2) = c/2 + \mathbb{Z}.$$

See Problem 3.2.5 of Van der Vaart and Wellner (1996) for a general version of the above result. It follows that

$$P(g_{1,1}(0) \leq c) = P(\mathbb{Z} \geq -c/2) = P(-\mathbb{Z} \leq c/2) = P(\mathbb{Z} \leq c/2),$$

using the fact that  $\mathbb{Z}$  is distributed symmetrically about 0. Thus  $g_{1,1}(0) \equiv_d 2\mathbb{Z}$ . We will have occasion to refer to this result shortly.

The following theorem describes the limiting behavior of the unconstrained and constrained MLE’s of  $\psi$ , appropriately normalized.

**Theorem 2.1** *Let,*

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

*Let*

$$a = \sqrt{\frac{1}{I(\psi(z_0)) p_Z(z_0)}} \quad \text{and} \quad b = \frac{\psi'(z_0)}{2}.$$

*Then  $(X_n(h), Y_n(h)) \rightarrow_d (g_{a,b}(h), g_{a,b}^0(h))$  finite dimensionally and also in the space  $\mathcal{L} \times \mathcal{L}$ .*

Thus,

$$X_n(0) = n^{1/3} (\hat{\psi}_n(z_0) - \psi(z_0)) \rightarrow g_{a,b}(0).$$

Using Brownian scaling it follows that the following distributional equality holds in the space  $\mathcal{L} \times \mathcal{L}$ :

$$(g_{a,b}(h), g_{a,b}^0(h)) =_d \left( a (b/a)^{1/3} g_{1,1} \left( (b/a)^{2/3} h \right), a (b/a)^{1/3} g_{1,1}^0 \left( (b/a)^{2/3} h \right) \right). \quad (2.2)$$

For a proof of this proposition, see for example, Banerjee (2000). Using the fact that  $g_{1,1}(0) \equiv_d 2\mathbb{Z}$ , we get:

$$n^{1/3} (\hat{\psi}_n(z_0) - \psi(z_0)) \rightarrow_d a (b/a)^{1/3} g_{1,1}(0) \equiv_d (8 a^2 b)^{1/3} \mathbb{Z}.$$

This is precisely the phenomenon described in (1.1).

Our next theorem concerns the limit distribution of the likelihood ratio statistic for testing  $H_0 : \psi(z_0) = \theta_0$  when the null distribution is true.

**Theorem 2.2** Under assumptions A.1 – A.7 and (a), (b),

$$2 \log \lambda_n \rightarrow_d \mathbb{D},$$

when  $H_0$  is true.

**Comments:** In this paper we work under the assumption that  $Z$  has a Lebesgue density on its support. However, Theorems 2.1 and 2.2, the main results of this paper, continue to hold under the weaker assumption that the distribution function of  $Z$  is continuously differentiable (and hence has a Lebesgue density) in a neighborhood of  $z_0$  with non-vanishing derivative at  $z_0$ . Also, subsequently we tacitly assume that MLE's always exist which is stronger than the assumption following (A.7). This makes the presentation simpler without compromising the ideas involved. In this paper, we will focus on the case when  $\psi$  is increasing. The case where  $\psi$  is decreasing is incorporated into this framework by replacing  $Z$  by  $-Z$  and considering the (increasing) function  $\bar{\psi}(z) = \psi(-z)$ .

In order to study the asymptotic properties of the statistics of interest it is necessary to characterize the MLE's  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$  and this is what we do below.

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

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

We now discuss the characterization of the maximum likelihood estimators of  $\psi$ . We first write down the log-likelihood function for the data. This is given by:

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

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

$$\tilde{\psi}(u_1, u_2, \dots, u_n) = \sum_{i=1}^n \phi(X_{(i)}, u_i),$$

over all  $u_1 \leq u_2 \leq \dots \leq u_n$ . Once, we obtain the (unique) minimizer  $\hat{u} \equiv (\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n)$ , the MLE,  $\hat{\psi}_n$  is given by:  $\hat{\psi}_n(Z_{(i)}) = \hat{u}_i$  for  $i = 1, 2, \dots, n$ .

By our assumptions,  $\tilde{\psi}$  is a (continuous) convex function defined on  $\mathbb{R}^n$  and assuming values in  $\mathbb{R} \cup \{\infty\}$ . Let  $R = \tilde{\psi}^{-1}(\mathbb{R})$ . From the continuity and convexity of  $\tilde{\psi}$ , it follows that  $R$  is an open convex subset of  $\mathbb{R}^n$ . We will minimize  $\tilde{\psi}(u)$  over  $R$  subject to the constraints  $u_1 \leq u_2 \leq \dots \leq u_n$ . Note that  $\tilde{\psi}$  is finite and differentiable on  $\mathbb{R}$ . Necessary and sufficient conditions characterizing

the minimizer are obtained readily, using the Kuhn-Tucker theorem. We write the constraints as  $g(u) \leq 0$  where  $g(u) = (g_1(u), g_2(u), \dots, g_{n-1}(u))$  and

$$g_i(u) = u_i - u_{i+1}, \quad i = 1, 2, \dots, n-1.$$

Then, there exists an  $n-1$  dimensional vector  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{n-1})$  with  $\lambda_i \geq 0$  for all  $i$ , such that, if  $\hat{u}$  is the minimizer in  $R$ , satisfying the constraints,  $g(\hat{u}) \leq 0$ , then,

$$\sum_{i=1}^n \lambda_i (\hat{u}_{i-1} - \hat{u}_i) = 0,$$

and furthermore,

$$\nabla \psi(\hat{u}) + G^T \lambda = 0,$$

where  $G$  is the  $(n-1) \times (n)$  matrix of partial derivatives of  $g$ . The conditions displayed above are often referred to as *Fenchel conditions*. In this case, with  $\lambda_0 \equiv 0$ , the second condition boils down to:

$$\nabla_i \tilde{\psi}(\hat{u}) + (\lambda_i - \lambda_{i-1}) = 0, \quad \text{for } i = 1, 2, \dots, n-1,$$

and

$$\nabla_n \tilde{\psi}(\hat{u}) - \lambda_{n-1} = 0.$$

Solving recursively to obtain the  $\lambda_i$ 's (for  $i = 1, 2, \dots, n-1$ ), we get

$$\lambda_i \equiv \sum_{j=i+1}^n \nabla_j \tilde{\psi}(\hat{u}) \geq 0, \quad \text{for } i = 1, 2, \dots, (n-1) \quad (2.3)$$

and

$$\sum_{j=1}^n \nabla_j \tilde{\psi}(\hat{u}) = 0. \quad (2.4)$$

Now, let  $B_1, B_2, \dots, B_k$  be the blocks of indices on which the solution  $\hat{u}$  is constant and let  $w_j$  be the common value on block  $B_j$ . The equality:  $\sum_{i=1}^n \lambda_i (\hat{u}_{i-1} - \hat{u}_i) = 0$  forces  $\lambda_i = 0$  whenever  $\hat{u}_{i-1} < \hat{u}_i$ . Noting that  $\nabla_r \psi(\hat{u}) = \phi'(X_{(r)}, \hat{u}_r)$ , this implies that on each  $B_j$ :

$$\sum_{r \in B_j} \phi'(X_{(r)}, w_j) = 0.$$

Thus  $w_j$  is the unique solution to the equation

$$\sum_{r \in B_j} \phi'(X_{(r)}, w) = 0. \quad (2.5)$$

The solution  $\hat{u}$  can be characterized as the vector of left derivatives of the greatest convex minorant (GCM) of a (random) cumulative sum (cusum) diagram, as will be shown below. In the general formulation, the cusum diagram will itself be characterized in terms of the solution  $\hat{u}$ , giving us,

what is called a “self-consistent” characterization. However, under sufficient structure on the underlying parametric model (for example, one parameter exponential families under the natural parametrization), the self-consistent characterization can be avoided and the MLE is characterized as the slope of the convex minorant of a cusum diagram that can be explicitly computed from the data. In such cases the asymptotics for the limit distribution of the MLE are somewhat simpler. A special case is dealt with in an earlier paper (Banerjee and Wellner (2001)). In our current general setting this will not be possible and the “self-consistent” conditions are *crucial* to understanding the long run behavior of the MLE.

Before proceeding further, we introduce some notation. For points  $\{(x_0, y_0), (x_1, y_1), \dots, (x_k, y_k)\}$  where  $x_0 = y_0 = 0$  and  $x_0 < x_1 < \dots < x_k$ , 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  $(x_1, x_2, \dots, x_n)$  by  $\text{slogcm} \{(x_i, y_i)\}_{i=0}^n$ .

Here is the idea behind the self-consistency characterization: Suppose that  $\hat{u}$  is the (unique) minimizer of  $\tilde{\psi}$  over the region  $R$  subject to the given constraints  $u_1 \leq u_2 \leq \dots \leq u_n$ . Consider now, the following (quadratic) function,

$$\xi(u) = \frac{1}{2} \left[ u - \hat{u} + \mathcal{K}^{-1} \nabla \tilde{\psi}(\hat{u}) \right]^T \mathcal{K} \left[ u - \hat{u} + \mathcal{K}^{-1} \nabla \tilde{\psi}(\hat{u}) \right],$$

where  $\mathcal{K}$  is some positive definite matrix. Note that  $\text{Hess}(\xi) = \mathcal{K}$  which is positive definite; thus  $\xi$  is a strictly convex function. It is also finite and differentiable over  $\mathbb{R}^n$ . Also,

$$\nabla \xi(u) = \mathcal{K} \left( u - \hat{u} + \mathcal{K}^{-1} \nabla \tilde{\psi}(\hat{u}) \right).$$

Now, consider the problem of minimizing  $\xi$  over  $\mathbb{R}^n$ , subject to the constraints:  $g(u) \leq 0$ . If  $u^*$  is the global minimizer, then necessary and sufficient conditions are given by conditions (2.3) (for  $i = 1, 2, \dots, n-1$ ) and (2.4), with  $\tilde{\psi}$  replaced by  $\xi$  and  $\hat{u}$  replaced by  $u^*$ . Now,  $\nabla \xi(\hat{u}) = \nabla \tilde{\psi}(\hat{u})$ , so that  $u^* = \hat{u}$  does indeed satisfy the conditions (2.3) (for  $i = 1, 2, \dots, n-1$ ) and (2.4), with  $\tilde{\psi}$  replaced by  $\xi$ . Also,  $u^*$  is the unique minimizer of  $\xi$  subject to the (convex) constraints  $g(u) \leq 0$  by virtue of the fact that the Hessian of  $\xi$  is always positive definite. A formal proof could be devised in the following manner: Suppose there exists  $u^{**}$  different from  $u^*$  satisfying the constraints and minimizing  $\xi$ . Since  $\xi$  is convex,

$$\xi(\lambda u^* + (1 - \lambda) u^{**}) \leq \lambda \xi(u^*) + (1 - \lambda) \xi(u^{**}) = \xi(u^*) = \xi(u^{**}),$$

for any  $0 \leq \lambda \leq 1$ . On the other hand, for any  $0 \leq \lambda \leq 1$ ,

$$\xi(\lambda u^* + (1 - \lambda) u^{**}) \geq \xi(u^*).$$

This implies

$$r(\lambda) = \xi(\lambda u^* + (1 - \lambda) u^{**}) = \xi(u^*) \text{ for all } \lambda \in [0, 1].$$

Hence, for any  $0 < \lambda < 1$ ,  $r''(\lambda) = 0$ . But

$$r''(\lambda) = (u^* - u^{**})^T \mathcal{K} (u^* - u^{**}) > 0,$$

since  $u^* - u^{**} \neq 0$  by supposition and  $\mathcal{K}$  is positive definite. This gives a contradiction.

It now suffices to try to minimize  $\xi$ ; of course the problem here is that  $\hat{u}$  is unknown and  $\xi$  is defined in terms of  $\hat{u}$ . However, an iterative scheme can be developed along the following lines. Choosing  $\mathcal{K}$  to be a diagonal matrix with the  $i, i$ 'th entry being  $d_i \equiv \nabla_{ii} \tilde{\psi}(\hat{u})$  ( $\mathcal{K}$  thus defined is a p.d. matrix, since the diagonal entries of the Hessian of  $\psi$  at the minimizer  $\hat{u}$ , which is a positive definite matrix are positive), we see that the above quadratic form reduces to,

$$\begin{aligned} \xi(u) &= \sum_{i=1}^n \left[ u_i - \hat{u}_i + \nabla_i \tilde{\psi}(\hat{u}) d_i^{-1} \right]^2 d_i \\ &= \sum_{i=1}^n \left[ u_i - \left( \hat{u}_i - \nabla_i \tilde{\psi}(\hat{u}) d_i^{-1} \right) \right]^2 d_i. \end{aligned}$$

Thus  $\hat{u}$  minimizes,

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

subject to the constraints that  $u_1 \leq u_2 \leq \dots \leq u_n$  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  $\{1, 2, \dots, n\}$  with weight function  $d_i$ . It is well known that the solution

$$(\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n) = \text{slogcm} \left\{ \sum_{j=1}^i d_i, \sum_{j=1}^i g(j) d_i \right\}_{i=0}^n.$$

See, for example Theorem 1.2.1 of Robertson, Wright and Dykstra (1988). In terms of the function  $\phi$  the solution can be written as:

$$(\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n) \equiv \left[ \text{slogcm} \left\{ \sum_{j=1}^i \phi''(X_{(j)}, \hat{u}_j), \sum_{j=1}^i \hat{u}_j \phi''(X_{(j)}, \hat{u}_j) - \phi'(X_{(i)}, \hat{u}_i) \right\}_{i=0}^n \right]. \quad (2.6)$$

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.

Since  $\hat{u}$  is unknown, we need to iterate. Thus, we pick an initial guess for  $\hat{u}$ , say  $u^{(0)}$  (belonging to  $R$ ) and satisfying the constraints imposed by  $g$ , compute  $u^{(1)}$  using the recipe (2.6), plug in  $u^{(1)}$  as an updated guess for  $\hat{u}$ , obtain  $u^{(2)}$  and proceed thus, till convergence (this can be checked for example by testing whether the conditions (2.3) and (2.4) are nearly satisfied).

Note however that  $\hat{u}$  belongs to the (maximal) convex region on which  $\tilde{\psi}$  is finite and the

iteration procedure does not guarantee that we will not move outside  $R$ , in which case  $\tilde{\psi}$  would hit  $\infty$  and fail to be differentiable. Hence, a detection procedure needs to be installed to keep the iteration in the right track. Jongbloed (1998) deals with a modified version of the iterative convex minorant (ICM) algorithm discussed above that is guaranteed to converge to  $\hat{u}$ .

**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  $u_1 \leq u_2 \leq \dots \leq u_m \leq \theta_0 \leq u_{m+1} \leq \dots \leq u_n$ . This can be reduced to solving two separate optimization problems. These are:

$$(1) \quad \text{Minimize } \sum_{i=1}^m \phi(X_{(i)}, u_i) \text{ over } u_1 \leq u_2 \leq \dots \leq u_m \leq \theta_0.$$

$$(2) \quad \text{Minimize } \sum_{i=m+1}^n \phi(X_{(i)}, u_i) \text{ over } \theta_0 \leq u_{m+1} \leq u_{m+2} \leq \dots \leq u_n.$$

Consider (1) first. As in the unconstrained minimization problem one can write down the Kuhn–Tucker conditions characterizing the minimizer. It is then easy to see that the solution  $(\hat{u}_1^0, \hat{u}_2^0, \dots, \hat{u}_m^0)$  can be obtained through the following recipe : Minimize  $\sum_{i=1}^m \phi'(X_{(i)}, u_i)$  over  $u_1 \leq u_2 \leq \dots \leq u_m$  to get  $(\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_m)$ . Then,

$$(\hat{u}_1^0, \hat{u}_2^0, \dots, \hat{u}_m^0) = (\tilde{u}_1 \wedge \theta_0, \tilde{u}_2 \wedge \theta_0, \dots, \tilde{u}_m \wedge \theta_0).$$

The solution vector to (2), say  $(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \dots, \hat{u}_n^0)$  is similarly given by

$$(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \dots, \hat{u}_n^0) = (\tilde{u}_{m+1} \vee \theta_0, \tilde{u}_{m+2} \vee \theta_0, \dots, \tilde{u}_n \vee \theta_0),$$

where

$$(\tilde{u}_{m+1}, \tilde{u}_{m+2}, \dots, \tilde{u}_n) = \operatorname{argmin}_{u_{m+1} \leq u_{m+2} \leq \dots \leq u_n} \sum_{i=m+1}^n \phi'(X_{(i)}, u_i).$$

It will be useful to examine how the constrained solution and the unconstrained solution relate to one another. With  $B_1, B_2, \dots, B_k$  denoting the blocks for the unconstrained solution, as before, let  $B_l = [l_1, l_2]$  be the block of indices containing  $m$ . Let  $B_{l,1}$  denote the block  $[l_1, m]$  and  $B_{l,2}$  the block  $[m+1, l_2]$ . The blocks for the vector  $(\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_m)$  will then be given by  $(B_1, B_2, \dots, B_{l-1}, C_1, C_2, \dots, C_r)$  where  $\cup_{i=1}^r C_i = B_{l,1}$ . Also, if  $w_{c,j}$  denotes the common value of the solution on the block  $C_j$ , we have  $w_l \leq w_{c,1} < w_{c,2} < \dots < w_{c,r}$ . Similarly, the blocks for the vector  $(\tilde{u}_{m+1}, \tilde{u}_{m+2}, \dots, \tilde{u}_n)$  will be given by  $(D_1, D_2, \dots, D_s, B_{l+1}, B_{l+2}, \dots, B_n)$  where  $\cup_{i=1}^s D_i = B_{l,2}$ . Also, if  $w_{d,j}$  denotes the common value of the solution on the block  $D_j$ , we have  $w_{d,1} < w_{d,2} \leq \dots \leq w_{d,s} \leq w_l$ .

Now suppose that the unconstrained solution exceeds  $\theta_0$  for some block  $B_i$  with  $i \leq l$ . Then for  $j \leq m$  the constrained solution will agree with the unconstrained solution on blocks  $B_1$  through  $B_{i-1}$  and will be identically equal to  $\theta_0$  on the blocks  $B_i, \dots, B_{l-1}, C_1, \dots, C_r$ . In this case, for  $j > m$ , the unconstrained solution will agree with the constrained solution on blocks  $B_{l+1}, B_{l+2}, \dots, B_k$ . On blocks  $D_1$  through  $D_s$  the constrained solution will be given by the vector  $(w_{d,1} < w_{d,2} \dots < w_{d,s}) \vee \theta_0$ . Thus, the solution can potentially be equal to  $w_l$  on the block  $D_s$  (if  $w_{d,s} = w_l$ ) in which case the set on which the unconstrained and constrained solutions differ will correspond to the set of indices  $\cup_{j=i}^l B_j - D_s$ . If  $w_{d,s} < w_l$ , then the difference set is simply  $\cup_{j=i}^l B_j$ .

A similar characterization can be given when the unconstrained solution only exceeds  $\theta_0$  for some block  $B_i$  with  $i > m$ . In either case, we can conclude that the following holds:

$$\hat{\psi}_n(z) \neq \hat{\psi}_n^0(z) \Rightarrow \hat{\psi}_n^0(z_0) = \theta_0 \text{ or } \hat{\psi}_n(z) = \hat{\psi}_n(z_0). \quad (2.7)$$

An important property of the constrained solution is that on any block  $B$  of indices where it is constant and not equal to  $\theta_0$ , the constant value, say  $w_B$  is the unique solution to the equation:

$$\sum_{i \in B} \phi'(X_{(i)}, w) = 0. \quad (2.8)$$

The constrained solution also has a “self-consistent” characterization in terms of the slope of the greatest convex minorant of a cumulative sum diagram. This follows in the same way as for the unconstrained solution by using the Kuhn–Tucker theorem and formulating a quadratic optimization problem based on the Fenchel conditions arising from this theorem. We skip the details but give the self-consistent characterization.

The constrained solution  $\hat{u}^0$  minimizes,

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

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

$$(\hat{u}_1^0, \hat{u}_2^0, \dots, \hat{u}_m^0) \equiv \left[ \text{slogcm} \left\{ \sum_{j=1}^i \phi''(X_{(i)}, \hat{u}_i^0), \sum_{j=1}^i \hat{u}_i^0 \phi''(X_{(i)}, \hat{u}_i^0) - \phi'(X_{(i)}, \hat{u}_i^0) \right\}_{i=0}^m \right] \wedge \theta_0, \quad (2.9)$$

and

$$(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \dots, \hat{u}_n^0) \equiv \left[ \text{slogcm} \left\{ \sum_{j=m+1}^i \phi''(X_{(i)}, \hat{u}_i^0), \sum_{j=1}^i \hat{u}_i^0 \phi''(X_{(i)}, \hat{u}_i^0) - \phi'(X_{(i)}, \hat{u}_i^0) \right\}_{i=m}^n \right] \vee \theta_0. \quad (2.10)$$

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  $i = 1, 2, \dots, n$ ,  $\hat{\psi}_n^0(z_0) = \theta_0$  and having no jump points outside the set

$\{Z_{(i)}\}_{i=1}^n \cup \{z_0\}$ .

The MLE's  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$  are uniformly consistent for  $\psi$  in a closed neighborhood of  $z_0$ . This is the content of the following lemma.

**Lemma 2.1** *There exists a neighborhood  $[\sigma, \tau]$  of  $z_0$  such that*

$$\sup_{z \in [\sigma, \tau]} \left| \hat{\psi}_n(z) - \psi(z) \right| \rightarrow_{a.s.} 0,$$

and

$$\sup_{z \in [\sigma, \tau]} \left| \hat{\psi}_n^0(z) - \psi(z) \right| \rightarrow_{a.s.} 0.$$

This lemma can be proved using arguments similar to those in the proof of Theorem 3.2 of Huang (1996). For the purposes of deducing the limit distribution of the MLE's and the likelihood ratio statistic we are more interested in the following lemma that guarantees local consistency at an appropriate rate.

**Lemma 2.2** *For any  $M_0 > 0$ , we have:*

$$\sup_{h \in [-M_0, M_0]} \left| \hat{\psi}_n(z_0 + h n^{-1/3}) - \psi(z_0) \right| = O_p(n^{-1/3}),$$

and

$$\sup_{h \in [-M_0, M_0]} \left| \hat{\psi}_n^0(z_0 + h n^{-1/3}) - \psi(z_0) \right| = O_p(n^{-1/3}).$$

We next state a number of preparatory lemmas required in the proofs of Theorems 2.1 and 2.2. But before that we need to introduce further notation. For a monotone function  $\Lambda$  taking values in  $\Theta$  define the following processes:

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

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

and

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

We will denote by  $W_n, G_n, B_n$  the above processes when  $\Lambda = \hat{\psi}_n$ , and by  $W_{n,0}, G_{n,0}, B_{n,0}$  the above processes when  $\Lambda = \hat{\psi}_n^0$ . Also, define normalized processes  $\tilde{B}_{n,\Lambda}(h)$  and  $\tilde{G}_{n,\Lambda}(h)$  in the following manner:

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

and

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

**Lemma 2.3** Consider the process:

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

Then  $\tilde{B}_{n,\Psi}(h) \rightarrow_d X_{a,b}(h)$  in the space  $B_{loc}(\mathbb{R})$ , where,

$$a = \sqrt{\frac{1}{I(\psi(z_0)) p_Z(z_0)}} \quad \text{and} \quad b = \frac{\psi'(z_0)}{2}.$$

**Lemma 2.4** For every  $K > 0$ , the following asymptotic equivalences hold:

$$\sup_{h \in [-K, K]} \left| \tilde{B}_{n,\psi}(h) - \tilde{B}_{n,\hat{\psi}_n}(h) \right| \rightarrow_p 0,$$

and

$$\sup_{h \in [-K, K]} \left| \tilde{B}_{n,\psi}(h) - \tilde{B}_{n,\hat{\psi}_n^0}(h) \right| \rightarrow_p 0.$$

**Lemma 2.5** The processes  $\tilde{G}_{n,\hat{\psi}_n}(h)$  and  $\tilde{G}_{n,\hat{\psi}_n^0}(h)$  both converge uniformly (in probability) to the deterministic function  $h$  on the compact interval  $[-K, K]$ .

The next lemma characterizes the set  $D_n$  on which  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$  vary.

**Lemma 2.6** Given any  $\epsilon > 0$  we can find an  $M > 0$  such that for all sufficiently large  $n$ ,

$$P(D_n \subset [z_0 - M n^{-1/3}, z_0 + M n^{-1/3}]) \geq 1 - \epsilon.$$

For proofs of Lemmas 2.2, 2.3, 2.4 and 2.6, see Section 5.

**Proof of Theorem 2.1:** There are two different ways of establishing this result. The first of these that we sketch here relies on continuous mapping for slopes of greatest convex minorant estimators and is intuitively more appealing. The second uses an artifice originally introduced by Groeneboom and is included in the appendix. Which route one desires to take is largely a matter of taste.

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 (2.6). Consequently,

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

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 *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:

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

Let  $h \equiv n^{1/3}(z - z_0)$  be the local variable and recall the normalized processes that were defined before the statement of Lemma 2.3. In terms of the local variable and the normalized processes, it is not difficult to see that:

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

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 (2.9) and (2.10)) and the definitions of the normalized processes it follows that:

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

Thus,

$$(X_n(h), Y_n(h)) = \left\{ \text{slogcm} \left( \tilde{B}_{n, \hat{\psi}_n} \cdot \tilde{G}_{n, \hat{\psi}_n}^{-1} \right) \left( \tilde{G}_{n, \hat{\psi}_n}(h) \right), \text{slogcm}^0 \left( \tilde{B}_{n, \hat{\psi}_n^0} \cdot \tilde{G}_{n, \hat{\psi}_n^0}^{-1} \right) \left( \tilde{G}_{n, \hat{\psi}_n^0}(h) \right) \right\}. \quad (2.11)$$

By Lemma 2.4, 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. Furthermore, by Lemma 2.3, the process  $\tilde{B}_{n, \psi_0}(h)$  converges to the process  $X_{a,b}(h)$  in  $B_{loc}(\mathbb{R})$ . It follows that the processes

$$(\tilde{B}_{n, \hat{\psi}_n^0}(h), \tilde{B}_{n, \hat{\psi}_n}(h)) \rightarrow_d (X_{a,b}(h), X_{a,b}(h)),$$

in the space  $B_{loc}(\mathbb{R}) \times B_{loc}(\mathbb{R})$ , equipped with the product topology. Furthermore, by Lemma 2.5, the processes

$$(\tilde{G}_{n, \hat{\psi}_n}(h), \tilde{G}_{n, \hat{\psi}_n^0}(h)) \rightarrow_p (h, h)$$

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 (2.11) by their limits. It follows that for any  $(h_1, h_2, \dots, h_k)$ ,

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

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 and Zhang (1994). If  $\phi_n, \phi$  are monotone functions such that

$$\phi_n(t) \mid (t_1, t_2, \dots, t_k) \rightarrow \phi(t) \mid (t_1, t_2, \dots, t_k)$$

for every  $t_1 < t_2 < \dots < t_k$ , then  $\phi_n$  converges to  $\phi$  in the  $L_2$  sense, on every compact set.  $\square$

**Proof of Theorem 2.2:** We have,

$$\begin{aligned} 2 \log \lambda_n &= -2 \left[ \sum_{i=1}^n \phi(X_{(i)}, \hat{\psi}_n(Z_{(i)})) - \sum_{i=1}^n \phi(X_{(i)}, \hat{\psi}_n^0(Z_{(i)})) \right] \\ &= -2 \left[ \sum_{i \in J_n} \phi(X_{(i)}, \hat{\psi}_n(Z_{(i)})) - \sum_{i \in J_n} \phi(X_{(i)}, \hat{\psi}_n^0(Z_{(i)})) \right] \\ &= -2 S_n. \end{aligned}$$

Here  $J_n$  is the set of indices for which  $\hat{\psi}_n(Z_{(i)})$  and  $\hat{\psi}_n^0(Z_{(i)})$  are different. By Taylor expansion about  $\psi(z_0)$  we can write:

$$\begin{aligned} S_n &= \left[ \sum_{i \in J_n} \phi'(X_{(i)}, \psi(z_0)) (\hat{\psi}_n(Z_{(i)}) - \psi(z_0)) + \sum_{i \in J_n} \frac{\phi''(X_{(i)}, \psi(z_0))}{2} (\hat{\psi}_n(Z_{(i)}) - \psi(z_0))^2 \right] \\ &\quad - \left[ \sum_{i \in J_n} \phi'(X_{(i)}, \psi(z_0)) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0)) + \sum_{i \in J_n} \frac{\phi''(X_{(i)}, \psi(z_0))}{2} (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2 \right] + R_n \end{aligned}$$

with  $R_n = R_{n,1} - R_{n,2}$ , where

$$R_{n,1} = \frac{1}{6} \sum_{i \in J_n} \phi'''(X_{(i)}, \psi_{n,i}^*) (\hat{\psi}_n(Z_{(i)}) - \psi(z_0))^3 \quad \text{and} \quad R_{n,2} = \frac{1}{6} \sum_{i \in J_n} \phi'''(X_{(i)}, \psi_{n,i}^{**}) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^3,$$

for points  $\psi_{n,i}^*$  (lying between  $\hat{\psi}_n(Z_{(i)})$  and  $\psi(z_0)$ ) and  $\psi_{n,i}^{**}$  (lying between  $\hat{\psi}_n^0(Z_{(i)})$  and  $\psi(z_0)$ ). Under our assumptions  $R_n$  is  $o_p(1)$  as will be established later. Thus, we can write:

$$S_n = I_n + II_n + o_p(1)$$

where

$$\begin{aligned} I_n &= \sum_{i \in J_n} \phi'(X_{(i)}, \psi(z_0)) (\hat{\psi}_n(Z_{(i)}) - \psi(z_0)) - \sum_{i \in J_n} \phi'(X_{(i)}, \psi(z_0)) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0)) \\ &\equiv I_{n,1} - I_{n,2}, \end{aligned}$$

and

$$II_n = \sum_{i \in J_n} \frac{\phi''(X_{(i)}, \psi(z_0))}{2} (\hat{\psi}_n(Z_{(i)}) - \psi(z_0))^2 - \sum_{i \in J_n} \frac{\phi''(X_{(i)}, \psi(z_0))}{2} (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2.$$

Consider the term  $I_{n,2}$ . Now,  $J_n$  can be written as the union of blocks of indices, say  $B_1^0, B_2^0, \dots, B_l^0$ , such that the constrained solution  $\hat{\psi}_n^0$  is constant on each of these blocks. Let  $B$  denote a typical block and let  $w_B^0$  denote the constant value of the constrained MLE on this block; thus  $\hat{\psi}_n^0(Z_{(j)}) = w_B^0$  for each  $j \in B$ . Also, on each block  $B$  where  $w_B^0 \neq \theta_0$ , we have,

$$\sum_{j \in B} \phi'(X_{(j)}, w_B^0) = 0,$$

from (2.8). Thus, for any block  $B$  where  $w_B^0 \neq \theta_0$  we have

$$\begin{aligned} \sum_{i \in B} \phi'(X_{(i)}, \psi(z_0)) (w_B^0 - \psi(z_0)) &= \sum_{i \in B} [\phi'(X_{(i)}, w_B^0) + (\psi(z_0) - w_B^0) \phi''(X_{(i)}, w_B^0) \\ &= \quad + \frac{1}{2} (\psi(z_0) - w_B^0)^2 \phi'''(X_{(i)}, w_B^{0,*})] (w_B^0 - \psi(z_0)) \\ &= - \sum_{i \in B} (\psi(z_0) - w_B^0)^2 \phi''(X_{(i)}, w_B^0) \\ &\quad - \frac{1}{2} \sum_{i \in B} (\psi(z_0) - w_B^0)^3 \phi'''(X_{(i)}, w_B^{0,*}), \end{aligned}$$

where  $w_B^{0,*}$  is once again a point between  $w_B^0$  and  $\psi(z_0)$ . We conclude that,

$$I_{n,2} = - \sum_{i \in J_n} \phi''(X_{(i)}, \hat{\psi}_n^0(Z_{(i)})) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2 - \frac{1}{2} \sum_{i \in J_n} \phi'''(X_{(i)}, \hat{\psi}_n^{0,*}(Z_{(i)})) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^3,$$

where  $\hat{\psi}_n^{0,*}(Z_{(i)})$  is a point between  $\hat{\psi}_n^0(Z_{(i)})$  and  $\psi(z_0)$ . The second term in the above display is shown to be  $o_p(1)$  by the exact same reasoning as used for  $R_{n,1}$  or  $R_{n,2}$ . Hence,

$$\begin{aligned} I_{n,2} &= - \sum_{i \in J_n} \phi''(X_{(i)}, \hat{\psi}_n^0(Z_{(i)})) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2 + o_p(1) \\ &= - \sum_{i \in J_n} \phi''(X_{(i)}, \psi(z_0)) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2 + o_p(1), \end{aligned}$$

this last step following from a one-step Taylor expansion about  $\psi(z_0)$ . Similarly,

$$I_{n,1} = - \sum_{i \in J_n} \phi''(X_{(i)}, \psi(z_0)) (\hat{\psi}_n(Z_{(i)}) - \psi(z_0))^2 + o_p(1).$$

Now, using the fact that  $S_n = I_n + II_n + o_p(1) \equiv I_{n,1} - I_{n,2} + II_n + o_p(1)$  and using the representations for these terms derived above, we get:

$$S_n = -\frac{1}{2} \left\{ \sum_{i \in J_n} \phi''(X_{(i)}, \psi(z_0)) (\hat{\psi}_n(Z_{(i)}) - \psi(z_0))^2 - \sum_{i \in J_n} \phi''(X_{(i)}, \psi(z_0)) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2 \right\} + o_p(1)$$

whence

$$\begin{aligned}
2 \log \lambda_n &= \sum_{i \in J_n} \phi''(X_{(i)}, \psi(z_0)) (\hat{\psi}_n(Z_{(i)}) - \psi(z_0))^2 - \sum_{i \in J_n} \phi''(X_{(i)}, \psi(z_0)) (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0))^2 + o_p(1) \\
&= n^{1/3} \mathbb{P}_n \left[ \phi''(X, \psi(z_0)) \left\{ (n^{1/3} (\hat{\psi}_n(Z_{(i)}) - \psi(z_0)))^2 - (n^{1/3} (\hat{\psi}_n^0(Z_{(i)}) - \psi(z_0)))^2 \right\} 1(Z \in D_n) \right] \\
&\quad + o_p(1) \\
&= n^{1/3} (\mathbb{P}_n - P) \xi_n(X, Z) + n^{1/3} P \xi_n(X, Z) + o_p(1),
\end{aligned}$$

where  $\xi_n(X, Z)$  is the random function given by:

$$\xi_n(X, Z) = \phi''(X, \psi(z_0)) \left\{ (n^{1/3} (\hat{\psi}_n(Z) - \psi(z_0)))^2 - (n^{1/3} (\hat{\psi}_n^0(Z) - \psi(z_0)))^2 \right\} 1(Z \in D_n).$$

The term  $n^{1/3} (\mathbb{P}_n - P) \xi_n(X, Z) \rightarrow_p 0$ ; this is deduced using standard preservation properties for Donsker classes of functions. We outline the argument when  $\phi''(X, \psi(z_0))$  for example is a bounded function. With arbitrarily high probability  $D_n$  is eventually contained in an interval of the form  $[z_0 - M n^{-1/3}, z_0 + M n^{-1/3}]$  (by Lemma 2.6) and the monotone functions  $n^{1/3} (\hat{\psi}_n(z) - \psi(z_0))$  and  $n^{1/3} (\hat{\psi}_n^0(z) - \psi(z_0))$  are  $O_p(1)$  on sets of the form  $[z_0 - M n^{-1/3}, z_0 + M n^{-1/3}]$ . Next, note that any class of uniformly bounded monotone functions on a compact set is universally Donsker, that indicator functions of intervals on the line form a bounded Donsker class, that the Donsker property is preserved on squaring a bounded Donsker class, taking differences of bounded Donsker classes, on forming pairwise products of uniformly bounded Donsker classes and finally on multiplying a Donsker class by a fixed bounded function. As a consequence of the above facts, the function  $\xi_n(X, Z)$  is eventually contained in a uniformly bounded Donsker class of functions with arbitrarily high probability; hence  $\sqrt{n} (\mathbb{P}_n - P) \xi_n(X, Z)$  is  $O_p(1)$ , implying that  $n^{1/3} (\mathbb{P}_n - P) \xi_n(X, Z) \rightarrow_p 0$ . It now remains to deal with the term  $n^{1/3} P(\xi_n(X, Z))$  and as we shall see, it is this term that contributes to the likelihood ratio statistic in the limit. Thus,

$$2 \log \lambda_n = n^{1/3} P(\xi_n(X, Z)) + o_p(1).$$

The first term on the right side of the above display can be written as

$$n^{1/3} \int_{D_n} E_{\psi(z)} (\phi''(X, \psi(z_0))) \left\{ (n^{1/3} (\hat{\psi}_n(z) - \psi(z_0)))^2 - (n^{1/3} (\hat{\psi}_n^0(z) - \psi(z_0)))^2 \right\} p_Z(z) dz.$$

On changing to the local variable  $h = n^{1/3} (z - z_0)$ , the above becomes

$$\int_{\tilde{D}_n} E_{\psi(z_0+h n^{-1/3})} (\phi''(X, \psi(z_0))) (X_n^2(h) - Y_n^2(h)) p_Z(z_0 + h n^{-1/3}) dh \equiv A_n + B_n$$

where

$$A_n \equiv \int_{\tilde{D}_n} E_{\psi(z_0)} (\phi''(X, \psi(z_0))) (X_n^2(h) - Y_n^2(h)) p_Z(z_0 + h n^{-1/3}) dh$$

and

$$B_n \equiv \int_{\tilde{D}_n} \left[ E_{\psi(z_0+h n^{-1/3})} (\phi''(X, \psi(z_0))) - E_{\psi(z_0)} (\phi''(X, \psi(z_0))) \right] (X_n^2(h) - Y_n^2(h)) p_Z(z_0 + h n^{-1/3}) dh.$$

The term  $B_n$  converges to 0 in probability on using the facts that eventually, with arbitrarily high probability,  $\tilde{D}_n$  is contained in an interval of the form  $[-M, M]$  on which the processes  $X_n$  and  $Y_n$  are  $O_p(1)$  and that for every  $M > 0$ ,

$$\sup_{|h| \leq M} |E_{\psi(z_0+h n^{-1/3})}(\phi''(X, \psi(z_0))) - E_{\psi(z_0)}(\phi''(X, \psi(z_0)))| \rightarrow 0,$$

by (A.6). Thus,

$$\begin{aligned} 2 \log \lambda_n &= I(\psi(z_0)) \int_{\tilde{D}_n} (X_n^2(h) - Y_n^2(h)) p_Z(z_0 + h n^{-1/3}) dh + o_p(1) \\ &= I(\psi(z_0)) p_Z(z_0) \int_{\tilde{D}_n} (X_n^2(h) - Y_n^2(h)) dh + o_p(1) \\ &= \frac{1}{a^2} \int_{\tilde{D}_n} (X_n^2(h) - Y_n^2(h)) dh + o_p(1) \end{aligned}$$

by similar arguments. We now deduce the asymptotic distribution of the expression on the right side of the above display. To this end, consider first the following lemma.

**Lemma 2.7** *Suppose that  $\{W_{n\epsilon}\}$ ,  $\{W_n\}$  and  $\{W_\epsilon\}$  are three sets of random variables such that*

(i)  $\lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} P[W_{n\epsilon} \neq W_n] = 0,$

(ii)  $\lim_{\epsilon \rightarrow 0} P[W_\epsilon \neq W] = 0,$

(iii) *For every  $\epsilon > 0$ ,  $W_{n\epsilon} \rightarrow_d W_\epsilon$  as  $n \rightarrow \infty$ .*

*Then  $W_n \rightarrow_d W$ , as  $n \rightarrow \infty$ .*

A version of this lemma can be found in Prakasa Rao (1969). Set

$$W_n = \frac{1}{a^2} \int_{\tilde{D}_n} (X_n^2(h) - Y_n^2(h)) dh \quad \text{and} \quad W = \frac{1}{a^2} \int_{\tilde{D}_n} \{(g_{a,b}(h))^2 - (g_{a,b}^0(h))^2\} dh.$$

Using Lemma 2.6, for each  $\epsilon > 0$ , we can find a compact set  $M_\epsilon$  of the form  $[-K_\epsilon, K_\epsilon]$  such that eventually,

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

Here  $D_{a,b}$  is the set on which the processes  $g_{a,b}$  and  $g_{a,b}^0$  vary. Now let

$$W_{n\epsilon} = \frac{1}{a^2} \int_{[-K_\epsilon, K_\epsilon]} (X_n^2(h) - Y_n^2(h)) dh \quad \text{and} \quad W_\epsilon = \int_{[-K_\epsilon, K_\epsilon]} \frac{1}{a^2} ((g_{a,b}(h))^2 - (g_{a,b}^0(h))^2) dh.$$

Since  $[-K_\epsilon, K_\epsilon]$  contains  $\tilde{D}_n$  with probability greater than  $1 - \epsilon$  eventually ( $\tilde{D}_n$  is the left-closed, right-open interval over which the processes  $X_n$  and  $Y_n$  differ) we have  $P[W_{n\epsilon} \neq W_n] < \epsilon$

eventually. Similarly  $P[W_\epsilon \neq W] < \epsilon$ . Also  $X_{n\epsilon} \rightarrow_d W_\epsilon$  as  $n \rightarrow \infty$ , for every fixed  $\epsilon$ . This is so because by Theorem 2.1  $(X_n(h), Y_n(h)) \rightarrow_d (g_{a,b}(h), g_{a,b}^0(h))$  as a process in  $\mathcal{L} \times \mathcal{L}$  and  $(f, g) \mapsto \int_{[-c,c]} (f^2(h) - g^2(h)) dh$  is a continuous real-valued function defined from  $\mathcal{L} \times \mathcal{L}$  to the reals. Thus all conditions of Lemma (2.7) are satisfied, leading to the conclusion that  $W_n \rightarrow_d W$ . The fact that the limiting distribution is actually independent of the constants  $a$  and  $b$ , thereby showing universality, falls out from Brownian scaling. Using (2.2) we obtain,

$$\begin{aligned} W &= \frac{1}{a^2} \int \{(g_{a,b}(h))^2 - (g_{a,b}^0(h))^2\} dh \\ &\equiv_d \frac{1}{a^2} a^2 (b/a)^{2/3} \int \{(g_{1,1}((b/a)^{2/3} h))^2 - (g_{1,1}^0((b/a)^{2/3} h))^2\} dh \\ &= \int \{(g_{1,1}(w))^2 - (g_{1,1}^0(w))^2\} dh, \end{aligned}$$

on making the change of variable:  $w = (b/a)^{2/3} h$ .

It only remains to show that  $R_n$  is  $o_p(1)$  as stated earlier. We outline the proof for  $R_{n,1}$ ; the proof for  $R_{n,2}$  is similar. We can write

$$R_{n,1} = \frac{1}{6} \mathbb{P}_n \left[ \phi'''(X, \hat{\psi}_n^*(Z)) \{n^{1/3} (\hat{\psi}_n(Z) - \psi(z_0))\}^3 \mathbf{1}(Z \in D_n) \right],$$

where  $\hat{\psi}_n^*(Z)$  is some point between  $\hat{\psi}_n(Z)$  and  $\psi(z_0)$ . On using the facts that  $D_n$  is eventually contained in a set of the form  $[z_0 - M n^{-1/3}, z_0 + M n^{-1/3}]$  with arbitrarily high probability on which  $\{n^{1/3} (\hat{\psi}_n(Z) - \psi(z_0))\}^3$  is  $O_p(1)$  and (A.5), we conclude that eventually, with arbitrarily high probability,

$$\begin{aligned} |R_{n,1}| &\leq \tilde{C} (\mathbb{P}_n - P) [B(X) \mathbf{1}(Z \in [z_0 - M n^{-1/3}, z_0 + M n^{-1/3}])] \\ &\quad + \tilde{C} P [B(X) \mathbf{1}(Z \in [z_0 - M n^{-1/3}, z_0 + M n^{-1/3}])], \end{aligned}$$

for some constant  $\tilde{C}$ . That the first term on the right side goes to 0 in probability is a consequence of an extended Glivenko-Cantelli theorem, whereas the second term goes to 0 by direct computation.  $\square$

### 3 Applications of the general theorems

In this section, we discuss some interesting special cases of the general theorem.

#### 3.1 Exponential Family Models

Consider the case of a one-parameter exponential family model, naturally parametrized. Thus,

$$p(x, \theta) = \exp[\theta T(x) + d(\theta)] h(x),$$

where  $\theta$  varies in an open interval  $\Theta$ . The function  $d$  possesses derivatives of all orders. Suppose we have  $Z \sim p_Z(\cdot)$  and  $X \mid Z = z \sim p(x, \psi(z))$  where  $\psi$  is increasing or decreasing in  $z$ . We are interested in making inference on  $\psi(z_0)$ , where  $z_0$  is an interior point in the support of  $Z$ . If  $p_Z$  and  $\psi$  satisfy conditions (a) and (b) of Section 2, the likelihood ratio statistic for testing  $\psi(z_0) = \theta_0$  converges to  $\mathbb{D}$  under the null hypothesis, since conditions (A.1) – (A.7) are readily satisfied for exponential family models. Note that:

$$l(x, \theta) = \theta T(x) + d(\theta) + \log h(x),$$

so that

$$\dot{l}(x, \theta) = T(x) + d'(\theta) \quad \text{and} \quad \ddot{l}(x, \theta) = d''(\theta).$$

Since we are in an exponential family setting, differentiation under the integral sign is permissible up to all orders. We note that  $l(x, \theta)$  is infinitely differentiable with respect to  $\theta$  for all  $x$ ; since  $E_\theta(\dot{l}(X, \theta)) = 0$  we have  $E_\theta(T(X)) = -d'(\theta)$  and  $I(\theta) = -E_\theta(\ddot{l}(X, \theta)) = -d''(\theta)$ . Also note that  $I(\theta) = E_\theta(\dot{l}(X, \theta)^2) = \text{Var}_\theta(T(X))$ . Since  $I(\theta) > 0$ , we get  $d''(\theta) < 0$  which implies the concavity of  $l(x, \theta)$ . Clearly  $I(\theta)$  is continuous everywhere. Hence conditions (A.1) – (A.4) are satisfied readily. Condition (A.6) is also satisfied easily:  $f_3(\theta) = d''(\theta)^2$  is clearly continuous in  $\theta$ ;  $f_2(\theta_1, \theta_2) = -d''(\theta_2)$  is also clearly differentiable in a neighborhood of  $(\theta_0, \theta_0)$  and that  $f_1(\theta_1, \theta_2)$  is continuous follows on noting that  $f_1(\theta_1, \theta_2) = I(\theta_1) + (d'(\theta_2) - d'(\theta_1))^2$ .

To check conditions (A.5) and (A.7) fix a neighborhood of  $\theta_0$ , say  $(\theta_0 - \epsilon, \theta_0 + \epsilon)$ . Since, for all  $x$ ,  $l'''(x, \theta) = d'''(\theta)$  which is continuous, we can actually choose  $B(x)$  in (A.5) to be a constant. It remains to verify condition (A.7). Since  $d'(\theta)$  and  $\dot{l}(x, \theta)$  are uniformly bounded for  $\theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon) \equiv \mathcal{N}$ , by choosing  $M$  sufficiently large, we can ensure that for some constant  $\gamma$  and  $\theta \in \mathcal{N}$ ,

$$H(\theta, M) \leq \xi(\theta, M) = E_\theta [(2T(X)^2 + \gamma) 1\{|T(X)| > M/2\}].$$

For  $\theta \in \mathcal{N}$ , consider

$$\xi(\theta, M) \leq \int 2(T(x)^2 + \gamma) e^{\theta T(x)} e^{d(\theta)} 1(|T(x)| > M/2) h(x) d\mu(x),$$

which in turn is dominated by

$$\sup_{\theta \in \mathcal{N}} e^{d(\theta)} \left[ \int 2(T(x)^2 + \gamma) (e^{(\theta_0 + \epsilon)T(x)} + e^{(\theta_0 - \epsilon)T(x)}) 1(|T(x)| > M/2) h(x) d\mu(x) \right].$$

The expression above is not dependent on  $\theta$  and hence serves as a bound for  $\sup_{\theta \in \mathcal{N}} H(\theta, M)$ . As  $M$  goes to  $\infty$  the above expression goes to 0; this is seen by an appeal to the DCT and the fact that  $T^2(X) + \gamma$  is integrable at parameter values  $\theta_0 - \epsilon$  and  $\theta_0 + \epsilon$ .

The nice structure of exponential family models actually leads to a somewhat simpler characterization of the MLE's  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$ . If  $B$  is a block of indices on which  $\hat{\psi}_n(Z_{(i)})$  is constant with common value equal to  $w$ , then we have  $\sum_{i \in B} (T(X_{(i)}) + d'(w)) = 0$ ; hence

$$-d'(w) = n_B^{-1} \sum_{i \in B} T(X_{(i)}) \tag{3.12}$$

where  $n_B$  is the number of indices in the block  $B$ . It follows from this that the unconstrained MLE  $\hat{\psi}_n$  can actually be written as:

$$\left\{ -d'(\hat{\psi}_n(Z_{(i)})) \right\}_{i=1}^n = \text{slogcm} \left\{ G_n(Z_{(i)}), V_n(Z_{(i)}) \right\}_{i=0}^n,$$

where

$$G_n(z) = \frac{1}{n} \sum_{i=1}^n 1(Z_i \leq z) \quad \text{and} \quad V_n(z) = \frac{1}{n} \sum_{i=1}^n T(X_i) 1(Z_i \leq z),$$

and  $G_n(Z_{(0)}) \equiv V_n(Z_{(0)}) = 0$ . The MLE  $\hat{\psi}_n^0$  is characterized in a similar fashion but as constrained slopes of the cumulative sum diagram formed by the points  $\{G_n(Z_{(i)}), V_n(Z_{(i)})\}_{i=0}^n$ . Thus, it is possible to avoid the “self-consistent” characterization for these models and the asymptotic distribution of the MLE’s may be obtained by more direct methods.

However, the self-consistent characterization for MLE’s is unavoidable in general, even within the context of exponential family models. Consider, for example, curved one-dimensional exponential families of the form:

$$\log p(x, \theta) = c(\theta) T(x) + d(\theta) S(x) + B(\theta)$$

with no linear dependence between  $c(\theta)$  and  $d(\theta)$  or  $T(x)$  and  $S(x)$ . Also suppose that the conditions (A.1) – (A.7) are satisfied (this happens under fairly mild conditions on the components of the log-density). The likelihood ratio statistic for testing  $\psi(z_0) = \theta_0$  then converges to  $\mathbb{D}$  and if  $B$  is a block of indices on which  $\hat{\psi}_n(Z_{(i)})$  assumes a common value  $w$ , then the equation  $\sum_{i \in B} \phi'(X_{(i)}, w) = 0$  is seen to boil down to

$$c'(w) \sum_{i \in B} T(X_{(i)}) + d'(w) \sum_{i \in B} S(X_{(i)}) + B'(w) n_B = 0. \quad (3.13)$$

Without an explicit knowledge of the functional forms of  $c, d$  and  $B$  there is no way to represent  $w$  (or a monotone transformation of it) in terms of an average (in contrast to the situation in 3.12); consequently, an explicit unified representation as the slope of the convex minorant of a known cusum diagram cannot be achieved in general. The self-consistent characterization for the MLE’s has the merit of providing a unified characterization of MLE’s across all (sufficiently regular) parametric models, which is precisely why it has been exploited in the general approach. In the context of the current status model (discussed in Banerjee and Wellner (2001)) the explicit characterization could indeed be used, but this was possible owing to the connection of that model to the Bernoulli densities which form a one parameter exponential family. This will be explained in Subsection 3.2.

We provide two examples of curved exponential family models to further illustrate the points made above. Consider first the  $N(\theta^{-1}, \theta^{-2})$  parametric family (with  $\theta > 0$ ). This is a regular parametric model with

$$l(x, \theta) \equiv \log p(x, \theta) = \log \theta - \theta^2 x^2 / 2 + \theta x + C$$

where  $C$  is a constant. Hence,

$$\dot{l}(x, \theta) = 1/\theta - \theta x^2 + x.$$

This is a decreasing function of  $\theta$ , showing that  $l$  is concave. The conditions (A.1) – (A.7) can be verified fairly easily and the likelihood ratio statistic for testing  $\psi(z_0) = \theta_0$  (for  $\theta_0 > 0$ ) will converge to  $\mathbb{D}$ . In this case, (3.13) reduces to:

$$\frac{n_B}{w} - w \sum_{i \in B} X_{(i)}^2 + \sum_{i \in B} X_{(i)} = 0.$$

One can now solve the above quadratic to obtain an explicit but cumbersome representation for  $w$ . This could be used as an alternative (to the self-consistent characterization) means of deriving the asymptotic distribution of  $\hat{\psi}_n$ , which would involve studying the limit distribution of normalized versions of

$$A_n(z) = \frac{1}{n} \sum_{i=1}^n X_i 1(Z_i \leq z) \quad \text{and} \quad B_n(z) = \frac{1}{n} \sum_{i=1}^n X_i^2 1(Z_i \leq z).$$

This approach however heavily relies on the specific structure of the model at hand and provide little insight about the situation for a different model. Consider, for example, the following parametric model:

$$p(x, \theta) = e^{-\theta x} (e^\theta - 1) 1(x < k) + e^{-\theta x} 1(x = k), \quad x = 0, 1, 2, \dots, k,$$

for  $\theta > 0$ . We have,

$$l(x, \theta) = -\theta x + \log(e^\theta - 1) 1(x < k).$$

This is concave in  $\theta$ ; as with the previous model, conditions (A.1) – (A.7) may be verified in a fairly straightforward manner. Now,

$$\dot{l}(x, \theta) = -x + \frac{e^\theta}{e^\theta - 1} 1(x < k).$$

For this model, (3.13) translates to:

$$-\sum_{i \in B} X_{(i)} + \frac{e^w}{e^w - 1} \sum_{i \in B} 1(X_{(i)} < k) = 0.$$

We can now solve explicitly for  $w$  to get:

$$e^w = \frac{\sum_{i \in B} X_{(i)}}{\sum_{i \in B} X_{(i)} - \sum_{i \in B} 1(X_{(i)} < k)}.$$

As before, defining the processes

$$A_n(z) = \frac{1}{n} \sum_{i=1}^n X_i 1(Z_i \leq z) \quad \text{and} \quad B_n(z) = \frac{1}{n} \sum_{i=1}^n (X_i - 1(X_i < k)) 1(Z_i \leq z),$$

it can be inferred that

$$\{e^{\hat{\psi}_n(Z_{(i)})}\}_{i=1}^n = \text{slogcm} \{B_n(Z_{(i)}), A_n(Z_{(i)})\}_{i=0}^n.$$

Thus, for this model, an explicit “slope of convex minorant” characterization is available (in contrast to the previous one) and the asymptotic distribution of the MLE’s may be obtained by simpler methods.

Thus, explicit representations for the MLE’s will typically involve case by case considerations beyond one–parameter exponential families, since the representations for different models will differ quite considerably. Consequently, methods for establishing distributional convergence using such representations will typically be model–specific and not provide a unified perspective which is what we have sought in this paper.

### 3.2 The Motivating Examples Revisited.

Here we show how the general theory applies to the motivating examples considered in Section 2.

- (a) **Monotone Regression Model:** To test  $\psi(z_0) = \theta_0$  for an interior point  $z_0$  in the domain of  $\psi$  in the monotone regression model, assume that conditions (a) and (b) of Section 2 are satisfied.

Let  $p(x, \theta)$  denote the  $N(\theta, \sigma^2)$  density. Then:

$$p(x, \theta) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -x^2/(2\sigma^2) + \theta x/\sigma^2 - \theta^2/(2\sigma^2) \right].$$

For fixed  $\sigma^2$ , this is a one–parameter exponential family with natural parameter  $\theta$  and sufficient statistic  $T(X) = X/\sigma^2$ . It follows immediately from Theorem 2.2 that  $2 \log \lambda_n$  in this problem converges to  $\mathbb{D}$ . It is instructive to write down the form of the likelihood ratio statistic in this problem. We have:

$$2 \log \lambda_n = \frac{1}{\sigma^2} \left( \sum_{i=1}^n (X_i - \hat{\psi}_n^0(Z_i))^2 - \sum_{i=1}^n (X_i - \hat{\psi}_n(Z_i))^2 \right).$$

To actually get a confidence interval for  $\psi(z_0)$  using this result a consistent estimator of  $\sigma^2$  should be plugged into the above expression. A standard estimator is one given by Rice (1984). Note further that the above distributional convergence holds for non–normal i.i.d. mean 0 errors as well. In this case, one may interpret  $p(x, \theta)$  above as representing a “pseudo” or a “working” likelihood. The pseudo likelihood ratio statistic can be interpreted as a scaled residual sum of squares statistic and the MLE’s  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$  are the unconstrained and constrained least squares estimate of the function  $\psi$ . To see that the limit distributions of the MLE’s and the likelihood ratio statistic remain unaltered under non–Gaussian errors one can check that the steps of the proofs of the general theorems go through in exactly the same way as for Gaussian errors; the underlying distribution of the errors enters into the proofs only through conditional expectations of the form  $E(\dot{l}^2(X, \psi(Z)) \mid Z = z)$  or  $E(\ddot{l}(X, \psi(Z)) \mid Z = z)$  and other similar quantities which are unaffected by the actual form of the error distribution and only depend upon  $\psi(z)$  and  $\sigma^2$ . Alternatively, a direct proof

can be given using the simple structure of the Gaussian likelihood.

(b) **Binary Choice Model:** We note that under Assumptions (a) and (b) of Section 2, this falls in the framework of our general theorems:  $\Delta \mid Z = z \sim p(x, \psi(z))$  where  $p(\delta, \theta) = \theta^\delta (1 - \theta)^{1-\delta}$ . Here  $\delta \in \{1, 0\}$  and  $0 < \theta < 1$ . Set

$$\eta = \log \left( \frac{\theta}{1 - \theta} \right) \quad \text{and} \quad \tilde{\psi} = \log \left( \frac{\psi}{1 - \psi} \right).$$

Since  $\psi$  is monotone, so is  $\tilde{\psi}$ . Under this reparametrization,  $\Delta \mid Z = z \sim q(x, \tilde{\psi}(z))$  where  $q(\delta, \eta)$  is the one-parameter exponential family given by

$$\log q(\delta, \eta) = \delta \eta - \log(1 + e^\eta).$$

Testing  $\psi(z_0) = \theta_0$  is the same as testing  $\tilde{\psi}(z_0) = \eta_0$  where  $\eta_0 = \theta_0/(1 - \theta_0)$ . It follows immediately from Theorem 2.2 that the likelihood ratio statistic in this model converges to  $\mathbb{D}$ .

(c) **Current Status Model:** Suppose that we want to test  $F(t_0) = \theta_0$  in the current status model, for  $\theta_0 \in (0, 1)$  and  $t_0$ , an interior point in the domain of  $T$ . Assuming that  $F$  is continuously differentiable in a neighborhood of  $t_0$  with  $F'(t_0) > 0$  and that the density  $g$  of  $T$  is continuous and positive in a neighborhood of  $t_0$  we can show that the likelihood ratio statistic for testing  $F(t_0) = \theta_0$  based on  $\{\Delta_i, T_i\}_{i=1}^n$  converges in distribution to  $\mathbb{D}$ . This is the key result in Banerjee and Wellner (2001) which can now be derived as a special application of our general theorems. As noted in Section 2, the current status model can be viewed as a binary regression model with the survival time distribution,  $F$ , assuming the role of  $\psi$ . Hence, the result on distributional convergence of the likelihood ratio statistic follows directly from Example (b).

(d) **Poisson Regression Model:** Under appropriate assumptions on  $\psi$  and  $p_Z$  this model falls within the scope of our general theorems; note that in this case

$$\log p(x, \theta) = -\theta + x \log \theta - x!.$$

That this model satisfies conditions (A.1) – (A.7) can be checked directly, or more efficiently, by reparametrizing to  $\eta = \log \theta$  (and correspondingly changing  $\psi$  to  $\tilde{\psi} \equiv \log \psi$ ) whence the density can be written as:

$$\log q(x, \eta) = -e^\eta + \eta x - x!.$$

This is a one-parameter exponential family naturally parametrized. It follows that the likelihood ratio statistic for testing  $\psi(z_0) = \theta_0$  ( $\theta_0 > 0$ ), or equivalently  $\tilde{\psi}(z_0) = \log \theta_0$  converges in distribution to  $\mathbb{D}$ .

## 4 Concluding Remarks

In this paper, we have developed a unified approach for studying the asymptotics of likelihood based inference in monotone response models. A crucial aspect of these models is the fact that

conditional on the covariate  $Z$ , the response  $X$  is generated from a parametric family that is regular in the usual sense; consequently, the conditional score functions, their derivatives and the conditional information play a key role in describing the asymptotic behavior of the maximum likelihood estimates of the function  $\psi$ . It must be noted though that there are monotone function models that asymptotically exhibit similar behavior though they are not monotone response models in the sense of this paper. The problem of estimating a monotone instantaneous hazard/density function based on i.i.d. observations from the underlying probability distribution or from right-censored data (with references provided in Section 1) are cases in point. While methods similar to ones developed in this paper apply to the hazard estimation problems and show that  $\mathbb{D}$  still arises as the limit distribution of the pointwise likelihood ratio statistic, the asymptotics of the likelihood ratio for the density estimation problems still remain to be established.

The formulation that we have developed in this paper admits some natural extensions which we briefly indicate. For example, one would like to investigate what happens in the case of (conditionally) multidimensional parametric models. Consider a  $k$  parameter exponential family model of the form

$$p(x, \eta_1, \eta_2, \dots, \eta_k) = \exp \left( \sum_{i=1}^k \eta_i T_i(x) - B(\eta_1, \eta_2, \dots, \eta_k) \right) h(x).$$

Suppose we have  $n$  i.i.d. observations from the distribution of  $(X, Z)$  where  $Z = (Z_1, Z_2, \dots, Z_k)$  is a random vector such that  $Z_1 < Z_2 < \dots < Z_k$  with probability 1 and given  $Z = z$ ,  $X$  is distributed as  $p(x, \psi(z_1), \psi(z_2) - \psi(z_1), \dots, \psi(z_k) - \psi(z_{k-1}))$  where  $\psi$  is strictly increasing. Would the likelihood ratio statistic for testing  $\psi(z_0) = \theta_0$  converge in this case to  $\mathbb{D}$  as well? The above model is motivated by the Case  $k$  interval censoring problem. Here  $X$  is the survival time of an individual and  $T = (T_1, T_2, \dots, T_k)$  is a random vector of follow-up times. One only observes  $(\Delta_1, \Delta_2, \dots, \Delta_{k+1})$  where  $\Delta_i = 1(X \in (T_{i-1}, T_i])$  (interpret  $T_0 = 0$  and  $T_{k+1} = \infty$ ). The goal is to estimate the monotone function  $F$  (the distribution function of  $X$ ) based on the above data. It is not difficult to see that with  $Z = T$  and  $X = (\Delta_1, \Delta_2, \dots, \Delta_{k+1})$ , this is a special case of the  $k$  parameter model with

$$p(x, \eta_1, \eta_2, \dots, \eta_k) = \prod_{i=1}^{k+1} \eta_i^{\delta_i},$$

where  $\eta_{k+1} = 1 - (\sum_{i=1}^k \eta_i)$ . This is the multinomial distribution. There are reasons to believe that  $\mathbb{D}$  should still arise as the limit of the likelihood ratio statistic for testing  $F(t_0) = \theta_0$  in the Case  $k$  interval censoring model. We conjecture that the limit distribution of the likelihood ratio statistic in the general case will also be  $\mathbb{D}$ .

Another potential extension is to semiparametric models where the infinite dimensional component is a monotone function. Here is a general formulation: Consider a random vector  $(Y, X, Z)$  where  $Z$  is unidimensional. Suppose that the distribution of  $Y$  conditional on  $(X, Z) = (x, z)$  is given by  $p(y, \beta^T x, \psi(z))$  where  $p(y, \theta, \eta)$  is a parametric model. We are interested in making inference on both  $\beta$  and  $\psi$ . The above formulation is fairly general and includes for example the partially linear regression model:  $Y = \beta^T X + \psi(Z) + \epsilon$  where  $\psi$  a monotone function (certain aspects

of this model have been studied by Huang(2002)), or the Cox Proportional Hazards Model with Current Status data (also studied by Huang (1994), Huang (1996)) and others of interest. In the light of previous results, we expect that under appropriate conditions on  $p$ ,  $\sqrt{n}(\hat{\beta}_{MLE} - \beta)$  will converge to a normal distribution with asymptotic dispersion given by the inverse of the efficient information matrix and the likelihood ratio statistic for testing  $\beta = \beta_0$  will be asymptotically  $\chi^2$ . The theory developed in Murphy and Van der Vaart (1997,2000) should prove very useful in this regard. As far as estimation of the nonparametric component goes,  $\hat{\psi}_n$ , the MLE of  $\psi$  should exhibit an  $n^{1/3}$  rate of convergence to a non-normal limit and the likelihood ratio for testing  $\psi$  pointwise should still converge to  $\mathbb{D}$ . This will be explored elsewhere and the ideas of the current paper should prove extremely useful in dealing with the nonparametric component of the model.

## 5 Proofs of lemmas

**Proof of Lemma 2.2:** We will prove the first assertion. The second follows from the first on noting that for all values of  $z$ ,

$$\left| \hat{\psi}_n^0(z) - \psi(z_0) \right| \leq \left| \hat{\psi}_n(z) - \psi(z_0) \right|.$$

We only show that

$$\sup_{h \in [-M_0, 0]} \left| \hat{\psi}_n(z_0 + h n^{-1/3}) - \psi(z_0) \right| = O_p(n^{-1/3}); \quad (\star)$$

the fact that  $\sup_{h \in [0, M_0]} \left| \hat{\psi}_n(z_0 + h n^{-1/3}) - \psi(z_0) \right| = O_p(n^{-1/3})$  is established similarly. Fix  $M_0 > 0$ . We show that there exists an  $M > 0$  such that eventually

$$P(\hat{\psi}_n(z_0 - M n^{-1/3}) \geq \psi(z_0)) < \epsilon/2;$$

using the monotonicity of  $\hat{\psi}_n$ ,  $M$  can be chosen to be larger than  $M_0$ . It follows that eventually, with probability at least  $1 - \epsilon/2$ ,  $\hat{\psi}_n$  has a jump in the interval  $(z_0 - M n^{-1/3}, z_0]$ .

Let  $K > 0$ ; for  $z \leq z_0 - \frac{1}{2} K n^{-1/3}$  define:

$$H_n(z, u) = \mathbb{P}_n \left( \dot{l}(X, \psi(z_0)) 1(z \leq Z \leq u) \right).$$

Also, let

$$H(z, u) = P \left( \dot{l}(X, \psi(z_0)) 1(z \leq Z \leq u) \right).$$

By a standard Glivenko-Cantelli argument,

$$\sup_{z \leq z_0 - K n^{-1/3/2}} |H_n(z, z_0) - H(z, z_0)| \rightarrow_{a.s} 0. \quad (5.14)$$

We now show that  $\sup_{z \leq z_0 - K n^{-1/3/2}} H(z, z_0) < 0$  for any  $n$ . To see this, note that for any  $t < z_0$ ,  $\dot{l}(x, \psi(z_0)) < \dot{l}(x, \psi(t))$ ; hence

$$\int_{\mathcal{X}} \dot{l}(x, \psi(z_0)) p(x, \psi(t)) d\mu(x) < \int_{\mathcal{X}} \dot{l}(x, \psi(t)) p(x, \psi(t)) d\mu(x) = 0.$$

It follows that for any  $z \leq z_0 - K n^{-1/3}/2$ ,

$$\begin{aligned} H(z, z_0) &= \int_z^{z_0} \left\{ \int_{\mathcal{X}} \dot{l}(x, \psi(z_0)) p(x, \psi(t)) d\mu(x) \right\} p_Z(t) dt \\ &\leq \int_{z_0 - K n^{-1/3}/2}^{z_0} \left\{ \int_{\mathcal{X}} \dot{l}(x, \psi(z_0)) p(x, \psi(t)) d\mu(x) \right\} p_Z(t) dt < 0. \end{aligned}$$

This implies that  $\sup_{z \leq z_0 - K n^{-1/3}/2} H(z, z_0) < 0$  for any  $n$ ; in conjunction with (5.14) this shows that

$$\sup_{z \leq z_0 - K n^{-1/3}/2} H_n(z, z_0) < 0 \text{ eventually a.s.} \quad (5.15)$$

Now, suppose that there is some  $\epsilon_0 > 0$ , such that for any  $\tilde{M} > 0$  we can find a subsequence  $\{n'\}$ , such that  $P(A_{n'}) > \epsilon/2$ , where

$$A_{n'} = \{\hat{\psi}_{n'}(z_0 - \tilde{M} n'^{-1/3}) \geq \psi(z_0)\}.$$

Let  $\tau_{n'}$  be the last jump time of  $\hat{\psi}_{n'}$  before  $z_0 - \tilde{M} n'^{-1/3}$ . From the characterization of the MLE  $\hat{\psi}_{n'}$ , it follows that for any  $Z_{(j)} > \tau_{n'}$ ,

$$\sum_{\tau_{n'} \leq Z_i \leq Z_{(j)}} \dot{l}(X_i, \hat{\psi}_{n'}(Z_i)) \geq 0. \quad (5.16)$$

Now, consider  $H_{n'}(\tau_{n'}, Z_{(j)})$  for any  $Z_{(j)} > \tau_{n'}$ . This can be written as:

$$H_{n'}(\tau_{n'}, Z_{(j)}) = \frac{1}{n'} \sum_{\tau_{n'} \leq Z_i \leq Z_{(j)}} \dot{l}(X_i, \psi(z_0)).$$

Now, on the set  $A_{n'}$ , for  $\tau_{n'} \leq Z_i \leq Z_{(j)}$ , we have:

$$\hat{\psi}_{n'}(Z_i) \geq \hat{\psi}_{n'}(\tau_{n'}) = \hat{\psi}_{n'}(z_0 - \tilde{M} n'^{-1/3}) \geq \psi(z_0).$$

This implies that, on the set  $A_{n'}$ , for  $\tau_{n'} \leq Z_i \leq Z_{(j)}$ ,

$$\dot{l}(X_i, \hat{\psi}_{n'}(Z_i)) \leq \dot{l}(X_i, \psi(z_0)).$$

In conjunction with (5.16) this implies that  $H_{n'}(\tau_{n'}, Z_{(j)}) \geq 0$  for any  $Z_{(j)} > \tau_{n'}$ . This, in turn, implies that on the set  $A_{n'}$ ,

$$\sup_{z \leq z_0 - \tilde{M} n'^{-1/3}/2} H_{n'}(z, z_0) \geq 0,$$

which is a contradiction to (5.15). It follows that given any  $\epsilon > 0$ , there exists  $M > 0$  (which may be chosen to be larger than  $M_0$ ), such that, eventually,

$$P(\hat{\psi}_n(z_0 - M n^{-1/3}) < \psi(z_0)) \geq 1 - \epsilon/2.$$

Using analogous arguments, we can find an  $M'$  sufficiently large such that the probability of  $\hat{\psi}_n$  having a jump in the interval  $(z_0 - (M + M') n^{-1/3}, z_0 - M n^{-1/3}]$  is eventually larger than  $1 - \epsilon/2$ .

Let  $B_n$  denote the set on which  $\hat{\psi}_n$  has a jump time in the interval  $[z_0 - (M + M')n^{-1/3}, z_0 - Mn^{-1/3}]$  and let  $\tau_n$  be such a jump-time (say, the first). We claim that by choosing  $c$  to be sufficiently large (and larger than  $M + M'$ ), we can ensure that the sets

$$B_n \cap \left\{ \hat{\psi}_n(\tau_n) > \psi(z_0 - cn^{-1/3}) \right\}$$

have probability larger than  $1 - \epsilon$ , eventually. Now, note that the sets

$$C_n \equiv \left\{ \hat{\psi}_n(z_0 - Mn^{-1/3}) < \psi(z_0) \right\} \cap B_n \cap \left\{ \hat{\psi}_n(\tau_n) > \psi(z_0 - cn^{-1/3}) \right\}$$

are contained in the sets

$$D_n \equiv \left\{ \psi(z_0 - cn^{-1/3}) \leq \hat{\psi}(z_0 - Mn^{-1/3}) < \psi(z_0) \right\}.$$

This observation, along with Bonferroni's inequality, implies that eventually with probability at least  $1 - 3\epsilon/2$

$$\psi(z_0 - cn^{-1/3}) \leq \hat{\psi}_n(z_0 - Mn^{-1/3}) \leq \psi(z_0).$$

Since,

$$\sup_{h \in [-M_0, 0]} \left| \hat{\psi}_n(z_0 + hn^{-1/3}) - \psi(z_0) \right| \leq \sup_{h \in [-M, 0]} \left| \hat{\psi}_n(z_0 + hn^{-1/3}) - \psi(z_0) \right|$$

and by the monotonicity of  $\hat{\psi}_n$ ,

$$\sup_{h \in [-M, 0]} \left| \hat{\psi}_n(z_0 + hn^{-1/3}) - \psi(z_0) \right| = \left| \hat{\psi}_n(z_0 - Mn^{-1/3}) - \psi(z_0) \right|$$

and eventually, with probability at least  $1 - 3\epsilon/2$ ,

$$\left| \hat{\psi}_n(z_0 - Mn^{-1/3}) - \psi(z_0) \right| \leq \left| \psi(z_0 - cn^{-1/3}) - \psi(z_0) \right|$$

which is  $O(n^{-1/3})$ , the assertion  $(\star)$  follows.

It now remains to prove the claim made above. We claimed that by choosing  $c > M + M'$  to be sufficiently large, we could ensure that the sets

$$B_n \cap \left\{ \hat{\psi}_n(\tau_n) > \psi(z_0 - cn^{-1/3}) \right\}$$

would have probability larger than  $1 - \epsilon$ , eventually. Suppose this is not true; then for every  $c > M + M'$  there exists a  $\tilde{c} > c$  and a subsequence  $\{n'\}$  such that

$$P \left( B_{n'} \cap \left\{ \hat{\psi}_{n'}(\tau_{n'}) > \psi(z_0 - \tilde{c}n'^{-1/3}) \right\} \right) \leq 1 - \epsilon.$$

Since  $B_{n'}$  eventually has probability greater than  $1 - \epsilon/2$ , this implies that

$$P \left( B_{n'} \cap \left\{ \hat{\psi}_{n'}(\tau_{n'}) \leq \psi(z_0 - \tilde{c}n'^{-1/3}) \right\} \right) \geq \epsilon/2.$$

We show that this will cause a contradiction. Let  $D_{n'}$  denote the set on the left side of the above display. From the characterization of the MLE  $\hat{\psi}_{n'}$ , we know that for all  $Z_{(j)} < \tau_{n'}$ :

$$\sum_{Z_{(j)} \leq Z_i < \tau_{n'}} \dot{l}(X_i, \hat{\psi}_{n'}(Z_i)) \leq 0. \quad (5.17)$$

Now, on the set  $D_{n'}$ , for any  $Z_i$  with  $Z_i < \tau_{n'}$ , we have

$$\hat{\psi}_n(Z_i) \leq \hat{\psi}_n(\tau_{n'}) \leq \psi(z_0 - \tilde{c} n'^{-1/3})$$

and this shows that,

$$\dot{l}(X_i, \psi(z_0 - c n'^{-1/3})) \leq \dot{l}(X_i, \hat{\psi}_{n'}(Z_i)).$$

In conjunction with (5.17) this implies that on  $D_{n'}$

$$\sum_{z_0 - \tilde{c} n'^{-1/3} \leq Z_i < \tau_{n'}} \dot{l}(X_i, \psi(z_0 - \tilde{c} n'^{-1/3})) \leq 0.$$

Now, for each  $n$  and  $z_0 - (M + M') n^{-1/3} \leq z \leq z_0 - M n^{-1/3}$ , define

$$H_n(z_0 - c n^{-1/3}, z) = n^{2/3} \mathbb{P}_n \left( \dot{l}(X, \psi(z_0 - c n^{-1/3})) 1(Z \in [z_0 - c n^{-1/3}, z]) \right).$$

Thus, for any  $c$ , we can produce a  $\tilde{c} > c$ , a subsequence  $\{n'\}$  and a sequence of events  $D_{n'}$  with probability at least  $\epsilon/2$  such that on  $D_{n'}$ ,  $H_{n'}(z_0 - c n'^{-1/3}, \tau_{n'}) \leq 0$ . This implies that on  $D_{n'}$

$$\inf_{z \in [z_0 - (M+M') n'^{-1/3}, z_0 - M n'^{-1/3}]} H_{n'}(z_0 - \tilde{c} n'^{-1/3}, z) \leq 0.$$

However, we will show that if  $c$  is chosen to be sufficiently large, we can ensure that with arbitrarily high probability, eventually,

$$\inf_{z \in [z_0 - (M+M') n^{-1/3}, z_0 - M n^{-1/3}]} H_n(z_0 - c n^{-1/3}, z) > 0.$$

This will give the desired contradiction. To establish our last claim, for any  $z \in [z_0 - (M + M') n^{-1/3}, z_0 - M n^{-1/3}]$ , consider

$$\begin{aligned} E \left[ H_n(z_0 - c n^{-1/3}, z) \right] &= E \left[ n^{2/3} \mathbb{P}_n \left[ \dot{l}(X, \psi(z_0 - c n^{-1/3})) 1 \left( Z \in [z_0 - c n^{-1/3}, z] \right) \right] \right] \\ &= n^{2/3} E \left[ \dot{l}(X, \psi(z_0 - c n^{-1/3})) 1 \left( Z \in [z_0 - c n^{-1/3}, z] \right) \right] \\ &= n^{2/3} \int_{z_0 - c n^{-1/3}}^z E_{\psi(t)}(\dot{l}(X, \psi(z_0 - c n^{-1/3}))) p_Z(t) dt \\ &\geq n^{2/3} \int_{z_0 - c n^{-1/3}}^{z_0 - (M+M') n^{-1/3}} E_{\psi(t)}(\dot{l}(X, \psi(z_0 - c n^{-1/3}))) p_Z(t) dt \equiv \xi_n, \end{aligned}$$

since

$$n^{2/3} \int_{z_0 - (M+M') n^{-1/3}}^z E_{\psi(t)}(\dot{l}(X, \psi(z_0 - c n^{-1/3}))) p_Z(t) dt \geq 0.$$

This follows from the fact that for  $t$  in the above range  $\psi(z_0 - cn^{-1/3}) \leq \psi(t)$ , implying that  $\dot{l}(X, \psi(z_0 - cn^{-1/3})) \geq \dot{l}(X, \psi(t))$ . Hence,

$$E_{\psi(t)} \dot{l}(X, \psi(z_0 - cn^{-1/3})) \geq E_{\psi(t)} \dot{l}(X, \psi(t)) = 0.$$

Hence,

$$\inf_{z \in [z_0 - (M+M')n^{-1/3}, z_0 - Mn^{-1/3}]} E \left[ H_n(z_0 - cn^{-1/3}, z) \right] \geq \xi_n.$$

Next,

$$\xi_n = n^{2/3} \int_{z_0 - cn^{-1/3}}^{z_0 - (M+M')n^{-1/3}} E_{\psi(t)} [\dot{l}(X, \psi(z_0 - cn^{-1/3})) - \dot{l}(X, \psi(t))] p_Z(t) dt.$$

Changing to the local variable,  $h = n^{1/3}(t - z_0)$ , we can write  $\xi_n$  as

$$-n^{1/3} \int_{-c}^{-(M+M')} E_{\psi(z_0 + un^{-1/3})} [\dot{l}(X, \psi(z_0 + un^{-1/3})) - \dot{l}(X, \psi(z_0 - cn^{-1/3}))] p_Z(z_0 + un^{-1/3}) du;$$

a Taylor expansion of  $\dot{l}(X, \psi(z_0 + un^{-1/3}))$  about  $\dot{l}(X, \psi(z_0 - cn^{-1/3}))$  allows this to be written as  $S_n + r_n$  where

$$\begin{aligned} S_n &= -n^{1/3} \int_{-c}^{-(M+M')} E_{\psi(z_0 + un^{-1/3})} \ddot{l}(X, \psi(z_0 - cn^{-1/3})) \\ &\quad \times (\psi(z_0 + un^{-1/3}) - \psi(z_0 - cn^{-1/3})) p_Z(z_0 + un^{-1/3}) du \end{aligned}$$

and

$$\begin{aligned} r_n &= -\frac{n^{1/3}}{2} \int_{-c}^{-(M+M')} E_{\psi(z_0 + un^{-1/3})} l'''(X, \psi_n^*(u, c)) \\ &\quad \times (\psi(z_0 + un^{-1/3}) - \psi(z_0 - cn^{-1/3}))^2 p_Z(z_0 + un^{-1/3}) du, \end{aligned}$$

where  $\psi_n^*(u, c)$  is some point between  $\psi(z_0 - cn^{-1/3})$  and  $\psi(z_0 + un^{-1/3})$ . Now, using the fact that

$$n^{1/3}(\psi(z_0 + un^{-1/3}) - \psi(z_0 - cn^{-1/3})) = (u + c) \psi'(z_0) + o(1) \quad (\star\star)$$

uniformly for  $u \in [-c, -(M + M')]$  we have

$$S_n = - \int_{-c}^{-(M+M')} E_{\psi(z_0 + un^{-1/3})} \ddot{l}(X, \psi(z_0 - cn^{-1/3})) (u + c) \psi'(z_0) p_Z(z_0 + un^{-1/3}) du + o(1);$$

using Assumption (A.6) it is now easily concluded that

$$S_n \rightarrow \frac{(c - (M + M'))^2}{2} I(\psi(z_0)) p_z(z_0) \psi'(z_0) > 0.$$

Using Assumption (A.5) in conjunction with  $(\star\star)$  we have

$$\left| n^{1/3} r_n \right| \leq K \int_{-c}^{-(M+M')} (\psi'(z_0) (u + c))^2 p_Z(z_0 + un^{-1/3}) du + o(1)$$

implying that  $r_n = O(n^{-1/3})$ . It follows that

$$\liminf_{n \rightarrow \infty} \inf_{z \in [z_0 - (M+M')n^{-1/3}, z_0 - Mn^{-1/3}]} E \left[ H_n(z_0 - cn^{-1/3}, z) \right] \geq \liminf_{n \rightarrow \infty} \xi_n.$$

But

$$\liminf_{n \rightarrow \infty} \xi_n = \frac{(c - (M + M'))^2}{2} I(\psi(z_0)) p_Z(z_0) \psi'(z_0),$$

which converges to  $\infty$  as  $c \rightarrow \infty$  at a quadratic rate. On the other hand, we can show that

$$\sup_{z \in [z_0 - (M+M')n^{-1/3}, z_0 - Mn^{-1/3}]} \text{Var} \left[ H_n(z_0 - cn^{-1/3}, z) \right] \leq \delta_n,$$

where

$$\delta_n = n^{1/3} E \left[ i^2(X, \psi(z_0 - cn^{-1/3})) 1(Z \in [z_0 - cn^{-1/3}, z_0 - Mn^{-1/3}]) \right].$$

Now  $\delta_n \rightarrow (c - M) I(\psi(z_0)) p_Z(z_0)$ . Hence,

$$\limsup_{n \rightarrow \infty} \sup_{z \in [z_0 - (M+M')n^{-1/3}, z_0 - Mn^{-1/3}]} \text{Var} \left[ H_n(z_0 - cn^{-1/3}, z) \right] \leq (c - M) I(\psi(z_0)) p_Z(z_0),$$

which grows to  $\infty$  only at a linear rate as  $c \rightarrow \infty$ . As in the proof of Lemma 6.2.2 of Huang (1994), we conclude that with arbitrarily high probability

$$\inf_{z \in [z_0 - (M+M')n^{-1/3}, z_0 - Mn^{-1/3}]} H_n(z_0 - cn^{-1/3}, z) > 0,$$

eventually.  $\square$

**Proof of Lemma 2.3:** It suffices to show that  $\tilde{B}_{n,\Psi}(h)$  converges to the process  $aW(h) + bh^2$  in  $l^\infty[-K, K]$ , the space of uniformly bounded functions on  $[-K, K]$  equipped with the topology of uniform convergence, for every  $K > 0$ . We can write,

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

where

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

To establish the above convergence, we invoke Theorem 2.11.22 of Van der Vaart and Wellner (1996). This requires verification of Conditions (2.11.21) and the convergence of the entropy integral in the statement of the theorem. Provided these conditions are satisfied, the sequence  $\sqrt{n} (\mathbb{P}_n - P) f_{n,h}$  is asymptotically tight in  $l^\infty[-K, K]$  and converges in distribution to a Gaussian process, the covariance kernel of which is given by:

$$K(s, t) = \lim_{n \rightarrow \infty} (P f_{n,s} f_{n,t} - P f_{n,s} P f_{n,t}).$$

We first compute  $P f_{n,s} f_{n,t}$ . It is easy to see that this is 0 if  $s$  and  $t$  are of opposite signs, so we need only consider the cases where they both have the same sign. So let  $s, t > 0$ . We have:

$$\begin{aligned} P f_{n,s} f_{n,t} &= \frac{E \left[ n^{1/3} (\phi'(X, \psi(Z)) - (\psi(Z) - \psi(z_0)) \phi''(X, \psi(Z)))^2 \mathbf{1}(Z \in (z_0, z_0 + (s \wedge t) n^{-1/3})) \right]}{(I(\psi(z_0)) p_Z(z_0))^2} \\ &= n^{1/3} \int_{z_0}^{z_0 + (s \wedge t) n^{-1/3}} H(z) p_Z(z) dz, \end{aligned}$$

where

$$H(z) = \frac{1}{(I(\psi(z_0)) p_Z(z_0))^2} E_{\psi(z)} (\phi'(X, \psi(z)) - (\psi(z) - \psi(z_0)) \phi''(X, \psi(z)))^2.$$

Now,  $H(z)$  converges to  $H(z_0)$  as  $z \rightarrow z_0$ . To see this, consider the function

$$\begin{aligned} G(\theta) &= E_{\theta} [\phi'(X, \theta) - (\theta - \theta_0) \phi''(X, \theta)]^2 \\ &= I(\theta) + (\theta - \theta_0)^2 E_{\theta} (\phi''(X, \theta))^2 - 2(\theta - \theta_0) E_{\theta} (\phi'(X, \theta) \phi''(X, \theta)) \\ &= I(\theta) + (\theta - \theta_0)^2 f_3(\theta) - 2(\theta - \theta_0) E_{\theta} (\phi'(X, \theta) \phi''(X, \theta)). \end{aligned}$$

As  $\theta \rightarrow \theta_0 \equiv \psi(z_0)$ , the first term converges to  $I(\theta_0)$  by (A.4) and the second term converges to 0, since  $f_3(\theta)$  is bounded in a neighborhood of  $\theta_0$  (by A.6). The third term also converges to 0, because, by the Cauchy–Schwarz inequality

$$| E_{\theta} (\phi'(X, \theta) \phi''(X, \theta)) | \leq \sqrt{I(\theta) f_3(\theta)},$$

which is bounded in a neighborhood of  $\theta_0$ . It follows that  $G(\theta)$  converges to  $I(\theta_0) \equiv G(\theta_0)$ . But  $H(z)$  is simply a constant times  $G(\psi(z))$  and by the continuity of  $\psi$  at  $z_0$  the result follows.

We conclude that

$$\lim_{n \rightarrow \infty} P f_{n,s} f_{n,t} = \frac{1}{(I(\psi(z_0)) p_Z(z_0))^2} H(z_0) p_Z(z_0) s \wedge t = \frac{1}{I(\psi(z_0)) p_Z(z_0)} s \wedge t,$$

on observing that  $H(z_0) = E_{\psi(z_0)} (\phi'(X, \psi(z_0)))^2 = I(\psi(z_0))$ . It is easily shown that  $P f_{n,s}$  and  $P f_{n,t}$  both converge to 0 as  $n \rightarrow \infty$ , showing that for  $s, t > 0$ ,  $K(s, t) = [I(\psi(z_0)) p_Z(z_0)]^{-1} s \wedge t$ . Similarly, we can show that  $K(s, t) = [I(\psi(z_0)) p_Z(z_0)]^{-1} |s| \wedge |t|$ , for  $s, t < 0$ . But this is the covariance kernel of the Gaussian process  $aW(h)$  with  $a = [I(\psi(z_0)) p_Z(z_0)]^{-1/2}$ . So the process  $\sqrt{n} (\mathbb{P}_n - P) f_{n,h}$  converges in  $l^\infty[-K, K]$  to the process  $aW(h)$ . We next show that  $\sqrt{n} P f_{n,h} \rightarrow (\psi'(z_0)/2) h^2$  uniformly on every  $[-K, K]$ . This implies that the process  $\tilde{B}_{n,\Psi}(h) \equiv \sqrt{n} \mathbb{P}_n f_{n,h}$  converges in distribution to  $X_{a,b}(h) \equiv aW(h) + bh^2$  in  $l^\infty[-K, K]$ . To show the convergence of  $\sqrt{n} P f_{n,h}$  to the desired limit, we restrict ourselves to the case where  $h > 0$ ; the case  $h < 0$  can be handled similarly. Let  $\xi_n(h) = I(\psi(z_0)) p_Z(z_0) \sqrt{n} P f_{n,h}$ . Then, we have,

$$\begin{aligned} \xi_n(h) &= n^{2/3} E \left\{ [(\psi(Z) - \psi(z_0)) \phi''(X, \psi(Z)) - \phi'(X, \psi(Z))] \mathbf{1}(z_0 < Z \leq z_0 + h n^{-1/3}) \right\} \\ &= n^{2/3} E \left[ (\psi(Z) - \psi(z_0)) \phi''(X, \psi(Z)) \mathbf{1}(z_0 < Z \leq z_0 + h n^{-1/3}) \right], \end{aligned}$$

on using the fact that  $E_{\psi(z)} \phi'(X, \psi(z)) = 0$ . It follows that

$$\begin{aligned}
\xi_n(h) &= n^{2/3} \int_{z_0}^{z_0+h n^{-1/3}} (\psi(z) - \psi(z_0)) E_{\psi(z)} (\phi''(X, \psi(z))) p_Z(z) dz \\
&= n^{1/3} \int_0^h (\psi(z_0 + u n^{-1/3}) - \psi(z_0)) I(\psi(z_0 + u n^{-1/3})) p_Z(z_0 + u n^{-1/3}) du \\
&= \int_0^h u \psi'(z_0) I(\psi(z_0 + u n^{-1/3})) p_Z(z_0 + u n^{-1/3}) du \\
&\quad + \int_0^h \left[ n^{1/3} (\psi(z_0 + u n^{-1/3}) - \psi(z_0)) - \psi'(z_0) u \right] I(\psi(z_0 + u n^{-1/3})) p_Z(z_0 + u n^{-1/3}) du.
\end{aligned}$$

The second term in the above display converges to 0 uniformly for  $0 \leq h \leq K$  on noting that:

$$\sup_{0 \leq h \leq K} \left| \frac{\psi(z_0 + u n^{-1/3}) - \psi(z_0)}{u n^{-1/3}} - \psi'(z_0) \right| \rightarrow 0.$$

The first term can be written as

$$\int_0^h u \psi'(z_0) I(\psi(z_0)) p_Z(z_0) du + o(1)$$

where  $o(1)$  goes to 0 uniformly over  $h \in [0, K]$  (by similar arguments). But

$$\int_0^h u \psi'(z_0) I(\psi(z_0)) p_Z(z_0) du = \frac{\psi'(z_0) I(\psi(z_0)) p_Z(z_0)}{2} h^2.$$

It follows that  $\sqrt{n} P f_{n,h} \rightarrow (\psi'(z_0)/2) h^2$  uniformly over  $0 \leq h \leq K$ .

We next check conditions (2.11.21). First, we construct a square integrable envelope function. It is easily checked that:

$$\begin{aligned}
f_{n,h} &\leq C_0 n^{1/6} \left[ |\phi'(X, \psi(Z)) - \phi''(X, \psi(Z)) (\psi(Z) - \psi(z_0))| \mathbf{1} \left( Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \right) \right] \\
&\leq \tilde{C}_0 n^{1/6} \left[ (|\phi'(X, \psi(Z))| + |\phi''(X, \psi(Z))|) \mathbf{1} \left( Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \right) \right],
\end{aligned}$$

where  $C_0$  and  $\tilde{C}_0$  are constants. Thus,

$$F_n = \tilde{C}_0 n^{1/6} \left[ (|\phi'(X, \psi(Z))| + |\phi''(X, \psi(Z))|) \mathbf{1} \left( Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \right) \right]$$

can be chosen as the envelope. Thus,

$$\begin{aligned}
F_n^2 &\leq 2 \tilde{C}_0^2 n^{1/3} (\phi'(X, \psi(Z))^2 + \phi''(X, \psi(Z))^2) \mathbf{1} \left( Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \right) \\
&= 2 \tilde{C}_0^2 n^{1/3} \int_{z_0 - K n^{-1/3}}^{z_0 + K n^{-1/3}} (I(\psi(z)) + f_3(\psi(z))) p_Z(z) dz;
\end{aligned}$$

that this is square integrable follows from the continuity of  $\psi$  at  $z_0$  and Assumptions (A.4) and (A.6).

Next to show that

$$E(F_n^2 \mathbf{1}\{F_n > \eta \sqrt{n}\}) \rightarrow 0$$

it suffices to show that for every  $\eta > 0$ ,

$$\begin{aligned} & n^{1/3} E \left[ (\phi'(X, \psi(Z))^2 + \phi''(X, \psi(Z))^2) \mathbf{1} \left( Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \right) \right. \\ & \quad \left. \times \left( \mathbf{1}(|\phi'(X, \psi(Z))| > \eta n^{1/6}/2) + \mathbf{1}(|\phi''(X, \psi(Z))| > \eta n^{1/6}/2) \right) \right] \rightarrow 0. \end{aligned}$$

Now, for any  $M > 0$  the sequence, say  $\xi_n$ , in the above display is eventually bounded by

$$\begin{aligned} & n^{1/3} E \left[ (\phi'(X, \psi(Z))^2 + \phi''(X, \psi(Z))^2) \mathbf{1} \left( Z \in [z_0 - K n^{-1/3}, z_0 + K n^{-1/3}] \right) \right. \\ & \quad \left. \times \left( \mathbf{1}(|\phi'(X, \psi(Z))| > M) + \mathbf{1}(|\phi''(X, \psi(Z))| > M) \right) \right] \rightarrow 0. \end{aligned}$$

This can be written as:

$$n^{1/3} \int_{z_0 - K n^{-1/3}}^{z_0 + K n^{-1/3}} H(\psi(z), M) p_Z(z) dz;$$

by (A.7) and the continuity of  $\psi$ , the above can be eventually made smaller than any pre-assigned  $\epsilon > 0$  by choosing  $M$  to be large enough. It follows that  $\limsup \xi_n \leq \epsilon$  for any given  $\epsilon > 0$  and hence equals 0.

Next we need to verify that

$$\sup_{|s-t| < \delta_n, -K \leq s, t \leq K} P(f_{n,s} - f_{n,t})^2 \rightarrow 0$$

if  $\delta_n \rightarrow 0$ . This is verified by straightforward computation. We restrict ourselves to the case where  $K > s > 0$  and  $-K < t < 0$  and  $s - t < \delta_n$ . Other cases are handled similarly. In this case, we can write:

$$\begin{aligned} P(f_{n,s} - f_{n,t})^2 &= C_0^2 n^{1/3} E \left[ (\phi'(X, \psi(Z)) - \phi''(X, \psi(Z)) (\psi(Z) - \psi(z_0)))^2 \right. \\ & \quad \left. \mathbf{1}\{Z \in [z_0 + t n^{-1/3}, z_0 + s n^{-1/3}]\} \right] \\ &\leq 2C_0^2 n^{1/3} \int_{z_0 + t n^{-1/3}}^{z_0 + s n^{-1/3}} (I(\psi(z)) + (\psi(z) - \psi(z_0))^2 f_3(\psi(z))) p_Z(z) dz. \end{aligned}$$

For all sufficiently large  $n$  (not depending on  $s$  and  $t$ )  $I(\psi(z)) + (\psi(z) - \psi(z_0))^2 f_3(\psi(z))$  is bounded by some constant  $\kappa$ ; consequently, for all sufficiently large  $n$  (not depending on  $s$  and  $t$ ),

$$P(f_{n,s} - f_{n,t})^2 \leq 2C_0^2 \kappa n^{1/3} \int_{z_0 + t n^{-1/3}}^{z_0 + s n^{-1/3}} p_Z(z) dz \leq 2C_0^2 \kappa \kappa' (s - t) \leq 2C_0^2 \kappa \kappa' \delta_n,$$

where  $\kappa'$  is an upper bound for  $p_Z(z)$  in a pre-fixed neighborhood of  $z_0$ . In the above display we have used the mean value theorem to arrive at the second inequality. Note that the last expression in the above display converges to 0 as  $n \rightarrow \infty$ .

It finally remains to verify the entropy integral condition. In other words, we need to check that

$$\sup_Q \int_0^{\delta_n} \sqrt{\log N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q))} d\epsilon \rightarrow 0 \quad \forall \delta_n \rightarrow 0, \quad (5.18)$$

where  $\mathcal{F}_n = \{f_{n,h} : h \in [-K, K]\}$ . Now,

$$N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q)) \leq N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_{n,1}, L_2(Q)) + N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_{n,2}, L_2(Q)).$$

Here  $\mathcal{F}_{n,1} = \{f_{n,h} : h \in [0, K]\}$  and  $\mathcal{F}_{n,2} = \{f_{n,h} : h \in [-K, 0]\}$ . Consider the class of functions

$$\mathcal{F}_{\delta_0} = \{[(\psi(Z) - \psi(z_0)) \phi''(X, \psi(Z)) - \phi'(X, \psi(Z))] 1(Z \in [z_0, z_0 + \delta]) : \delta \leq \delta_0\}.$$

Since  $\mathcal{F}_{\delta_0}$  is a fixed function times a class of indicator functions, it is a VC class with a fixed VC dimension  $V_0$ . Thus, any constant times  $\mathcal{F}_{\delta_0}$  also has VC dimension  $V_0$ . It is now easy to see that for all sufficiently large  $n$ ,  $\mathcal{F}_{n,1} \subset n^{1/6} \times \mathcal{F}_K$ ; it follows that each  $\mathcal{F}_{n,1}$  is a VC class with VC-dimension bounded above by  $V_0 (\geq 2)$ . It is well known (see, for example, Theorem 2.6.7 of Van der Vaart and Wellner (1996)) that

$$\log N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_{n,1}, L_2(Q)) \leq K V_n (16e)^{V_n} \left(\frac{1}{\epsilon}\right)^{2(V_n-1)},$$

for some universal constant  $K$ . Here  $V_n$  is the VC-dimension of  $\mathcal{F}_{n,1}$ . But  $V_n \leq V_0$  and hence, the above inequality implies

$$\log N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_{n,1}, L_2(Q)) \leq \tilde{K} \left(\frac{1}{\epsilon}\right)^s$$

where  $s = 2(V_0 - 1) < \infty$  and  $\tilde{K}$  is a universal constant not depending upon  $n$  and  $Q$ . A similar bound applies to  $\log N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_{n,2}, L_2(Q))$ . It follows that for a sufficiently large integer  $s' \geq 1$  we can ensure that

$$N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q)) \leq K^* \left(\frac{1}{\epsilon}\right)^{s'},$$

for some universal constant  $K^*$ . To check (5.18) it therefore suffices to check that

$$\int_0^{\delta_n} \sqrt{-\log \epsilon} d\epsilon \rightarrow 0$$

as  $\delta_n \rightarrow 0$ . But this is trivial.  $\square$

**Proof of Lemma 2.4:** We only prove the first assertion. The second one follows similarly. For

the first assertion, we write the proof for  $h > 0$ ; the proof for  $h < 0$  is similar. So, let  $0 \leq h \leq K$ . Recall that,

$$\tilde{B}_{n,\hat{\psi}_n}(h) = C n^{2/3} \mathbb{P}_n \left[ \left\{ (\hat{\psi}_n(Z) - \psi(z_0)) \phi''(X, \psi(Z)) - \phi'(X, \psi(Z)) \right\} 1(Z \in (z_0, z_0 + h n^{-1/3})) \right],$$

where  $C$  is a constant, and  $\tilde{B}_{n,\psi}(h)$  has the same form as above but with  $\hat{\psi}_n$  replaced by  $\psi$ . Now, for any  $Z \in (z_0, z_0 + K n^{-1/3}]$  we can write:

$$\phi'(X, \psi(z_0)) = \phi'(X, \psi(Z)) + \phi''(X, \psi(Z)) (\psi(z_0) - \psi(Z)) + \frac{1}{2} \phi'''(X, \psi^*(Z)) (\psi(Z) - \psi(z_0))^2,$$

where  $\psi^*(Z)$  is a point between  $\psi(Z)$  and  $\psi(z_0)$ . Also,

$$\phi'(X, \psi(z_0)) = \phi'(X, \hat{\psi}_n(Z)) + \phi''(X, \hat{\psi}_n(Z)) (\psi(z_0) - \hat{\psi}_n(Z)) + \frac{1}{2} \phi'''(X, \hat{\psi}_n^*(Z)) (\hat{\psi}_n(Z) - \psi(z_0))^2,$$

where  $\hat{\psi}_n^*(Z)$  is a point between  $\hat{\psi}_n(Z)$  and  $\psi(z_0)$ . It follows that we can write:

$$\begin{aligned} \tilde{B}_{n,\psi}(h) - \tilde{B}_{n,\hat{\psi}_n}(h) &= C \frac{1}{2} \mathbb{P}_n \left[ (n^{1/3} (\psi(Z) - \psi(z_0)))^2 \phi'''(X, \psi^*(Z)) 1(Z \in (z_0, z_0 + h n^{-1/3})) \right] \\ &\quad - C \frac{1}{2} \mathbb{P}_n \left[ (n^{1/3} (\psi(Z) - \psi(z_0)))^2 \phi'''(X, \hat{\psi}_n^*(Z)) 1(Z \in (z_0, z_0 + h n^{-1/3})) \right]. \end{aligned}$$

We will show that the second term in the above display converges to 0 uniformly in  $h$ ; the proof for the first term is similar. Up to a constant, the second term is bounded in absolute value by:

$$\mathbb{P}_n \left[ (n^{1/3} (\hat{\psi}_n(Z) - \psi(z_0)))^2 | \phi'''(X, \hat{\psi}_n^*(Z)) | 1(Z \in (z_0, z_0 + K n^{-1/3})) \right] \equiv \mathbb{P}_n(\xi_n) \quad (\star\star).$$

For any  $z \in (z_0, z_0 + K n^{-1/3}]$ , we have

$$\left[ n^{1/3} (\hat{\psi}_n(z) - \psi(z_0)) \right]^2 \leq (n^{1/3} (\hat{\psi}_n(z_0 + K n^{-1/3}) - \psi(z_0)))^2$$

which, with arbitrarily high probability is eventually bounded by a constant  $C$  (by Lemma 2.2). Also, since for any such  $z$ ,  $\hat{\psi}_n^*(z)$  lies between  $\hat{\psi}_n(z_0 + K n^{-1/3})$  and  $\psi(z_0)$  to which the former converges in probability, with arbitrarily high probability  $| \phi'''(X, \hat{\psi}_n^*(Z)) |$  is eventually bounded by  $B(X)$  (by Assumption (A.5)). It follows that with arbitrarily high probability, the random function  $\xi_n$  in  $(\star\star)$  is eventually bounded up to a constant by  $B(X) 1(Z \in [z_0, z_0 + K n^{-1/3}])$ . It follows that eventually, with arbitrarily high probability,

$$\begin{aligned} \mathbb{P}_n(\xi_n) &\leq \tilde{C} (\mathbb{P}_n - P) \left[ B(X) 1(Z \in [z_0, z_0 + K n^{-1/3}]) \right] \\ &\quad + \tilde{C} P \left[ C B(X) 1(Z \in [z_0, z_0 + K n^{-1/3}]) \right], \end{aligned}$$

for some constant  $\tilde{C}$ . The first term on the right side is  $o_p(1)$  using straightforward Glivenko-Cantelli type arguments and the second term is seen to go to 0 by direct computation. This shows

that the second term goes to 0 uniformly in  $h$ .  $\square$

**Comment on Lemma 2.5:** The proof of this lemma is skipped. The arguments used are very similar to those from Lemma 2.4; hence we provide only an outline. We can write

$$\tilde{G}_{n,\hat{\psi}_n}(h) = \tilde{G}_{n,\hat{\psi}_n}(h) - \tilde{G}_{n,\psi(h)} + \tilde{G}_{n,\psi}(h).$$

The third term can be shown to converge in probability uniformly to  $h$  on  $[-K, K]$  whereas the difference of the first two terms converges to 0 uniformly, under our assumptions. A similar proof works for  $\tilde{G}_{n,\hat{\psi}_n^0}(h)$ .

**Proof of Lemma 2.6:** Firstly note that  $D_n$  is either the null set, or it is an interval containing the point  $z_0$ . Let  $\tilde{D}_n$  denote the set  $n^{1/3}(D_n - z_0)$ . It suffices to show that given  $\epsilon > 0$  there exists  $M > 0$  such that  $P(\tilde{D}_n \subset [-M, M]) > 1 - \epsilon$  eventually.

To prove this, proceed in the following way. Let  $\tilde{D}_n = [\tilde{A}_n, \tilde{B}_n)$ . Now the event  $\{\tilde{D}_n \subset [-M, M]\}$  is the same as  $\{-M < \tilde{A}_n \leq \tilde{B}_n < M\}$ . It suffices to show that  $P(\tilde{B}_n < M) > 1 - \epsilon/2$  and  $P(-M < \tilde{A}_n) > 1 - \epsilon/2$  eventually for  $M$  sufficiently large. We shall prove the first assertion; the second follows similarly. To prove the first assertion it suffices to show that  $P(\tilde{B}_n > M) < \epsilon/2$  for  $M$  sufficiently large. But  $\tilde{B}_n > M$  means that  $z_0 + M n^{-1/3}$  is in the difference set  $D_n$  and this implies that either  $\hat{\psi}_n^0(z_0 + M n^{-1/3}) = \theta_0$  or  $\hat{\psi}_n(z_0 + M n^{-1/3}) = \hat{\psi}_n(z_0)$  by (2.7). This observation can be written in terms of the processes  $X_n$  and  $Y_n$  in the following way:

$$\{\tilde{B}_n > M\} \subset \{Y_n(M) = 0\} \cup \{X_n(0) = X_n(M)\}. \quad (5.19)$$

Now

$$P(Y_n(M) = 0) \rightarrow P(g_{a,b,R}(M) \leq 0)$$

by Theorem 2.1 where  $g_{a,b,R}(M)$  is the right derivative of the greatest convex minorant of the process  $aW(h) + bh^2$  on  $[0, \infty)$ . By choosing  $M$  large enough we can ensure that the probability on the right of the above display is strictly less than  $\epsilon/4$ . Consider now  $P(X_n(0) = X_n(M))$ . This is the same as  $P(X_n(0) - X_n(M) = 0)$ . Now by Theorem 2.1 again, we have

$$(X_n(0), X_n(M)) \rightarrow_d (g_{a,b}(0), g_{a,b}(M))$$

so that

$$\limsup P(X_n(0) - X_n(M) = 0) \leq P(g_{a,b}(0) - g_{a,b}(M) = 0)$$

and the right hand side is again less than  $\epsilon/4$  for  $M$  sufficiently large. It now follows from (5.19) that

$$P(\tilde{B}_n > M) < \epsilon/2$$

eventually.  $\square$

**Alternative proof of Theorem 2.1:** The alternative proof of this theorem relies on

extensive use of “switching relationships” which allow us to translate the behavior of the slope of the convex minorant of a random cumulative sum diagram (this is how the estimators  $\hat{\psi}_n$  and  $\hat{\psi}_n^0$  are characterized) in terms of the minimizer of a stochastic process. The limiting behavior of the slope process can then be studied in terms of the limiting behavior of this minimizer through argmax continuous mapping theorems. Switching relationships on the limit process then enable interpretation of the behavior of the minimizer of the limit process in terms of the slope of the convex minorant of the limiting version of the cumulative sum diagrams (appropriately normalized).

For the unconstrained MLE the switching relationship implies the following:

$$\hat{\psi}_n(z) \leq a \Leftrightarrow \operatorname{argmin}_{r \geq 0} [B_n(r) - a G_n(r)] \geq Z_z \quad (5.20)$$

where  $Z_z$  is the largest element in the set  $\{Z_i\}_{i=1}^n$  not exceeding  $z$ . By argmin we denote the largest element in the set of minimizers. This can be chosen to be one of the  $Z_i$ 's. The above equivalence is a direct characterization of the fact that the vector  $\{\hat{\psi}_n(Z_{(i)})\}_{i=1}^n$  is the vector of slopes (left-derivatives) of the cumulative sum diagram formed by the points  $\{G_n(Z_{(i)}), B_n(Z_{(i)})\}_{i=0}^n$ , computed at the points  $\{G_n(Z_{(i)})\}_{i=1}^n$ . The easiest way to verify this is by drawing a picture.

For the constrained MLE the switching relationship is slightly more involved. We need to isolate different cases.

(1) For  $z \leq z_0$  and  $a < \theta_0$ , we have:

$$\hat{\psi}_n^0(z) \leq a \Leftrightarrow \operatorname{argmin}_{r \leq z_0} [B_{n,0}(r) - a G_{n,0}(r)] \geq Z_z. \quad (5.21)$$

(2) For  $z \leq z_0$  and  $a = \theta_0$ , we have:

$$\hat{\psi}_n^0(z) = \theta_0 \Leftrightarrow \operatorname{Argmin}_{r \leq z_0} [B_{n,0}(r) - \theta_0 G_{n,0}(r)] < Z_z. \quad (5.22)$$

Here Argmin denotes the smallest element in the set of minimizers.

(3) For  $z \geq z_0$  and  $a > \theta_0$ , we have:

$$\hat{\psi}_n^0(z) \geq a \Leftrightarrow \operatorname{Argmin}_{r \geq z_0} [B_{n,0}(r) - a G_{n,0}(r)] < Z_z. \quad (5.23)$$

(4) For  $z \geq z_0$  and  $a = \theta_0$ , we have:

$$\hat{\psi}_n^0(z) = \theta_0 \Leftrightarrow \operatorname{argmin}_{r \geq z_0} [B_{n,0}(r) - \theta_0 G_{n,0}(r)] \geq Z_z. \quad (5.24)$$

We first prove finite dimensional convergence of the processes  $X_n(h)$  and  $Y_n(h)$ . To this end, let  $t_1, t_2, \dots, t_k, t_{k+1}, \dots, t_{k'}$  be strictly negative numbers. Let  $x_1, x_2, \dots, x_{k'}$  be real numbers and let  $y_1, y_2, \dots, y_k$  be strictly negative numbers. Let  $s_1, s_2, \dots, s_l, s_{l+1}, \dots, s_{l'}$  be strictly positive numbers. Let  $z_1, z_2, \dots, z_{l'}$  be real numbers and let  $v_1, v_2, \dots, v_{l'}$  be strictly positive numbers.

Next, define the following events:

$$A_{i,n} = \{n^{1/3} (\hat{\psi}_n(z_0 + t_i n^{-1/3}) - \psi(z_0)) \leq x_i\} \text{ for } i = 1, 2, \dots, k',$$

$$B_{i,n} = \{n^{1/3} (\hat{\psi}_n^0(z_0 + t_i n^{-1/3}) - \psi(z_0)) \leq y_i\} \text{ for } i = 1, 2, \dots, k,$$

and

$$B_{i,n} = \{n^{1/3} (\hat{\psi}_n^0(z_0 + t_i n^{-1/3}) - \psi(z_0)) = 0\} \text{ for } i > k.$$

Next,

$$C_{i,n} = \{n^{1/3} (\hat{\psi}_n(z_0 + s_i n^{-1/3}) - \psi(z_0)) \leq z_i\} \text{ for } i = 1, 2, \dots, l',$$

$$D_{i,n} = \{n^{1/3} (\hat{\psi}_n^0(z_0 + s_i n^{-1/3}) - \psi(z_0)) \geq v_i\} \text{ for } i = 1, 2, \dots, l,$$

and

$$D_{i,n} = \{n^{1/3} (\hat{\psi}_n^0(z_0 + s_i n^{-1/3}) - \psi(z_0)) = 0\} \text{ for } i > l.$$

We will compute the limiting value of the probability

$$P(\{A_{i,n}\}_{i=1}^{k'}, \{B_{i,n}\}_{i=1}^{k'}, \{C_{i,n}\}_{i=1}^{l'}, \{D_{i,n}\}_{i=1}^{l'}). \quad (\star)$$

Now, consider the event  $A_{i,n}$ . We have

$$\begin{aligned} n^{1/3} ((\hat{\psi}_n(z_0 + t_i n^{-1/3}) - \psi(z_0)) \leq x_i &\Leftrightarrow \hat{\psi}_n(z_0 + t_i n^{-1/3}) \leq \psi(z_0) + x_i n^{-1/3} \\ &\Leftrightarrow \operatorname{argmin}_r \left[ B_n(r) - (\psi(z_0) + x_i n^{-1/3}) G_n(r) \right] \geq Z_{(z_0 + t_i n^{-1/3})} \\ &\Leftrightarrow \operatorname{argmin}_r \left[ V_n(r) - x_i n^{-1/3} G_n(r) \right] \geq Z_{(z_0 + t_i n^{-1/3})}, \end{aligned}$$

where  $V_n(r) = B_n(r) - \psi(z_0) G_n(r)$ . Note that we have used (5.20) in the display above. Thus, for  $i = 1, 2, \dots, k'$ ,

$$\begin{aligned} A_{i,n} &= \left\{ n^{1/3} \left( \operatorname{argmin}_r \left[ V_n(r) - x_i n^{-1/3} G_n(r) \right] - z_0 \right) \geq n^{1/3} (Z_{(z_0 + t_i n^{-1/3})} - z_0) \right\} \\ &= \left\{ \operatorname{argmin}_h V_n(z_0 + h n^{-1/3}) - x_i n^{-1/3} G_n(z_0 + h n^{-1/3}) \geq t_i + o_p(1) \right\} \\ &= \left\{ \operatorname{argmin}_h \mathbb{M}_n(h) - x_i \mathbb{G}_n(h) \geq t_i + o_p(1) \right\} \\ &\equiv \left\{ \operatorname{argmin}_h P_{A_{i,n}}(h) \geq t_i + o_p(1) \right\}, \end{aligned}$$

where

$$\mathbb{M}_n(h) = \frac{1}{I(\psi(z_0)) p_Z(z_0)} n^{2/3} \left[ V_n(z_0 + h n^{-1/3}) - V_n(z_0) \right] \equiv \tilde{B}_{n, \hat{\psi}_n}(h)$$

and

$$\mathbb{G}_n(h) = \frac{1}{I(\psi(z_0)) p_Z(z_0)} n^{1/3} \left[ G_n(z_0 + h n^{-1/3}) - G_n(z_0) \right] \equiv \tilde{G}_{n, \hat{\psi}_n}(h).$$

Similarly, using (5.21) for  $i = 1, 2, \dots, k$ ,

$$B_{i,n} = \{\operatorname{argmin}_{h \leq 0} \mathbb{M}_n^0(h) - x_i \mathbb{G}_n^0(h) \geq t_i + o_p(1)\} \equiv \{\operatorname{argmin}_{h \leq 0} P_{B_{i,n}}(h) \geq t_i + o_p(1)\}$$

where

$$\mathbb{M}_n^0(h) = \frac{1}{I(\psi(z_0)) p_Z(z_0)} n^{2/3} \left[ V_{n,0}(z_0 + h n^{-1/3}) - V_{n,0}(z_0) \right] \equiv \tilde{B}_{n, \hat{\psi}_n^0}(h),$$

and

$$\mathbb{G}_n^0(h) = \frac{1}{I(\psi(z_0)) p_Z(z_0)} n^{1/3} \left[ G_{n,0}(z_0 + h n^{-1/3}) - G_{n,0}(z_0) \right] \equiv \tilde{G}_{n, \hat{\psi}_n^0}(h)$$

Using (5.22), for  $i > k$ ,

$$B_{i,n} = \{\operatorname{Argmin}_{h \leq 0} \mathbb{M}_n^0(h) < t_i + o_p(1)\} \equiv \{\operatorname{Argmin}_{h \leq 0} P_{C_{i,n}}(h) < t_i + o_p(1)\}.$$

Using (5.20), for  $i = 1, 2, \dots, l'$ ,

$$C_{i,n} = \{\operatorname{argmin}_h \mathbb{M}_n^0(h) - z_i \mathbb{G}_n^0(h) \geq s_i + o_p(1)\} \equiv \{\operatorname{argmin}_h P_{C_{i,n}}(h) \geq s_i + o_p(1)\}.$$

Using (5.23), for  $i = 1, 2, \dots, l$ ,

$$D_{i,n} = \{\operatorname{Argmin}_{h \geq 0} \mathbb{M}_n^0(h) - v_i \mathbb{G}_n^0(h) < s_i + o_p(1)\} \equiv \{\operatorname{Argmin}_{h \geq 0} P_{D_{i,n}}(h) < s_i + o_p(1)\}$$

and for  $i > l$ ,

$$D_{i,n} = \{\operatorname{argmin}_{h \geq 0} \mathbb{M}_n^0(h) \geq s_i + o_p(1)\} \equiv \{\operatorname{argmin}_{h \geq 0} P_{D_{i,n}}(h) \geq s_i + o_p(1)\}.$$

Next, note that the processes

$$\left( \{P_{A_{i,n}}(h)\}_{i=1}^{k'}, \{P_{B_{i,n}}(h)\}_{i=1}^{k'}, \{P_{C_{i,n}}(h)\}_{i=1}^{l'}, \{P_{D_{i,n}}(h)\}_{i=1}^{l'} \right)$$

converge jointly in the space  $l^\infty[-M, M]^{2(l'+k')}$  for every  $M > 0$  to the processes

$$\left( \{P_{A_i}(h)\}_{i=1}^{k'}, \{P_{B_i}(h)\}_{i=1}^{k'}, \{P_{C_i}(h)\}_{i=1}^{l'}, \{P_{D_i}(h)\}_{i=1}^{l'} \right), \quad (\star\star)$$

where

$$P_{A_i}(h) = a W(h) + b h^2 - x_i h, \quad i = 1, 2, \dots, k',$$

$$P_{B_i}(h) = a W(h) + b h^2 - y_i h, \quad i = 1, 2, \dots, k,$$

$$P_{B_i}(h) = a W(h) + b h^2, \quad i = k + 1, \dots, k',$$

$$P_{C_i}(h) = a W(h) + b h^2 - z_i h, \quad i = 1, 2, \dots, l',$$

$$P_{D_i}(h) = a W(h) + b h^2 - v_i h, \quad i = 1, 2, \dots, l,$$

and

$$P_{D_i}(h) = a W(h) + b h^2, \quad i = l + 1, \dots, l'.$$

This result is obtained by using the fact that the processes  $\mathbb{M}_n(h)$  and  $\mathbb{M}_n^0(h)$  converge (jointly) to the same limiting process  $aW(h) + bh^2$  under the topology of uniform convergence on compact sets, by Lemmas 2.3 and 2.4. Moreover, the processes  $\mathbb{G}_n(h)$  and  $\mathbb{G}_n^0(h)$  converge uniformly in probability on every  $[-M, M]$  to the deterministic process  $h$ , by Lemma 2.5.

Consequently the vector,

$$\left( \{\operatorname{argmin}_h P_{A_{i,n}}(h)\}_{i=1}^{k'}, \{\operatorname{argmin}_{h \leq 0} P_{B_{i,n}}(h)\}_{i=1}^k, \{\operatorname{Argmin}_{h \leq 0} P_{B_{i,n}}(h)\}_{i=k+1}^{k'}, \right. \\ \left. \{\operatorname{argmin}_h P_{C_{i,n}}(h)\}_{i=1}^{l'}, \{\operatorname{Argmin}_{h \geq 0} P_{B_{i,n}}(h)\}_{i=1}^l, \{\operatorname{argmin}_{h \leq 0} P_{B_{i,n}}(h)\}_{i=l+1}^{l'} \right),$$

converges to the continuous random vector  $Z$  given by:

$$\left( \{\operatorname{argmin}_h P_{A_i}(h)\}_{i=1}^{k'}, \{\operatorname{argmin}_{h \leq 0} P_{B_i}(h)\}_{i=1}^{k'}, \{\operatorname{argmin}_h P_{C_i}(h)\}_{i=1}^{l'}, \{\operatorname{argmin}_{h \geq 0} P_{D_i}(h)\}_{i=1}^{l'} \right).$$

The above convergence is accomplished by appealing to an appropriate continuous mapping theorem for minimizers of stochastic processes in  $B_{loc}(\mathbb{R})$  (see, for example, Theorem 6.1 of Huang and Wellner (1995) or Lemma 3.6.10 of Banerjee(2000)). The key facts that guarantee the convergence of the minimizers are (i) the fact that almost surely, the limiting processes in  $(\star\star)$  restricted to  $\mathbb{R}$  or  $\mathbb{R}^-$  or  $\mathbb{R}^+$  diverge to  $\infty$  as the argument diverges to  $\infty$  in absolute value and possess a unique minimizer (this follows from Lemmas 2.5 and 2.6 of Kim and Pollard (1990)) and (ii) the minimizers of the converging (finite sample) processes are tight. Establishing (ii) involves application of an appropriate *rate theorem* for minimizers of stochastic processes (see, for example Theorem 3.2.5 or Theorem 3.4.1 of Van der Vaart and Wellner (1996)). The computations are tedious but straightforward and skipped here. For a flavor of the key steps involved in establishing tightness, we refer the reader to Sections 3.2.2 and 3.2.3 of Van der Vaart and Wellner (1996).

By Slutsky's theorem, the limiting value of the probability  $(\star)$  is,

$$P \left( \begin{array}{l} \operatorname{argmin}_{\mathbb{R}} [aW(h) + bh^2 - x_i h] \geq t_i, \quad i = 1, \dots, k' \\ \operatorname{argmin}_{\mathbb{R}^-} [aW(h) + bh^2 - y_i h] \geq t_i, \quad i = 1, \dots, k \\ \operatorname{argmin}_{\mathbb{R}^-} [aW(h) + bh^2] < t_i, \quad i = k + 1, \dots, k' \\ \operatorname{argmin}_{\mathbb{R}} [aW(h) + bh^2 - z_i h] \geq s_i, \quad i = 1, \dots, l' \\ \operatorname{argmin}_{\mathbb{R}^+} [aW(h) + bh^2 - v_i h] < s_i, \quad i = 1, \dots, l \\ \operatorname{argmin}_{\mathbb{R}^+} [aW(h) + bh^2] \geq s_i, \quad i = l + 1, \dots, l' \end{array} \right).$$

We now use the switching relationships on the limit processes. It is well known that:

$$\operatorname{argmin}_{\mathbb{R}} aW(h) + bh^2 - ch > t \Leftrightarrow g_{a,b}(t) < c,$$

with probability one. Also, with probability 1, we have:

$$\operatorname{argmin}_{\mathbb{R}^-} aW(h) + bh^2 - ch > t \Leftrightarrow g_{a,b,L}(t) < c,$$

$$\operatorname{argmin}_{\mathbb{R}^+} a W(h) + b h^2 - c h > t \Leftrightarrow g_{a,b,R}(t) < c.$$

Using the above switching relationships on the limiting processes and keeping in mind that the random variables involved are all continuous, we obtain the limiting probability as

$$P \left( \begin{array}{l} g_{ab}(t_i) \leq x_i, \quad i = 1, \dots, k' \\ g_{a,b,L}(t_i) \leq y_i, \quad i = 1, \dots, k \\ g_{a,b,L}(t_i) \geq 0, \quad i = k+1, \dots, k' \\ g_{a,b}(s_i) \leq z_i, \quad i = 1, \dots, l' \\ g_{a,b,R}(s_i) \geq v_i, \quad i = 1, \dots, l \\ g_{a,b,R}(s_i) \leq 0, \quad i = l+1, \dots, l' \end{array} \right).$$

On using the fact that the  $y_i$ 's are strictly negative and that the  $v_i$ ' are strictly positive, and the relationship between  $g_{a,b,L}, g_{a,b,R}$  and  $g_{a,b}^0$  the above expression can be simply written as:

$$P \left( \begin{array}{l} g_{ab}(t_i) \leq x_i, \quad i = 1, \dots, k' \\ g_{a,b}^0(t_i) \leq y_i, \quad i = 1, \dots, k \\ g_{a,b}^0(t_i) = 0, \quad i = k+1, \dots, k' \\ g_{a,b}(s_i) \leq z_i, \quad i = 1, \dots, l' \\ g_{a,b}^0(s_i) \geq v_i, \quad i = 1, \dots, l \\ g_{a,b}^0(s_i) = 0, \quad i = l+1, \dots, l' \end{array} \right).$$

This completes the proof of finite-dimensional convergence. To deduce the convergence in  $\mathcal{L} \times \mathcal{L}$ , note that  $X_n(h)$  and  $Y_n(h)$  are monotone functions. For a sequence  $(\psi_n, \phi_n)$  in  $L_2[-K, K] \times L_2[-K, K]$  satisfying (a)  $\psi_n$  and  $\phi_n$  are monotone functions and (b) for all vectors  $(h_1, h_2, \dots, h_k)$

$$(\psi_n(h), \phi_n(h)) |_{h=h_1, h_2, \dots, h_k} \rightarrow (\psi(h), \phi(h)) |_{h=h_1, h_2, \dots, h_k}$$

for monotone functions  $(\psi, \phi)$  in  $L_2[-K, K] \times L_2[-K, K]$ , it follows that  $(\psi_n, \phi_n) \rightarrow (\psi, \phi)$  in  $L_2[-K, K] \times L_2[-K, K]$ . In the wake of convergence of all the finite - dimensional marginals of  $(X_n, Y_n)$  to those of  $(g_{a,b}(h), g_{a,b}^0(h))$ , it follows that

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

in  $L_2[-K, K] \times L_2[-K, K]$  for every  $K > 0$  and consequently, in the space,  $\mathcal{L} \times \mathcal{L}$  (this parallels the result of Corollary 2 following Theorem 3 of Huang and Zhang (1994)).  $\square$

## References

- Banerjee, M. (2000). *Likelihood Ratio Inference in Regular and Nonregular Problems*. Ph.D. dissertation, University of Washington.
- Banerjee, M. and Wellner, J. A. (2001a). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699–1731.

- Brunk, H.D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference*. M.L. Puri, ed.
- Diggle, P., Morris, S. and Morton–Jones, T. (1999) Case–control isotonic regression for investigation of elevation in risk around a risk source. *Statistics in Medicine*, **18**, 1605–1613.
- Dunson, D.B., (2004) Bayesian isotonic regression for discrete outcomes. *Working Paper*, available at <http://ftp.isds.duke.edu/WorkingPapers/03-16.pdf>
- Ghosh, D., Banerjee, M. and Biswas, P. (2004) Binary isotonic regression procedures, with applications to cancer biomarkers. Technical report available on <http://www.stat.lsa.umich.edu/~moulib/gbb5.pdf>
- Groeneboom, P. (1986). *Some Current Developments in Density Estimation*. CWI Monographs, Amsterdam, **1**, 163 – 192.
- Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields* **81**, 79 - 109.
- Groeneboom, P. and Wellner J.A. (1992). *Information Bounds and Nonparametric Likelihood Estimation*. Birkhäuser, Basel.
- Groeneboom, P. and Wellner J.A. (2001). Computing Chernoff’s distribution. *Journal of Computational and Graphical Statistics*. **10**, 388-400.
- Huang, Y. and Zhang, C. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **24**, 1256 – 1274.
- Huang, J. (1994). *Estimation in Regression Models with Interval Censoring*. Ph.D. dissertation, University of Washington.
- Huang, J. (1996) Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics*, **24**, No. 2, 540–568.
- Huang, J. (2002) A note on estimating a partly linear model under monotonicity constraints. *Journal of Statistical Planning and Inference*, **107**, 343-351.
- Huang, J. and Wellner, J. (1995) . Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Stat.* **22** , 3 - 33.
- Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7**, 310-321.
- Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.
- Mammen, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19**, 724-740.
- Marshall, A.W. and Proschan, F. (1965). Maximum Likelihood Estimation for distributions with monotone failure rate. *Ann. Math. Statist.*, **36** , 69 - 77.
- Morton–Jones, T., Diggle, P. and Elliott, P. (1999) Investigation of excess environment risk around putative sources: Stone’s test with covariate adjustment. *Statistics in Medicine*, **18**, 189 – 197.

- Mammen, E. (1988). Monotone nonparametric regression. *Ann. Statist.* **16**, 741-750.
- Murphy, S.A. and Van der Vaart, A.W. (1997). Semiparametric Likelihood Ratio Inference. *Ann. Statist.* **25**, 1471 – 1509.
- Murphy, S. and Van der Vaart, A. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** , 449 - 465.
- Newton, M.A., Czado, C. and Chappell, R. (1996) Bayesian inference for semiparametric binary regression. *JASA*, **91**, 142-153.
- Politis, D.M., Romano, J.P., and Wolf, M. (1999) *Subsampling*, Springer-Verlag, New York.
- Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhya. Ser. A*, **31** , 23 - 36.
- Prakasa Rao, B.L.S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Statist.*, **36** , 69 - 77.
- Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215-1230.
- Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Wiley, New York
- Salanti, G. and Ulm, K. (2003) Tests for trend with binary response. *Biometrical Journal*, **45**, 277-291.
- Shiboski, S. (1998) Generalized additive models with current status data. *Lifetime Data Analysis*, **4**, 29 – 50.
- Stone, R. A. (1988) Investigations of excess environmental risks around putative sources: Statistical problems and a proposed test. *Statistics In Medicine*, **7**, 649–660.
- Sun, J. and Kalbfleisch, J.D. (1993) The analysis of current status data on point processes. *JASA*, **88**, 1449-1454.
- Sun, J. and Kalbfleisch, J.D. (1995) Estimation of the mean function of a point process based on panel count data. *Statistica Sinica*, **5**, 279–290.
- Sun, J. (1999) A nonparametric test for current status data with unequal censoring. *Jour. of the Royal Stat. Soc. B*, **61**, 243–250.
- Van der Vaart, A. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- Wellner, J. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.
- Wilks, S.S (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **19** , 60–62.