Likelihood Based Inference For Monotone Response Models

Moulinath Banerjee [1]

University of Michigan

May 1, 2006

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 framework for likelihood based estimation of monotone functions is developed and 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 χ^2 but can be explicitly characterized in terms of a functional of Brownian motion.

1 Introduction

A common problem in nonparametric statistics is the need to estimate a function, like a density, a distribution, a hazard or a regression function. 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 function of distance from a toxic source), a monotone density (inference in renewal theory and other applications) or a monotone hazard rate (reliability).

[1] Research supported in part by National Science Foundation grant DMS-0306235

Key words and phrases. greatest convex minorant, ICM, likelihood ratio statistic, monotone function, monotone response model, self–induced characterization, two–sided Brownian motion, universal limit, Wilks' phenomenon

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 ψ, then provided that $\psi'(z)$ does not vanish,

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

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, $\mathbb{Z} = \text{argmin}_h\,(W(h) + h^2)$, 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).

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 $W_1, W_2, \ldots, W_n$ are i.i.d. observations from a regular parametric model $p(x, \theta)$, and $\hat{\theta}_n$ denotes the MLE of θ based on this data, then $\sqrt{n}\,(\hat{\theta}_n - \theta) \rightarrow_d (I(\theta))^{-1/2}\,N$ where N follows $N(0,1)$ and $I(\theta)$ is the Fisher information. In monotone function estimation, the $\sqrt{n}$ rate slows down to $n^{1/3}$ and $\mathbb{Z}$ plays the role of N.

In this paper, we study a class of conditionally parametric models, of the covariate–response type, where the conditional distribution of the response given the covariate comes from a regular parametric model, with the parameter being given by a monotone function of the value of the covariate. We call these *monotone response models*. Here is a formal description:

Let $\{p(x, \theta) : \theta \in \Theta\}$ with Θ being an open subset of $\mathbb{R}$, be a one–parameter family of probability densities with respect to a dominating measure μ. Let ψ be an increasing or decreasing continuous function defined on an interval $\tilde{I}$ and taking values in Θ. Consider i.i.d. data $\{(X_i, Z_i)\}_{i=1}^n$ where

$Z_i \sim p_Z$, p_Z being a Lebesgue density defined on $\tilde{I}$ and $X_i \mid Z_i = z \sim p(x, \psi(z))$. We think of X_i as the covariate value for the i'th observation, and X_i as the value of the response. Interest focuses on estimating the function ψ, since it captures the nature of dependence between the response and the covariate. If the parametric family of densities, $p(x, \theta)$, is parametrized by its mean, then $\psi(z) = E(X \mid Z = z)$ is precisely the regression function. It will be seen later that when $p(x, \theta)$ is a one-parameter full rank exponential family model, the regression function $\tilde{B}(z) \equiv E(T(X) \mid Z = z)$ corresponding to the sufficient statistic $T(x)$, is a strictly increasing (known) transformation of ψ; consequently, estimating ψ is equivalent to estimating the regression of $T(X)$ on Z. In this paper, we study the asymptotics of the MLE of ψ and also the likelihood ratio statistic for testing ψ at a pre-fixed point of interest, with a view to obtaining pointwise confidence sets for ψ of a pre-assigned level of significance. Before we discuss this further, here are some motivating examples to illustrate the above framework.

(i) **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, ϵ_i is independent of Z_i, each ϵ_i has mean 0 and variance σ^2, each Z_i has a Lebesgue density $p_Z(\cdot)$ and ψ 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 (1988), Mammen (1991), Huang (2002)). Now suppose that the ϵ_i's are Gaussian. We are then in the above framework: $Z \sim p_Z(\cdot)$ and $X \mid Z = z \sim N(\psi(z), \sigma^2)$. We want to estimate ψ and test $\psi(z_0) = \theta_0$ for an interior point z_0 in the domain of ψ.

(ii) **Binary Choice Model:** Here we have a dichotomous response variable $X = 1$ or 0 and a continuous covariate Z with a Lebesgue density $p_Z(\cdot)$ such that $P(X = 1 \mid Z) \equiv \psi(Z)$ 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 X 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 ψ. As in (a), we want to make inference on ψ.

A special version of this model is the Case 1 interval censoring/current status model that is used extensively in biomedical studies and epidemiological contexts and has received much attention among biostatisticians and statisticians (see, for example, Groeneboom and Wellner (1992), Huang (1996), Sun and Kalbfleisch (1993), Shiboski (1998), Sun (1999), Banerjee and Wellner (2001)). Consider n individuals who are checked for infection at independent random times $Z_1, Z_2, \ldots, Z_n$; we set $X_i = 1$ if individual i is infected by time Z_i and 0 otherwise. We can think of X_i as $1\{U_i \leq Z_i\}$ where U_i is the (random) time to infection (measured from some baseline period). The U_i's are assumed to be independent and also independent of the Z_i's and are unknown. We are interested in making inference on F, the common (survival) distribution of the U_i's. We note that $\{X_i, Z_i\}_{i=1}^n$ is an i.i.d. sample from the distribution of (X, Z) where $Z \sim p_Z(\cdot)$ for some Lebesgue density p_Z and $X \mid Z \sim \text{Bernoulli}(F(Z))$. This is just a binary regression model with monotonicity constraints.

(iii) **Poisson Regression Model:** Suppose that $Z \sim p_Z(\cdot)$ and $X \mid Z = z \sim \text{Poisson}(\psi(z))$ where ψ is a monotone function. We have n i.i.d. observations from this model. Here one can think of Z as the distance 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 ψ 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)).

A common feature of all three models described above is the fact that the conditional distribution of the response in all three cases comes from a one parameter full rank exponential family (in (i), the variance σ^2 needs to be held fixed). One parameter full rank exponential family models have certain structural niceties that can be exploited to derive explicit solutions to the MLE of ψ. Our last example below considers a curved exponential family model for the response which is fundamentally of a different flavor.

(iv) **Conditional normality under a mean–variance relationship:** Consider the scenario, where Z has a Lebesgue density concentrated on an interval $[a, b]$ with $0 < a < b$ and given $Z = z$, $X \sim p(x, \psi(z))$ for an increasing function ψ, with $p(x, \theta)$ being the normal density with $\mu = c\,\theta^{-2m+1}$ and $\sigma^2 = d\,\theta^{-2m}$ for some real $m \geq 1$, and $\theta, c, d > 0$. For $m = 1$, this reduces to a normal density with a linear relationship between the mean and the standard deviation. Such a model could be postulated in a real–life setting based on say, exploratory plots of the mean–variance relationship using observed data, or background knowledge. The likelihood for the parametric model is analytically more complicated than the ones considered above and will be discussed later.

Based on existing work, one would expect $\hat{\psi}_n(z_0)$, the MLE of ψ at a pre–fixed point z_0 to satisfy (1.1), with z replaced by z_0. As will be seen, this indeed happens. This result permits the construction of (asymptotic) confidence intervals for $\psi(z_0)$ using the quantiles of $\mathbb{Z}$ which are well–tabulated. The constant $C(z_0)$ however needs to be estimated and involves nuisance parameters depending on the underlying model, and in particular, the derivative of ψ at z_0, estimating which is a tricky affair. 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 is suspect in this situation. This however is computationally intensive; see, for example, Banerjee and Wellner (2005) and Sen and Banerjee (2005) for a discussion of some of the issues involved.

Another likelihood based method of constructing confidence sets for $\psi(z_0)$ would involve testing a null hypothesis of the form $H_{0,\theta} : \psi(z_0) = \theta$, using the likelihood ratio test, for different values of θ, and then inverting the acceptance region of the likelihood ratio test; in other words, the confidence set for $\psi(z_0)$ is formed by compiling all values of θ for which the likelihood ratio statistic does not exceed a critical threshold. The threshold depends on $0 < \alpha < 1$, where $1 - \alpha$ is the level of confidence being sought, and the asymptotic distribution of the likelihood ratio statistic when the null hypothesis is correct. Thus, we are interested in studying the asymptotics of the likelihood ratio statistic for testing the true (null) hypothesis $H_{0,\theta_0} : \psi(z_0) = \theta_0$. It should be noted here that from the sole *perspective of statistical hypothesis testing*, a hypothesis of the above kind is not as important as, say, tests for monotonicity. However, pointwise null hypotheses are very important from the *perspective of estimation* since they serve as a conduit for setting confidence limits for the value of ψ, through inversion.

A key motivation behind studying the likelihood ratio based method in our framework comes from our experience with (regular) parametric models, where, under modest regularity conditions, the likelihood ratio statistic for testing the null hypothesis $\theta = \theta_0$, based on an i.i.d. sample $W_1, W_2, \ldots, W_n$ from the density $p(x, \theta)$ converges to a χ^2 distribution, under the null. This is a powerful result; since the limit distribution is free of nuisance parameters, this permits construction of confidence regions for the parameter with mere knowledge of the quantiles of the χ^2 distribution. Nuisance parameters like $I(\theta)$ do not need to be estimated from the data. This would however be indispensable if one tried to use the asymptotic normality of $\hat{\theta}$ to construct confidence intervals for the true value of the parameter. The question then naturally arises whether, similar to the classical parametric case, there exists a *universal limit distribution* that describes the limiting behavior of the likelihood ratio statistic when the null hypothesis H_{0,θ_0} holds, for the monotone response models introduced above. 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 χ^2 distribution for the limiting likelihood ratio statistic for monotone functions; the χ^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 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 (F) 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). 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 χ_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 χ_1^2 distribution. We will show that for our monotone response models, this is the case.

We are now in a position to describe the agenda for this paper. In Section 2, we stipulate regularity conditions on the monotone response models, under which the results in this paper are developed. We state and prove the main theorems describing the limit distributions of the MLE's and the likelihood ratio statistic. In particular, the emergence of a fixed limit law for the pointwise likelihood ratio statistic for testing ψ has some of the flavor of Wilks' classical result (Wilks (1938)) on the limiting χ^2 distribution of likelihood ratios in standard parametric models. Section 3 discusses applications of the main theorems and Section 4 contains some discussion and concluding remarks. Section 5 contains the proofs of the lemmas used to establish the main results in Section 2, and is followed by references.

2 Model Assumptions, Characterizations of Estimators and Main Results

Consider the general monotone response model introduced in the previous section. Let z_0 be an interior point of $\tilde{I}$ at which one seeks to estimate ψ. Assume that

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

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

The joint density of the data vector $\{X_i, Z_i\}_{i=1}^n$ (with respect to an appropriate dominating measure) can be written as:

$$p_n(\psi, \{X_i, Z_i\}_{i=1}^n) = \prod_{i=1}^n p(X_i, \psi(Z_i)) \times \prod_{i=1}^n p_Z(Z_i)\,.$$

The second factor on the right side of the above display does not involve ψ and hence is irrelevant as far as computation of MLE's is concerned. Absorbing this into the dominating measure, the likelihood function is given by the first factor on the right side of the display above. Denote by $\hat{\psi}_n$ the unconstrained MLE of ψ and by $\hat{\psi}_n^0$ the MLE of ψ under the constraint imposed by the pointwise null hypothesis $H_0 : \psi(z_0) = \theta_0$. We assume:

(A.0): With probability increasing to 1 as $n \to \infty$, the MLE's $\hat{\psi}_n$ and $\hat{\psi}_n^0$ exist.

Consider the likelihood ratio statistic for testing the hypothesis $H_0 : \psi(z_0) = \theta_0$, where θ_0 is an interior point of Θ. Denoting the likelihood ratio statistic by $2 \log \lambda_n$, we have

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

In what follows, assume that the null hypothesis H_0 holds.

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 θ and is denoted by $\mathcal{X}$.

(A.2) $l(x,\theta) = \log p(x,\theta)$ is at least three times differentiable with respect to θ and is strictly concave in θ for every fixed x in $\mathcal{X}$. The first, second and third partial derivatives of $l(x,\theta)$ with respect to θ will be denoted by $\dot{l}(x,\theta), \ddot{l}(x,\theta)$ and $l'''(x,\theta)$.

(A.3) If T is any statistic such that $E_\theta(\mid T \mid) < \infty$, 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 θ_0.

(A.5) There exists a neighborhood $\mathcal{N}$ of θ_0 such that for all x, $\sup_{\theta\in\mathcal{N}} \mid l'''(x,\theta) \mid \leq B(x)$ and $\sup_{\theta\in\mathcal{N}} E_\theta(B(X)) < \infty$.

(A.6) The functions:

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

are continuous in a neighborhood of (θ_0,θ_0). Also, the function $f_3(\theta_1,\theta_2) = E_{\theta_1}(\ddot{l}(X,\theta_2)^2)$ is uniformly bounded in a neighborhood of (θ_0,θ_0).

(A.7) Let $H(\theta, M)$ be defined as:

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

Then,

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

We are interested in 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 ϕ that are square integrable on every compact set and ϕ_n is said to converge to ϕ if

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

for every K. The space $\mathcal{L} \times \mathcal{L}$ denotes the cartesian product of two copies of $\mathcal{L}$ with the usual product topology. Also define $B_{loc}(\mathbb{R})$ 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 $\sup_{x\in[-K,K]} \mid h_n(x) - h(x) \mid \to 0$ for every $K > 0$.

For positive constants a and b define the process $X_{a,b}(h) := a\,W(h) + b\,h^2$, 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 the non-decreasing function $g_{a,b}$ is piecewise constant, 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 $g^0_{a,b}(h)$ 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, $g^0_{a,b}(h)$, like $g_{a,b}(h)$, is non-decreasing and piecewise constant, 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, $g^0_{a,b}(h)$ 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 length of the interval $D_{a,b}$ on which $g_{a,b}$ and $g^0_{a,b}$ differ is $O_p(1)$. For more detailed descriptions of the processes $g_{a,b}$ and $g^0_{a,b}$, see Groeneboom (1989), Banerjee (2000), Banerjee and Wellner (2001) and Wellner (2003). Thus, $g_{1,1} \equiv g$ and $g^0_{1,1} \equiv g^0$ are the unconstrained and constrained versions of the slope processes associated with the canonical process $X_{1,1}(h) \equiv X(h)$. Finally define,

$$\mathbb{D} := \int \left((g(z))^2 - (g^0(z))^2\right) \, dz \, .$$

The following theorem describes the limiting behavior of the unconstrained and constrained MLE's of ψ, appropriately normalized.

Theorem 2.1 *Let,*

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

Let

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

Under assumptions A.0 – A.7 and (a), (b), $(X_n(h), Y_n(h)) \to_d (g_{a,b}(h), g^0_{a,b}(h))$ *finite dimensionally and also in the space* $\mathcal{L} \times \mathcal{L}$.

Thus,

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

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

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

For a proof of this proposition, see for example, Banerjee (2000). Using the fact that $g(0) \equiv_d 2\,\mathbb{Z}$ (see, for example, Prakasa Rao (1969)), we get:

$$n^{1/3}\,(\hat{\psi}_n(z_0) - \psi(z_0)) \to_d a\,(b/a)^{1/3}\,g(0) \equiv_d (8\,a^2\,b)^{1/3}\,\mathbb{Z}\,. \tag{2.3}$$

This is precisely the phenomenon described in (1.1).

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

Theorem 2.2 *Under assumptions A.0 – A.7 and (a), (b),*

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

when H_0 is true.

Remark 1: 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; this is not really a stronger assumption than (A.0). Since our main results deal with convergences in distribution, we can, without loss of generality, restrict ourselves to sets with probability tending to 1. In this paper, we will focus on the case when ψ is increasing. The case where ψ is decreasing is incorporated into this framework by replacing Z by $-Z$ and considering the (increasing) function $\overline{\psi}(z) = \psi(-z)$.

Remark 2: We briefly discuss the implications of Theorems 2.1 and 2.2 for constructing confidence sets for $\psi(z_0)$. Theorem 2.1 leads to (2.3), based on which an asymptotic level $1-\alpha$ confidence interval for $\psi(z_0)$ is given by $[\hat{\psi}_n(z_0) - 2\,n^{-1/3}\,(\widehat{a^2\,b})^{1/3}\,p_{\alpha/2}, \hat{\psi}_n(z_0) + 2\,n^{-1/3}\,(\widehat{a^2\,b})^{1/3}\,p_{\alpha/2}]$, where $p_{\alpha/2}$ is the upper $\alpha/2$'th quantile of the (symmetric) distribution of $\mathbb{Z}$ (this is approximately 1 when $\alpha = .05$). Here $\widehat{a^2\,b}$ is an estimate of $a^2\,b$. To use Theorem 2.2, for constructing confidence sets, for each θ one constructs the likelihood ratio statistic for testing the null hypothesis $H_{0,\theta} : \psi(z_0) = \theta$. Denoting this by $2\,\log\,\lambda_n(\theta)$, the set $\{\theta : 2\,\log\,\lambda_n(\theta) \le q_\alpha\}$, where q_α is the upper α'th quantile of $\mathbb{D}$ (this is approximately 2.28 for $\alpha = .05$) gives an asymptotic level $1-\alpha$ confidence set for $\psi(z_0)$.

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

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

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

We now discuss the characterization of the maximum likelihood estimators of ψ. We first write down the log–likelihood function for the data. This is given by:

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

The goal is to maximize this expression over all increasing functions ψ. Let $Z_{(1)} < Z_{(2)} < \ldots < Z_{(n)}$ denote the ordered values of Z and $X_{(i)}$ denote the observed value of X corresponding to $Z_{(i)}$. Since ψ is increasing $\psi(Z_{(1)}) \leq \psi(Z_{(2)}) \leq \ldots \leq \psi(Z_{(n)})$. Finding $\hat{\psi}_n$ therefore reduces to minimizing $\tilde{\psi}(u_1,u_2,\ldots,u_n) = \sum_{i=1}^n \phi(X_{(i)},u_i)$ over all $u_1 \leq u_2 \leq \ldots \leq u_n$. Once, we obtain the (unique) minimizer $\hat{u} \equiv (\hat{u}_1,\hat{u}_2,\ldots,\hat{u}_n)$, the MLE $\hat{\psi}_n$ at the points $\{Z_{(i)}\}_{i=1}^n$ is given by: $\hat{\psi}_n(Z_{(i)}) = \hat{u}_i$ for $i = 1,2,\ldots,n$.

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 $u_1 \leq u_2 \leq \ldots \leq u_n$. Note that $\tilde{\psi}$ is finite and differentiable on 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 $g(u) = (g_1(u), g_2(u), \ldots, g_{n-1}(u))^T$ and $g_i(u) = u_i - u_{i+1}\,,\ i = 1,2,\ldots,n-1\,.$ Then, there exists an $n-1$ dimensional vector $\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_{n-1})^T$ 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,

$$\sum_{i=1}^{n-1} \lambda_i\,(\hat{u}_i - \hat{u}_{i+1}) = 0 \text{ and } \bigtriangledown\,\tilde{\psi}(\hat{u}) + G^T\,\lambda = 0\,,$$

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 $\bigtriangledown_i\,\tilde{\psi}(\hat{u}) + (\lambda_i - \lambda_{i-1}) = 0$ for $i = 1,2,\ldots,n-1$ and and $\bigtriangledown_n\,\tilde{\psi}(\hat{u}) - \lambda_{n-1} = 0\,.$ Solving recursively to obtain the λ_i's (for $i = 1,2,\ldots,n-1$), we get

$$\lambda_i \equiv \sum_{j=i+1}^{n} \bigtriangledown_j\,\tilde{\psi}(\hat{u}) = \sum_{i+1}^{n} \dot{\phi}(X_{(j)},\hat{u}_j) \geq 0 \quad \text{,for} \quad i = 1,2,\ldots,(n-1) \tag{2.4}$$

and

$$\sum_{j=1}^{n} \bigtriangledown_j\,\tilde{\psi}(\hat{u}) = \sum_{1}^{n} \dot{\phi}(X_{(j)},\hat{u}_j) = 0\,. \tag{2.5}$$

Now, let $B_1,B_2,\ldots,B_k$ 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: $\sum_{i=1}^{n-1} \lambda_i\,(\hat{u}_i - \hat{u}_{i+1}) = 0$ forces $\lambda_i = 0$ whenever

$\hat{u}_i < \hat{u}_{i+1}$. Noting that $\bigtriangledown_r\, \psi(\hat{u}) = \dot{\phi}(X_{(r)}, \hat{u}_r)$, this implies that on each B_j, $\sum_{r\in B_j} \dot{\phi}(X_{(r)}, w_j) = 0$. Thus w_j is the unique solution to the equation

$$\sum_{r\in B_j} \dot{\phi}(X_{(r)}, w) = 0\,. \tag{2.6}$$

Also, if S is a *head–subset* of the block B_j (i.e. S is the ordered subset of the first few indices of the ordered set B_j), then it follows that

$$\sum_{r\in S} \dot{\phi}(X_{(r)}, w_j) \leq 0\,. \tag{2.7}$$

Note that (2.4) and (2.5) jointly imply that $\sum_{j=1}^{i} \bigtriangledown_j \tilde{\psi}(\hat{u}) \leq 0$ for all i; this is equivalent to the inequality $\sum_{j=1}^{i} \dot{l}(X_{(j)}, \hat{\psi}_n(Z_{(j)})) \geq 0$ for all i. If $\hat{\psi}_n(Z_{(k)}) < \hat{\psi}_n(Z_{(k+1)})$ for some k, then $\sum_{j=1}^{k} \dot{l}(X_{(j)}, \hat{\psi}_n(Z_{(j)})) = 0$, so that $\sum_{j=k+1}^{i} \dot{l}(X_{(j)}, \hat{\psi}_n(Z_{(j)})) \geq 0$ for any $i \geq k+1$. It is also easily deduced that $\sum_{j=i}^{k} \dot{l}(X_{(j)}, \hat{\psi}_n(Z_{(j)})) \leq 0$ for any $i \leq k$.

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. The cusum diagram will itself be characterized in terms of the solution $\hat{u}$, giving us a *self-induced characterization*. Under sufficient structure on the underlying parametric model (for example, one parameter full rank exponential families), the self-induced characterization can be avoided and the MLE can characterized as the slope of the convex minorant of a cusum diagram that is explicitly computable from the data. This issue is discussed in greater detail in Section 3.

Before proceeding further, we introduce some notation. For points $\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_0 = y_0 = 0$ and $x_0 < x_1 < \ldots < x_n$, consider the left-continuous function $P(x)$ such that $P(x_i) = y_i$ and such that $P(x)$ is constant on (x_{i-1}, x_i). We will denote the vector of slopes (left–derivatives) of the GCM of $P(x)$ computed at the points $(x_1, x_2, \ldots, x_n)$ by $\mathrm{slogcm}\,\{(x_i, y_i)\}_{i=0}^{n}$.

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

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

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

Now,

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

Now, consider the problem of minimizing ξ over $\mathbb{R}^n$, subject to the constraints: $g(u) \leq 0$. If $u^\star$ is the global minimizer, then necessary and sufficient conditions are given by conditions (2.4) (for

$i = 1, 2, \ldots, n-1$) and (2.5), with $\tilde{\psi}$ replaced by ξ and $\hat{u}$ replaced by $u^\star$. Now, $\bigtriangledown \xi(\hat{u}) = \bigtriangledown \tilde{\psi}(\hat{u})$, so that $u^\star = \hat{u}$ does indeed satisfy the conditions (2.4) (for $i = 1, 2, \ldots, n-1$) and (2.5), with $\tilde{\psi}$ replaced by ξ. Also, $u^\star$ is the unique minimizer of ξ subject to the (convex) constraints $g(u) \leq 0$ by virtue of the fact that the Hessian of ξ is always positive definite. A formal proof could be devised in the following manner: Suppose there exists $u^{\star\star}$ different from $u^\star$ satisfying the constraints and minimizing ξ. Since ξ is convex,

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

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

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

This implies

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

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

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

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

To compute $\hat{u}$, it therefore suffices to try to minimize ξ. Choosing $\mathcal{K}$ to be a diagonal matrix with the i, i'th entry being $d_i \equiv \bigtriangledown_{ii}\, \tilde{\psi}(\hat{u}) = \ddot{\phi}(X_{(i)}, \hat{u}_i)$, we see $\xi(u)$ reduces to:

$$\xi(u) = \sum_{i=1}^{n} \left[u_i - \hat{u}_i + \bigtriangledown_i\, \tilde{\psi}(\hat{u})\, d_i^{-1}\right]^2 d_i = \sum_{i=1}^{n} \left[u_i - \left(\hat{u}_i - \dot{\phi}_i(X_{(i)}, u_i)\, d_i^{-1}\right)\right]^2 d_i\,.$$

Since $\hat{u}$ minimizes $\sum_{i=1}^n \left[u_i - \left(\hat{u}_i - \dot{\phi}_i(X_{(i)}, u_i)\, d_i^{-1}\right)\right]^2 d_i$ subject to the constraints that $u_1 \leq u_2 \leq \ldots \leq u_n$ and hence, is given by the isotonic regression of the function $g(i) = \hat{u}_i - \dot{\phi}_i(X_{(i)}, u_i)\, d_i^{-1}$ on the ordered set $\{1, 2, \ldots, n\}$ with weight function d_i. It is well known that the solution

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

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

$$(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n) \equiv \left[\text{slogcm} \left\{\sum_{j=1}^{i} \ddot{\phi}(X_{(j)}, \hat{u}_j), \sum_{j=1}^{i} \left(\hat{u}_j\, \ddot{\phi}(X_{(j)}, \hat{u}_j) - \dot{\phi}(X_{(j)}, \hat{u}_j)\right)\right\}_{i=0}^{n}\right]. \tag{2.8}$$

Recall that $\hat{\psi}_n(Z_{(i)}) = \hat{u}_i$; for a z that lies strictly between $Z_{(i)}$ and $Z_{(i+1)}$, we set $\hat{\psi}_n(z) = \hat{\psi}_n(Z_{(i)})$. The MLE $\hat{\psi}_n$ thus defined is a piecewise constant right–continuous function.

Characterizing $\hat{\psi}_n^0$: Let m be the number of Z_i's that are less than or equal to z_0. Finding $\hat{\psi}_n^0$ amounts to minimizing $\psi(u) = \sum_{i=1}^n \phi(X_{(i)}, u_i)$ over all $u_1 \leq u_2 \ldots \leq u_m \leq \theta_0 \leq u_{m+1} \leq \ldots \leq u_n$. This can be reduced to solving two separate optimization problems. These are: (1) Minimize $\sum_{i=1}^m \phi(X_{(i)}, u_i)$ over $u_1 \leq u_2 \leq \ldots \leq u_m \leq \theta_0$ and (2) Minimize $\sum_{i=m+1}^n \phi(X_{(i)}, u_i)$ over $\theta_0 \leq u_{m+1} \leq u_{m+2} \leq \ldots \leq u_n$.

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 $(\hat{u}_1^0, \hat{u}_2^0, \ldots, \hat{u}_m^0)$ can be obtained through the following recipe : Minimize $\sum_{i=1}^m \dot{\phi}(X_{(i)}, u_i)$ over $u_1 \leq u_2 \leq \ldots \leq u_m$ to get $(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_m)$. Then, $(\hat{u}_1^0, \hat{u}_2^0, \ldots, \hat{u}_m^0) = (\tilde{u}_1 \wedge \theta_0, \tilde{u}_2 \wedge \theta_0, \ldots, \tilde{u}_m \wedge \theta_0)$. The solution vector to (2), say $(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \ldots, \hat{u}_n^0)$ is similarly given by $(\hat{u}_{m+1}^0, \hat{u}_{m+2}^0, \ldots, \hat{u}_n^0) = (\tilde{u}_{m+1} \vee \theta_0, \tilde{u}_{m+2} \vee \theta_0, \ldots, \tilde{u}_n \vee \theta_0)$ where $(\tilde{u}_{m+1}, \tilde{u}_{m+2}, \ldots, \tilde{u}_n) = \text{argmin}_{u_{m+1} \leq u_{m+2} \leq \ldots \leq u_n} \sum_{i=m+1}^n \dot{\phi}(X_{(i)}, u_i)$.

An examination of the connections between the unconstrained and the constrained solutions reveals that the following holds:

$$\hat{\psi}_n(z) \neq \hat{\psi}_n^0(z) \Rightarrow \hat{\psi}_n^0(z) = \theta_0 \ \text{ or } \ \hat{\psi}_n(z) = \hat{\psi}_n(z_0)\,. \tag{2.9}$$

Furthermore, for any z in the set D_n on which the two solutions differ, we have: $\mid \hat{\psi}_n(z) - \psi(z_0) \mid \ \geq \ \mid \hat{\psi}_n^0(z) - \psi(z_0) \mid$. Another important property of the constrained solution is that on any block B of indices where it is constant and not equal to θ_0, the constant value, say w_B^0 is the unique solution to the equation:

$$\sum_{i \in B} \dot{\phi}(X_{(i)}, w) = 0\,. \tag{2.10}$$

The constrained solution also has a self-induced 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,

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

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

$$(\hat{u}_1^0, \hat{u}_2^0, \ldots, \hat{u}_m^0) \equiv \left[\text{slogcm} \left\{\sum_{j=1}^i \ddot{\phi}(X_{(j)}, \hat{u}_j^0), \sum_{j=1}^i \left(\hat{u}_j^0\, \ddot{\phi}(X_{(j)}, \hat{u}_j^0) - \dot{\phi}(X_{(j)}, \hat{u}_j^0)\right)\right\}_{i=0}^m\right] \wedge \theta_0\,, \tag{2.11}$$

and

$$(\hat{u}^0_{m+1}, \hat{u}^0_{m+2}, \ldots, \hat{u}^0_n) \equiv \left[\text{slogcm}\left\{\sum_{j=m+1}^{i} \ddot{\phi}(X_{(j)}, \hat{u}^0_j), \sum_{j=1}^{i} \left(\hat{u}^0_j\, \ddot{\phi}(X_{(j)}, \hat{u}^0_j) - \dot{\phi}(X_{(j)}, \hat{u}^0_j)\right)\right\}_{i=m}^{n}\right] \vee \theta_0 \,. \tag{2.12}$$

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

Remark 3: The characterization of the estimators above does not take into consideration boundary constraints on ψ. However, in certain models, the very nature of the problem imposes natural boundary constraints; for example, the parameter space Θ for the parametric model may be naturally non-negative (Example (d) discussed above), in which case the constraint $0 \leq u_1$ needs to be enforced. Similarly, there can be situations, where u_n is constrained to lie below some natural bound. In such cases, Fenchel conditions may be derived in the usual fashion by applying the Kuhn-Tucker theorem and self-induced characterizations may be derived similarly as above. However as the sample size n grows, with probability increasing to 1, the Fenchel conditions characterizing the estimator in a neighborhood of z_0, will remain unaffected by these additional boundary constraints, since $\psi(z_0)$ is assumed to lie in the interior of the parameter space, and the asymptotic distributional results will remain unaffected.

The MLE's $\hat{\psi}_n$ and $\hat{\psi}^0_n$ for ψ in a closed neighborhood of z_0. We state this formally below.

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

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

For the purposes of deducing the limit distribution of the MLE's and the likelihood ratio statistic the following lemma that guarantees local consistency at an appropriate rate, is crucial.

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

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

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. Let $\mathbb{P}_n$ denote the empirical measure based on the data, that assigns mass $1/n$ to each observation (X_i, Z_i). For a monotone function Λ taking values in Θ, define the following processes:

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

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

and

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

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

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

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 $[-K, K]$, for every $K > 0$.*

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,*

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

Lemma 2.7 *For any $M > 0$, if $\tau_{L,M,n}$ denotes the first jump point of $\hat{\psi}_n$ to the left of $z_0 - M\, n^{-1/3}$ and $\tau_{R,M,n}$ denotes the first jump point of $\hat{\psi}_n$ to the right of $z_0 + M\, n^{-1/3}$, then $\tilde{\tau}_{L,M,n} = n^{1/3}\, (\tau_{L,M,n} - z_0)$ and $\tilde{\tau}_{R,M,n} = n^{1/3}\, (\tau_{R,M,n} - z_0)$ are $O_p(1)$.*

Lemma 2.8 *Suppose that* $\{W_{n\epsilon}\}, \{W_n\}$ *and* $\{W_\epsilon\}$ *are three sets of random vectors such that*

(i) $\lim_{\epsilon\to 0} \ \limsup_{n\to\infty} \ P[W_{n\epsilon} \neq W_n] = 0$, *(ii)* $\lim_{\epsilon\to 0} \ P[W_\epsilon \neq W] = 0$ *and (iii) For every* $\epsilon > 0$, $W_{n\epsilon} \to_d W_\epsilon$ *as* $n \to \infty$.

Then $W_n \to_d W$ *, as* $n \to \infty$.

The above lemma is adapted from Prakasa Rao (1969).

Proof of Theorem 2.1: The proof presented here relies on continuous–mapping arguments for *slopes of greatest convex minorant* estimators.

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 (2.8). Consequently,

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

Recall that the normalized processes are defined in the following manner:

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

and

$$\tilde{B}_{n,\hat{\psi}_n}(h) = n^{2/3}\,\frac{(B_{n,\hat{\psi}_n}(z_0 + h\,n^{-1/3}) - B_{n,\hat{\psi}_n}(z_0)) - \psi(z_0)\,(G_{n,\hat{\psi}_n}(z_0 + h\,n^{-1/3}) - G_{n,\hat{\psi}_n}(z_0))}{I(\psi(z_0))\,p_Z(z_0)}\,.$$

Now, $X_n(h) = n^{1/3}\,(\hat{\psi}_n(z_0 + h\,n^{-1/3}) - \psi(z_0))$. Let $SGCM_n \equiv$ slope-gcm $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in \mathbb{R}\}$ denote the slope of the greatest convex minorant of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in \mathbb{R}\}$. This is a piecewise constant function and it is easy to see that $\{X_n(h_i) : 1 \leq i \leq k\} = \{SGCM_n(\tilde{G}_{n,\hat{\psi}_n}(h_i)) : 1 \leq i \leq k\}$. By Lemmas 2.3 and 2.4, the process $\tilde{B}_{n,\hat{\psi}_n}(h)$ converges in distribution to the process $X_{a,b}(h) = a\,W(h) + b\,h^2$ under the topology of uniform convergence on compacta; in other words, $\{\tilde{B}_{n,\hat{\psi}_n}(h) : h \in [-C, C]\}$ converges as a process in the space of bounded functions on $[-C, C]$ equipped with the topology of uniform convergence to $\{X_{a,b}(h) : h \in [-C, C]\}$. Furthermore, by Lemma 2.5, the process $\{\tilde{G}_{n,\hat{\psi}_n}(h) : h \in [-C, C]\}$ converges in probability (uniformly) to the function h on $[-C, C]$.

Denote by $X_{n,C}(w)$, the slope of the GCM of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [-C, C]\}$ evaluated at the point $\tilde{G}_{n,\hat{\psi}_n}(w)$, where $-C < w < C$. Similarly, denote by $g_{a,b,C}(w)$ the slope of the GCM of $\{X_{a,b}(h) : h \in [-C, C]\}$, evaluated at the point w. Now fix an $M > 0$ such that the points $h_1, h_2, \ldots, h_k$ are in the interior of the set $[-M, M]$. For every $\epsilon > 0$, find $M_\epsilon > 0$, such that with probability at least $1-\epsilon$, $-M_\epsilon < \tilde{\tau}_{L,M,n} \le -M < M \le \tilde{\tau}_{R,M,n} < M_\epsilon$. That this can be done is guaranteed by Lemma 2.7. So, consider a sample point for which the above happens. Noting that the points $\tilde{G}_{n,\hat{\psi}_n}(\tilde{\tau}_{L,M,n})$ and $\tilde{G}_{n,\hat{\psi}_n}(\tilde{\tau}_{R,M,n})$ are kink–points of the GCM of $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in \mathbb{R}\}$, it follows that the GCM of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [\tilde{\tau}_{L,M,n}, \tilde{\tau}_{R,M,n}]\}$ coincides with the GCM of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in \mathbb{R}\}$ restricted to the set $[\tilde{G}_{n,\hat{\psi}_n}(\tilde{\tau}_{L,M,n}), \tilde{G}_{n,\hat{\psi}_n}(\tilde{\tau}_{R,M,n})]$. Since, $-M_\epsilon < \tilde{\tau}_{L,M,n}$ and $M_\epsilon > \tilde{\tau}_{R,M,n}$, it is also the case that the GCM of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [-M_\epsilon, M_\epsilon]\}$ coincides on the set $[\tilde{G}_{n,\hat{\psi}_n}(\tilde{\tau}_{L,M,n}), \tilde{G}_{n,\hat{\psi}_n}(\tilde{\tau}_{R,M,n})]$ with the GCM of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [\tilde{\tau}_{L,M,n}, \tilde{\tau}_{R,M,n}]\}$. Consequently, the slope of the GCM of $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [-M_\epsilon, M_\epsilon]\}$ evaluated at $\tilde{G}_{n,\hat{\psi}_n}(h_i), i = 1, 2, \ldots, k$ must agree with the slope of the GCM of $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in \mathbb{R}\}$ evaluated at the points $\tilde{G}_{n,\hat{\psi}_n}(h_i), i = 1, 2, \ldots, k$. We conclude that with probability more than $1-\epsilon$, $\{X_{n,M_\epsilon}(h_i), i = 1, 2, \ldots, k\}$ eventually coincides with $\{X_n(h_i) : i = 1, 2, \ldots, k\}$.

Further, we can ensure that with probability at least $1-\epsilon$, $-M_\epsilon < \tilde{\tau}_{L,M} \le -M < M \le \tilde{\tau}_{R,M} < M_\epsilon$, where $\tau_{L,M}$ is the first kink–point of the GCM of the process $X_{a,b}(h)$ to the left of $-M$ and $\tau_{R,M}$ is the first kink–point of the GCM of $X_{a,b}(h)$ to the right of M. And, as in the previous paragraph, we can argue that with probability more than $1-\epsilon$, $\{g_{a,b,M_\epsilon}(h_i), i = 1, 2, \ldots, k\}$ coincides with $\{g_{a,b}(h_i), i = 1, 2, \ldots, k\}$.

Now, setting

$$\{X_{n,M_\epsilon}(h_i), i = 1, 2, \ldots, k\} \equiv W_{n,\epsilon}, \quad \{X_n(h_i) : i = 1, 2, \ldots, k\} \equiv W_n \; ,$$

$$W_\epsilon \equiv \{g_{a,b,M_\epsilon}(h_i), i = 1, 2, \ldots, k\} \;\text{ and }\; W = \{g_{a,b}(h_i), i = 1, 2, \ldots, k\}$$

we find that Conditions (i) and (ii) of Lemma 2.8 are satisfied. It only remains to check Condition (iii). Since, on $[-M_\epsilon, M_\epsilon]$, $\tilde{G}_{n,\hat{\psi}_n}(h)$ converges uniformly in probability to h and $\tilde{B}_{n,\hat{\psi}_n}(h)$ converges (in distribution) to $X_{a,b}(h)$ under the uniform topology, the GCM of the points $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [-M_\epsilon, M_\epsilon]\}$ converges in distribution to the GCM of the process $\{X_{a,b}(h) : h \in [-M_\epsilon, M_\epsilon]\}$ (under the uniform topology); consequently, slope-gcm $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in [-M_\epsilon, M_\epsilon]\}$ evaluated at the points $\{\tilde{G}_{n,\hat{\psi}_n}(h_i) : i = 1, 2, \ldots, k\}$ converges in distribution to slope-gcm $\{X_{a,b}(h) : h \in [-M_\epsilon, M_\epsilon]\}$, evaluated at the points $\{h_i : 1 \le i \le k\}$ (without loss, the set of M_ϵ's can be taken to be a subset of the positive integers and hence with probability 1, the GCM of $\{X_{a,b}(h) : h \in [-M_\epsilon, M_\epsilon]\}$ is differentiable at the h_i's for each $\epsilon > 0$). This establishes (iii) of Lemma 2.8. Hence, by Lemma 2.8, we conclude that $\{X_n(h_i) : i = 1, 2, \ldots, k\}$ converges to $\{g_{a,b}(h_i) : i = 1, 2, \ldots, k\}$ in distribution.

A similar argument applies to the constrained MLE with some modification. The crux of the argument, which involves applying Lemma 2.8, is similar. The modification comes from the fact that instead of dealing with the slope of the convex minorant of $\{\tilde{G}_{n,\hat{\psi}_n}(h), \tilde{B}_{n,\hat{\psi}_n}(h) : h \in \mathbb{R}\}$, we now need to deal with slope-gcm $\{\tilde{G}_{n,\hat{\psi}_n^0}(h), \tilde{B}_{n,\hat{\psi}_n^0}(h) : h \le 0\} \wedge 0$ (where the minimum is interpreted as being taken pointwise) on one hand and slope-gcm $\{\tilde{G}_{n,\hat{\psi}_n^0}(h), \tilde{B}_{n,\hat{\psi}_n^0}(h) : h > 0\} \vee 0$ on the other. The convergence of slope-gcm $\{\tilde{G}_{n,\hat{\psi}_n^0}(h), \tilde{B}_{n,\hat{\psi}_n^0}(h) : h \leq 0\}$ evaluated at points $\{\tilde{G}_{n,\hat{\psi}_n^0}(s_i) : i = 1, 2, \ldots, k\}$ (for a set of points $s_1, s_2, \ldots, s_k$, where the s_i's are all negative) to slope-gcm $\{X_{a,b}(h) : h \leq 0\}$, evaluated at the s_i's, is deduced similarly, whence $\{$slope-gcm $\{\tilde{G}_{n,\hat{\psi}_n^0}(h), \tilde{B}_{n,\hat{\psi}_n^0}(h) : h < 0\}$ *evaluated at the* $\tilde{G}_{n,\hat{\psi}_n}(s_i)$*'s* $\} \wedge 0$ (this is precisely $Y_n(s_i) \equiv n^{1/3}(\hat{\psi}_n^0(z_0 + s_i\, n^{-1/3}) - \psi(z_0)), i = 1, 2, \ldots, k$) converges to $\{$slope-gcm $\{X_{a,b}(h) : h \leq 0\}$ *evaluated at the* s_i*'s*$\} \wedge 0$. But this is precisely the vector $\{g_{a,b}^0(s_i) : i = 1, 2, \ldots, k\}$. Similarly, one argues the convergence for points s_i's which are larger than 0.

Finally, note that the above convergences happen jointly for the constrained and unconstrained estimators; so we can conclude that $(X_n(h_i), Y_n(h_i))_{i=1}^k$ converges finite–dimensionally to $(g_{a,b}(h_i), g_{a,b}^0(h_i))_{i=1}^k$. 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 strenghtening 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). □

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:

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

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

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

for points $\psi_{n,i}^\star$ (lying between $\hat{\psi}_n(Z_{(i)})$ and $\psi(z_0)$) and $\psi_{n,i}^{\star\star}$ (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

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

and

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

Consider the term $I_{n,2}$. Now, J_n can be written as the union of blocks of indices, say $B_1^0, B_2^0, \ldots, B_l^0$, 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, $\sum_{j \in B} \dot{\phi}(X_{(j)}, w_B^0) = 0$ from (2.10). Thus, for any block B where $w_B^0 \neq \theta_0$ we have

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

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

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

where $\hat{\psi}_n^{0,\star}(Z_{(i)})$ 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,

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

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

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

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:

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

whence

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

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

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

The term $n^{1/3}\,(\mathbb{P}_n - P)\,\xi_n(X, Z) \to_p 0$ by Lemma 2.9 stated and proved below. 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

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

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

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

where

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

and

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

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 $M > 0$,

$$\sup_{|h| \le M}\;|\,E_{\psi(z_0 + h\,n^{-1/3})}\,(\ddot{\phi}(X, \psi(z_0))) - E_{\psi(z_0)}(\ddot{\phi}(X, \psi(z_0)))\,| \to 0\,,$$

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, using Lemma 2.8 introduced above. Set

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

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

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

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

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

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 $W_{n\epsilon} \to_d W_\epsilon$ as $n \to \infty$, for every fixed ϵ. This is so because by Theorem 2.1 $(X_n(h), Y_n(h)) \to_d \left(g_{a,b}(h), g_{a,b}^0(h)\right)$ as a process in $\mathcal{L} \times \mathcal{L}$ and $(f, g) \mapsto \int_{[-c,c]} \left(f^2(h) - g^2(h)\right) dh$ is a continuous real–valued function defined from $\mathcal{L} \times \mathcal{L}$ to the reals. Thus all conditions of Lemma (2.8) are satisfied, leading to the conclusion that $W_n \to_d W$. 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,

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

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

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

where $\hat{\psi}_n^{\star}(Z)$ 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,

$$\mid R_{n,1} \mid \leq \tilde{C}\,(\mathbb{P}_n - P)\,[B(X)\,1(Z \in [z_0 - M\,n^{-1/3}, z_0 + M\,n^{-1/3}])]$$
$$+\tilde{C}\,P\,[B(X)\,1(Z \in [z_0 - M\,n^{-1/3}, z_0 + M\,n^{-1/3}])]\,,$$

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 (see, for example Proposition 2 or Theorem 3 of Van der Vaart and Wellner (1999)), whereas the second term goes to 0 by direct computation. □

Lemma 2.9 *With*

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

we have: $n^{1/3}\,(\mathbb{P}_n - P)\,\xi_n(X,Z) \to_p 0\,.$

Proof of Lemma 2.9: By virtue of Lemmas 2.2 and 2.6, it is easy to see that with arbitrarily high pre-assigned probability (say, probability greater than $1 - \epsilon$, where ϵ is pre-assigned and allowed to be arbitrarily small), the random function $\xi_n(X,Z)$ is eventually contained in the class of functions $\mathcal{F}\,\mathcal{G}$ (the class of functions formed by taking the product of a function in $\mathcal{F}$ with a function in $\mathcal{G}$), where $\mathcal{F} = \{(\ddot{\phi}(X, \psi(z_0))\,1(z \in [a_0, b_0])\,1(z \in [a, b)) : a_0 \leq a < z_0 < b \leq b_0\}$ and $\mathcal{G} = \{r_1^2(z) - r_2^2(z) : r_1, r_2 \in \mathcal{R}_M\}$, where $\mathcal{R}_M$ is the class of monotone increasing functions on the line bounded in absolute value by $M > 0$ (M depends on ϵ). We choose a_0, b_0 above in such a way that the function $(\ddot{\phi}(X, \psi(z_0))\,1(z \in [a_0, b_0])$ is square-integrable; this is doable by the assumption on f_3 in (A.6). To complete the proof, it suffices to show that the class $\mathcal{F}\mathcal{G}$ is Donsker. By Example 2.10.23 of Van der Vaart and Wellner (1996), this follows if we can show that there exist envelope functions F for $\mathcal{F}$ and G for $\mathcal{G}$ such that $P\,F^2\,G^2 < \infty$ and for which $\mathcal{F}$ and $\mathcal{G}$ satisfy the uniform–entropy condition (see, Section 2.5 of Van der Vaart and Wellner (1996), for a formal definition of this condition). Choosing F to be $\mid \ddot{\phi}(X, \psi(z_0)) \mid\; 1(Z \in [a_0, b_0])$ and G to be the constant function $2\,M^2$, we see that $P\,F^2\,G^2 < \infty$ holds by Assumption (A.6). That $\mathcal{F}$ satisfies the uniform entropy condition follows from the fact that the class of indicators of intervals of the form $[a, b)$ forms a VC–class of functions, and mulitplying this class by a fixed function, $\ddot{\phi}(X, \psi(z_0))\,1(Z \in [a_0, b_0])$, preserves the VC–property. A VC–class of functions readily satisfies the uniform entropy condition. To show that $\mathcal{G}$ satisfies the uniform entropy condition needs a little more work. If classes of functions $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy the uniform entropy condition for envelopes F_1 and F_2 it is not difficult to show that the class of functions $\mathcal{F}_1 - \mathcal{F}_2$ also satisfies the uniform entropy condition for the envelope $F_1 + F_2$. Now $\mathcal{G} = \mathcal{G}_1 - \mathcal{G}_2$, where $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{R}_M^2$. We will show that $\mathcal{G}_1$ satisfies the uniform entropy condition for the constant envelope function M^2, whence it will follow that $\mathcal{G}$ satisifes the uniform entropy condition for the constant envelope function $2\,M^2$.

Now, the constant function M is an envelope function for $\mathcal{R}_M$ and for every probability measure Q, we have:

$$\log\, N(\epsilon\, M, \mathcal{R}_M, L_2(Q)) \leq \log\, N_{[\,]}(2\,\epsilon\, M, \mathcal{R}_M, L_2(Q)) \leq \frac{K}{2\,M}\,\frac{1}{\epsilon}\,,$$

using Theorem 2.7.5 of Van der Vaart and Wellner (1996), where the constant K is completely free of Q. It follows by direct computation that

$$\int_0^\infty \sup_Q \sqrt{\log\, N(\epsilon\, M, \mathcal{R}_M, L_2(Q))}\, d\,\epsilon < \infty$$

showing that $\mathcal{R}_M$ satisifes the uniform entropy condition with envelope function M. Now, note that if f and g in $\mathcal{R}_M$ are at a distance less than $\epsilon\, M$ in the $L_2(Q)$ metric, then

$$\int (f^2 - g^2)\, dQ = \int (f+g)^2\,(f-g)^2\, dQ \leq 4\, M^2\, \epsilon^2\, M^2 = (2\, M^2\, \epsilon)^2\,.$$

It follows that $\log\, N(\epsilon\, 2\, M^2, \mathcal{R}_M^2, L_2(Q)) \leq K\,(2M\,\epsilon)^{-1}$ whence $\log\, N(\epsilon\, M^2, \mathcal{R}_M^2, L_2(Q)) \leq K\,(M\,\epsilon)^{-1}$, showing that $\mathcal{G}_1 \equiv \mathcal{R}_M^2$ indeed satisfies the uniform entropy condition for the constant envelope function M^2. $\Box$

3 Applications of the main results

In this section, we discuss some interesting special cases of monotone response models.

Consider the case of a one–parameter full rank exponential family model, naturally parametrized. Thus,

$$p(x, \theta) = \exp[\theta\, T(x) - C(\theta)]\, h(x)\,,$$

where θ varies in an open interval Θ. The function C possesses derivatives of all orders. Suppose we have $Z \sim p_Z(\cdot)$ and $X \mid Z = z \ \sim p(x, \psi(z))$ where ψ 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 ψ 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:

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

so that

$$\dot{l}(x, \theta) = T(x) - C'(\theta) \;\text{ and }\; \ddot{l}(x, \theta) = -C''(\theta)\,.$$

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 θ for all x; since $E_\theta(\dot{l}(X,\theta)) = 0$ we have $E_\theta(T(X)) = C'(\theta)$ and $I(\theta) = -E_\theta(\ddot{l}(X,\theta)) = C''(\theta)$. Also note that $I(\theta) = E_\theta(\dot{l}(X,\theta)^2) = \mathrm{Var}_\theta(T(X))$. Since $I(\theta) > 0$, we get $C''(\theta) > 0$ 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: $f_3(\theta_1,\theta_2) = C''(\theta_2)^2$ is clearly uniformly bounded in a neighborhood of(θ_0,θ_0); $f_2(\theta_1,\theta_2) = -C''(\theta_2)$ is differentiable in a neighborhood of (θ_0,θ_0) and that $f_1(\theta_1,\theta_2)$ is continuous follows on noting that $f_1(\theta_1,\theta_2) = I(\theta_1)+(C'(\theta_2)-C'(\theta_1))^2$.

To check conditions (A.5) and (A.7) fix a neighborhood of θ_0, say $(\theta_0-\epsilon,\theta_0+\epsilon)$. Since, for all x, $l'''(x,\theta) = -C'''(\theta)$ which is continuous, we can actually choose $B(x)$ in (A.5) to be a constant. It remains to verify condition (A.7). Since $C'(\theta)$ and $\ddot{l}(x,\theta)$ 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 γ and $\theta \in \mathcal{N}$,

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

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

$$\xi(\theta,M) \le \int 2(T(x)^2+\gamma)\,e^{\theta\,T(x)}\,e^{-C(\theta)}\,1\left(\mid T(x)\mid > M/2\right)h(x)\,d\,\mu(x)\,,$$

which in turn is dominated by

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

The expression above is not dependent on θ and hence serves as a bound for $\sup_{\theta\in\mathcal{N}} H(\theta,M)$. As M goes to ∞ 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 simpler characterization of the MLE's $\hat{\psi}_n$ and $\hat{\psi}_n^0$. For each block B of indices on which $\hat{\psi}_n(Z_{(i)})$ is constant with common value equal to, say, w, it follows by (2.6) that $\sum_{i\in B}(T(X_{(i)}-C'(w)) = 0$; hence

$$C'(w) = n_B^{-1}\sum_{i\in B} T(X_{(i)}) \tag{3.13}$$

where n_B is the number of indices in the block B. Furthermore, by (2.7), it follows that if S is a *head–subset* of the block B, then $\sum_{i\in B}(T(X_{(i)}-C'(w)) \ge 0$; i.e. $C'(w) \le n_S^{-1}\sum_{i\in S} T(X_{(i)})$, n_S denoting the cardinality of S. As a direct consequence of the above, we deduce that the unconstrained MLE $\hat{\psi}_n$ can actually be written as:

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

where

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

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, the MLE's have explicit characterizations for these models and their asymptotic distributions may also be obtained by direct methods. These would involve studying the asymptotic behavior of the unconstrained and constrained convex minorants of the cusum diagram given by $\{G_n(Z_{(i)}), V_n(Z_{(i)})\}_{i=0}^n$ in local $(n^{-1/3})$ neighborhoods of z_0 and subsequently transforming by the inverse of C'.

It is not difficult to check that Examples (i), (ii) and (iii) discussed in the introduction are special cases of the one-parameter full rank exponential family models discussed above. Theorems 2.1 and 2.2 therefore hold for these models, and MLE's have explicit characterizations and are easily computable. We discuss Example (ii) briefly and leave (i) and (iii) to the reader. Under Assumptions (a) and (b) of Section 2, Example (ii) falls in the monotone response framework: $X \mid Z = z \ \sim p(x, \psi(z))$ where $p(x, \theta) = \theta^x\,(1-\theta)^{1-x}$. Here $x \in \{1, 0\}$ and $0 < \theta < 1$. Set

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

Since ψ is monotone, so is $\tilde{\psi}$. Under this reparametrization, $X \mid Z = z \ \sim q(x, \tilde{\psi}(z))$ 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 $\eta_0 = \log(\theta_0/(1-\theta_0))$. It follows immediately from Theorem 2.2 that the likelihood ratio statistic in this model converges to $\mathbb{D}$. This is the key result that was derived in Banerjee and Wellner (2001), but now follows as a special case of our current results.

Consider now, Example (iv) with $c = d = 1$ for simplicity. In this case,

$$p(x, \theta) = \frac{1}{\sqrt{2\,\pi}}\,\theta^m \exp\left[-\frac{\theta^{2m}}{2}\left(x^2 - 2\,\frac{x}{\theta^{2m-1}} + \frac{1}{\theta^{4m-2}}\right)\right].$$

This is a one-parameter curved exponential family model with a two-dimensional sufficient statistic. It follows that

$$l(x, \theta) = \text{constant} + m\,\log(\theta) - \frac{\theta^{2m}\,x^2}{2} + \theta\,x - \frac{1}{2\,\theta^{2m-2}},$$

whence

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

For $m \geq 1$, this is clearly decreasing on $(0, \infty)$ showing that $l(x, \theta)$ is concave in θ. Also, note that as θ converges to 0 or diverges to ∞, $l(x, \theta)$ diverges to $-\infty$. Defining $l(x, \theta) = -\infty$ for $\theta = 0$, we can consider the problem of maximizing $\sum_{i=1}^n l(X_{(i)}, u_i)$ over all $0 \leq u_1 \leq u_2 \leq \ldots \leq u_n$. A unique maximum $\hat{u}$ exists, with $\hat{u}_1 > 0$. If B is a block of indices such that $\hat{u}_i$ assumes a constant value, say w, for $i \in B$ and S is a head-subset, then, as before, the Fenchel conditions imply that $\sum_{i \in B} \dot{l}(X_{(i)}, w) = 0$, which translates to:

$$n_B\,m\,\frac{1}{w} - m\,w^{2m-1}\sum_{i \in B} X_{(i)}^2 + \sum_{i \in B} X_{(i)} + (m-1)\,n_B\,\frac{1}{w^{2m-1}} = 0,$$

where n_B is the cardinality of B. Also, if n_S is the cardinality of S, then

$$n_S\, m\, \frac{1}{w} - m\, w^{2m-1} \sum_{i \in S} X_{(i)}^2 + \sum_{i \in S} X_{(i)} + (m-1)\, n_S\, \frac{1}{w^{2m-1}} \geq 0\,.$$

For a general (real) m, the above equations and inequalities do not translate into an explicit characterization of the MLE as the solution to an isotonic regression problem. The mileage that we get in the full rank exponential family models is no longer available to us owing to the more complex analytical structure of the current model. The self-induced characterization nevertheless allows us to write down the MLE as a slope of greatest convex minorant estimator, and determine its asymptotic behavior, following the route described in this paper (apart from log-concavity of $p(x, \theta)$ and the existence of a solution, which we argue, the other regularity conditions on the parametric model need to be verified; we leave this as an exercise for the reader).

Remark 4: Example (iv) attempts to illustrate the point that simple explicit characterizations of MLE's may not always be available for the monotone response models we have considered in this paper. While the analytical structure of a particular model may lead to an explicit representation for the MLE of ψ, as we have witnessed above, this typically will not provide any information about the characterization of the MLE for a structurally different model. For example, the explicit characterization of the MLE for the one parameter full rank exponential family models provides no insight into the characterization of the MLE in Example (iv). In our view, the advantage of the self-induced characterization lies in the fact that it does not depend on the analytical structure of the model and therefore provides an integrated representation of the MLE's as *slopes of greatest convex minorant estimators* across the entire class of monotone response models.

From a computational standpoint, an explicit characterization of the MLE, if available, clearly has its advantages. Applying the self-induced characterization involves iteration using, for example, the iterative convex minorant algorithm (Jongbloed (1998)), and can be time consuming. On the other hand, if the MLE also has an explicit characterization as the solution to an isotonic regression problem, then, for example, a simple (non-iterative) application of the PAVA will suffice, and will typically consume far less time.

4 Discussion

In this paper, we have studied 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 ψ. 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 the Introduction) 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. Suppose we have n i.i.d. observations from the distribution of (X, Z) where $Z = (Z_1, Z_2, \ldots, Z_k)$ is a random vector such that $Z_1 < Z_2 < \ldots < Z_k$ with probability 1 and given $Z = z$, X is distributed as $p(x, \psi(z_1), \psi(z_2) - \psi(z_1), \ldots, \psi(z_k) - \psi(z_{k-1}))$ where ψ is strictly increasing, and $p(x, \theta_1, \theta_2, \ldots, \theta_k)$ is a k–dimensional parametric model. Under appropriate regularity conditions (in particular, log-concavity in the parameters would appear to be essential), 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 $T = (T_1, T_2, \ldots, T_k)$ is a random vector of follow–up times. One only observes $(\Delta_1, \Delta_2, \ldots, \Delta_{k+1})$ 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 $Z = T$, $X = (\Delta_1, \Delta_2, \ldots, \Delta_{k+1})$ and $F = \psi$, this is a special case of the k parameter model with

$$p(x = (\delta_1, \delta_2, \ldots, \delta_{k+1}), \theta_1, \theta_2, \ldots, \theta_k) = \Pi_{i=1}^{k+1} \theta_i^{\delta_i}\,,$$

where $\theta_{k+1} = 1 - (\sum_{i=1}^{k} \theta_i)$. This is simply the multinomial distribution.

The above extension has a fundamental structural difference with the monotone response models treated in this paper. For the models of this paper, the values of the monotone function to be estimated appear in the log-likelihood function in a *separated* manner. Recall that the log-likelihood function can be written as $\sum_{i=1}^{n} l(X_{(i)}, u_i)$, where u_i is the value of the monotone function at $Z_{(i)}$, the i'th largest covariate value; furthermore, the additive components of the log-likelihood correspond to independent observations. These features simplify the characterization of the maximum likelihood estimates and the subsequent derivation of the asymptotics. However, the additive separated structure present here, is no longer encountered for the extended models of the previous paragraph. For simplicity, consider the Case 2 interval censoring model, where there are two observation times (U_i, V_i) for each individual and one records in which of the three mutually disjoint intervals $(0, U_i]$, $(U_i, V_i]$, (V_i, ∞), the individual fails. Letting $\{T_{(i)}\}_{i=1}^{k}$ denote the distinct ordered values of the $2n$ observation times $\{(U_i, V_i) : i = 1, 2, \ldots, n\}$ (here n is the number of individuals being observed), and u_l denote $F(T_{(l)})$, one can write down the log-likelihood for the data. It is seen that terms of the form $\log(u_i - u_j)$ immediately enter into the log-likelihood (see, for example, Groeneboom and Wellner (1992), for a detailed treatment). One important consequence of this, in particular, is the fact that the computation of the constrained MLE of the survival distribution F (under a hypothesis of the form $F(t_0) = \theta_0$) can no longer be decomposed into two separate optimization problems, in contrast to the monotone response models of this paper. Consequently, an analytical treatment of the constrained estimator will involve techniques beyond those presented in this paper. Regarding the unconstrained estimator of F in this model, Groeneboom (1996) uses some hard analysis to show that under a hypothesis of separation between U and V (the first and second observation times), the estimator converges to the truth at (pointwise) rate $n^{1/3}$, with limit distribution still given by $\mathbb{Z}$. A general treatment of the extended models is expected to involve applications and extensions of the ideas in Groeneboom (1996), and is left as a topic for future research.

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 (X, W, Z) where Z is unidimensional, but W can be vector–valued. Suppose that the distribution of X conditional on $(W, Z) = (w, z)$ is given by $p(x, \beta^T w + \psi(z))$ where $p(x, \theta)$ is a one-dimensional parametric model. We are interested in making inference on both β and ψ. The above formulation is fairly general and includes for example the partially linear regression model: $X = \beta^T W + \psi(Z) + \epsilon$ where ψ is a monotone function (certain aspects of this model have been studied by Huang(2002)), semiparametric logistic regression (with X denoting the binary response, and (W, Z) covariates of interest) where the log-odds of a positive outcome $(X = 1)$ is modelled as $\beta^T W + \psi(Z)$, and other models of interest . In the light of previous results, we expect that under appropriate conditions on $p(x, \theta)$, $\sqrt{n}\,(\hat{\beta}_{MLE} - \beta)$ 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 χ^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 ψ should exhibit an $n^{1/3}$ rate of convergence to a non–normal limit and the likelihood ratio for testing ψ 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 Appendix

Here, we present proofs of selected lemmas.

Proof of Lemma 2.1: Global consistency can be established in a variety of ways. Here is an approach adapted from Section 4.1 of Groeneboom and Wellner. Before proceeding further, we impose some further (mild) assumptions. These are:

$\tilde{A}(1)$ Let $L = \inf\{\psi(z) : z \in \tilde{I}\}$, and $U = \sup\{\psi(z) : z \in \tilde{I}\}$. Then $-\infty < L < U < \infty$. Also, $L, U \in \Theta$.

$\tilde{A}(2)$ The function f_1 from (A.6) is uniformly bounded in a compact rectangle containing $[L, U] \times [L, U]$.

$\tilde{A}(3)$ For $\theta, \theta_1, \theta_2 \in \Theta$, $E_\theta(\dot{l}(X, \theta_1)) \neq E_\theta(\dot{l}(X, \theta_2))$ whenever $\theta_1 \neq \theta_2$.

For this lemma, we will denote the true monotone function by ψ_0. To establish this lemma, note that by $\tilde{A}(1)$, the true function ψ_0 is bounded on $\tilde{I}$, the domain of Z (there are constants L and

U such that $L \leq \psi(z) \leq U$ on $\tilde{I}$). Also, we denote by $\tilde{\psi}_n$, the M.L.E. of ψ_0 among all monotone increasing functions that assume values between L and U. We first establish strong–consistency of $\tilde{\psi}_n$ for ψ_0. Now, $\tilde{\psi}_n$ maximizes $\mathbb{P}_n\, l(X, \psi(Z))$ among all monotone increasing ψ assuming values between L and U (recall that $l(x,\theta) = \log p(x,\theta)$ and is assumed to be jointly continuous in x and θ). For every $\epsilon > 0$, $(1-\epsilon)\,\tilde{\psi}_n + \epsilon\,\psi_0$ is a monotone increasing function assuming values between L and U. It follows that

$$\lim_{\epsilon \to 0+} \frac{\mathbb{P}_n\,[l(X, (1-\epsilon)\,\tilde{\psi}_n(Z) + \epsilon\,\psi_0(Z))] - \mathbb{P}_n[l(X, \tilde{\psi}_n(Z))]}{\epsilon} \leq 0\,.$$

But the above is simply,

$$\mathbb{P}_n \left[\lim_{\epsilon \to 0+} \frac{l(X, (1-\epsilon)\,\tilde{\psi}_n(Z) + \epsilon\,\psi_0(Z)) - l(X, \tilde{\psi}_n(Z))}{\epsilon} \right] = \mathbb{P}_n \left[\dot{l}(X, \tilde{\psi}_n(Z))\,(\psi_0(Z) - \tilde{\psi}_n(Z)) \right]\,.$$

We therefore obtain:

$$\mathbb{P}_n\,[\dot{l}(X, \tilde{\psi}_n(Z))\,(\psi_0(Z) - \tilde{\psi}_n(Z))] \leq 0\,. \tag{5.14}$$

Next, consider the class of functions

$$\mathcal{F} = \{\tilde{\xi}(x,z) \equiv \xi(z) : \xi \text{ is monotone;} ; L \leq \xi(z) \leq U,\ z \in \mathbb{R}\}\,.$$

The class $\mathcal{F}$ has finite $L_1(Q)$ bracketing entropy for every probability measure Q on $\mathbb{R}^2$, for every $\epsilon > 0$ and is therefore a P–Glivenko-Cantelli class of functions; here P refers to the joint distribution of (X, Z) on $\mathbb{R}^2$. Furthermore, the function $(x,z) \mapsto x$ is also P– Glivenko-Cantelli by the strong law. Since $(a,b,c,d) \mapsto \dot{l}(a,b)\,(c-d)$ is a continuous function, it follows by standard preservation properties for Glivenko–Cantelli classes of functions (see for example, Theorem 3 of Van der Vaart and Wellner (1999)) that the class of functions

$$\tilde{\mathcal{F}} \equiv \{\dot{l}(x, \psi_1(z))\,(\psi_2(z) - \psi_3(z)) : \psi_1, \psi_2, \psi_3 \in \mathcal{F}\}$$

is P–Glivenko-Cantelli, provided it has an integrable envelope. An integrable envelope is given, for example, by a positive constant times $\mid \dot{l}(x, L) \mid + \mid \dot{l}(x, U) \mid$ (that this is integrable is implied, for example, by the boundedness of the function $f_1(\theta_1, \theta_2)$ on the set $[L, U] \times [L, U]$, which we assume ($\tilde{A}(2)$). Thus, there exists a set Ω_0 with $P(\Omega_0) = 1$, such that for every $\omega \in \Omega_0$,

$$\lim_{n \to \infty} \sup_{\psi_1, \psi_2, \psi_3 \in \mathcal{F}} \mid (\mathbb{P}_{n,\omega} - P)\,(\dot{l}(X, \psi_1(Z))\,(\psi_2(Z) - \psi_3(Z)) \mid = 0\,. \tag{5.15}$$

Here $\mathbb{P}_{n,\omega}$ is the empirical measure of the observations $\{X_i(\omega), Z_i(\omega)\}_{i=1}^n$.

Next, fix an $\omega \in \Omega_0$. By the Helly selection theorem, $\{\tilde{\psi}_{n,\omega}(\cdot)\}$ has a subsequence $\{\tilde{\psi}_{n_k,\omega}(\cdot)\}$ that converges *vaguely* to a non–decreasing right–continuous function $\overline{\psi}_\omega$, such that $L \leq \overline{\psi}_\omega(\cdot) \leq U$ (by *vague* convergence here, we mean pointwise convergence at every continuity point of $\overline{\psi}_\omega$). Note that the set of discontinuity points of $\overline{\psi}_\omega$ are countable. This implies that

$$\lim_{k \to \infty} \int \dot{l}(x, \tilde{\psi}_{n_k,\omega}(z))\,(\psi_0(z) - \tilde{\psi}_{n_k,\omega}(z))\,d\,\mathbb{P}_{n_k,\omega}(x,z) = \int \dot{l}(x, \overline{\psi}_\omega(z))\,(\psi_0(z) - \overline{\psi}_\omega(z))\,d\,P(x,z)\,. \tag{5.16}$$

To show this, observe that the left side of the above dislay can be written as

$$(\mathbb{P}_{n_k,\omega} - P)\ \left[\dot{l}(x, \tilde{\psi}_{n_k,\omega}(z))\,(\psi_0(z) - \tilde{\psi}_{n_k,\omega}(z))\right]$$

$$+P\ \left[\dot{l}(x, \tilde{\psi}_{n_k,\omega}(z))\,(\psi_0(z) - \tilde{\psi}_{n_k,\omega}(z)) - \dot{l}(x, \overline{\psi}_\omega(z))\,(\psi_0(z) - \overline{\psi}_\omega(z))\right] .$$

The first term in the above display converges to 0, on noting that the functions $\{\tilde{\psi}_{n_k,\omega}(z)\}$ all lie in the class $\mathcal{F}$, in conjunction with (5.15). The second term can be written as:

$$\int E_{\psi_0(z)}(\dot{l}(x, \tilde{\psi}_{n_k,\omega}(z))\,(\psi_0(z) - \tilde{\psi}_{n_k,\omega}(z))\,p_Z(z)\,dz - \int E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z))\,(\psi_0(z) - \overline{\psi}_\omega(z))\,p_Z(z)\,dz\,.$$

Outside the set of discontinuity points of $\overline{\psi}_\omega$ (which has Lebesgue measure 0), $E_{\psi_0(z)}(\dot{l}(x, \tilde{\psi}_{n_k,\omega}(z))\,(\psi_0(z) - \tilde{\psi}_{n_k,\omega}(z))$ converges to $E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z))\,(\psi_0(z) - \overline{\psi}_\omega(z))$; furthermore, these functions are all uniformly bounded by some constant. Hence the dominated convergence theorem implies that the second term goes to 0. Hence (5.16) holds. In conjuction with (5.14) this implies that:

$$\int \dot{l}(x, \overline{\psi}_\omega(z))\,(\psi_0(z) - \overline{\psi}_\omega(z))\,d\,P(x,z) \equiv \int E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z))\,(\psi_0(z) - \overline{\psi}_\omega(z))\,p_Z(z)\,dz \leq 0\,.$$

This can be rewritten as:

$$\int \left[E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z)) - E_{\psi_0(z)}(\dot{l}(x, \psi_0(z))\right]\ (\overline{\psi}_\omega(z) - \psi_0(z))\,p_Z(z)\,dz \geq 0\,. \tag{5.17}$$

Now, suppose that $\overline{\psi}_\omega(z) > \psi_0(z)$. Then $\dot{l}(x, \overline{\psi}_\omega(z) \leq \dot{l}(x, \psi_0(z))$. Consequently, $E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z)) \leq E_{\psi_0(z)}(\dot{l}(x, \psi_0(z))$ and

$$\left[E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z)) - E_{\psi_0(z)}(\dot{l}(x, \psi_0(z))\right]\ (\overline{\psi}_\omega(z) - \psi_0(z)) \leq 0\,.$$

The same inequality continues to hold when $\overline{\psi}_\omega(z) < \psi_0(z)$ (and the inequality is an exact equality, when $\overline{\psi}_\omega(z) = \psi_0(z)$). This implies that:

$$\int \left[E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z)) - E_{\psi_0(z)}(\dot{l}(x, \psi_0(z))\right]\ (\overline{\psi}_\omega(z) - \psi_0(z))\,p_Z(z)\,dz \leq 0\,.$$

It follows that

$$\int \left[E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z)) - E_{\psi_0(z)}(\dot{l}(x, \psi_0(z))\right]\ (\overline{\psi}_\omega(z) - \psi_0(z))\,p_Z(z)\,dz = 0\,. \tag{5.18}$$

In the above integral z varies in the support of z. Now, fix $a < b$ in the interior of the support, such that (i) $a < z_0 < b$ and (ii) g is positive and continuous on $[a, b]$. The function:

$$\eta(z) = \left[E_{\psi_0(z)}(\dot{l}(x, \overline{\psi}_\omega(z)) - E_{\psi_0(z)}(\dot{l}(x, \psi_0(z))\right]\ (\overline{\psi}_\omega(z) - \psi_0(z))\,p_Z(z)$$

is right continuous on (a, b); if it is non–zero at any point $\tilde{z} \in (a, b)$ it has to be strictly negative at $\tilde{z}$, whence by right–continuity, it is negative on an interval of the form $[\tilde{z}, c)$, whence the integral in (5.18) is strictly negative. Hence we conclude that η is identically 0 on (a, b), whence $\overline{\psi}_\omega$ is identically equal to ψ_0 on (a, b) (here, we use ($\tilde{A}(3)$) which says that $E_\theta(\dot{l}(x, \theta_1)) \neq E_\theta(\dot{l}(x, \theta_2))$ if $\theta_1 \neq \theta_2$).

Thus, we have shown that for each ω in a set Ω_0 having probability 1, each subsequence of $\tilde{\psi}_{n,\omega}$ has a convergent subsequence and all these subsequences convergence pointwise to the same function ψ_0 on (a, b). It follows that for each $z \in (a, b)$, $\tilde{\psi}_n(z)$ converges to $\psi_0(z)$. Since ψ_0 is continuous, it follows that on any compact subset $[\sigma, \tau]$ of (a, b), $\sup_{\sigma \leq z \leq \tau} \mid \tilde{\psi}_{n,\omega}(z) - \psi_0(z) \mid$ converges to 0. This shows strong–consistency of $\tilde{\psi}_n$.

Since ψ_0' has a strictly positive derivative in a neighborhood of z_0, by choosing σ, τ to be sufficiently close to z_0 we can ensure that $L < \psi_0(z) < U$ for all $z \in [\sigma, \tau]$. Thus, for each $\omega \in \Omega_0$, $\tilde{\psi}_{n,\omega}(z)$ must eventually lie strictly between L and U for each $z \in [\sigma, \tau]$. Now, $\tilde{\psi}_{n,\omega}$ is obtained by first computing $\hat{\psi}_{n,\omega}$ and then setting $\tilde{\psi}_{n,\omega}(u) = \hat{\psi}_{n,\omega}(u)$ if $L < \hat{\psi}_{n,\omega}(u) < U$, $\tilde{\psi}_{n,\omega}(u) = L$ if $\hat{\psi}_{n,\omega}(u) < L$, and $\tilde{\psi}_{n,\omega}(u) = U$ if $\hat{\psi}_{n,\omega}(u) > U$. Consequently, $L < \tilde{\psi}_{n,\omega}(u) < U$ implies that $\hat{\psi}_{n,\omega}(u) = \tilde{\psi}_{n,\omega}(u)$. Thus, for every fixed $\omega \in \Omega_0$, $\hat{\psi}_{n,\omega}(u)$ converges to $\psi_0(u)$ for every $u \in [\sigma, \tau]$. It follows that on $[\sigma, \tau]$, $\sup_{\sigma \leq x \leq \tau} \mid \hat{\psi}_{n,\omega}(x) - \psi_0(x) \mid$ converges to 0.

The strong consistency of $\hat{\psi}_n^0$ for ψ_0 follows from the fact that for any $z \neq z_0$ that is sufficiently close to z_0, $\hat{\psi}_n(z)$ and $\hat{\psi}_n^0(z)$ are eventually equal with probability 1 (this is a consequence of the consistency of $\hat{\psi}_n$ and the relation between the constrained estimator and the unconstrained estimator). □

Proof of Lemma 2.2: We seek to establish that

$$\sup_{h \in [-M,M]} n^{1/3} \mid \hat{\psi}_n(z_0 + h\, n^{-1/3}) - \psi(z_0) \mid \text{ is } O_p(1)\,.$$

The above fact implies that

$$\sup_{h \in [-M,M]} n^{1/3} \mid \hat{\psi}_n^0(z_0 + h\, n^{-1/3}) - \psi(z_0) \mid \text{ is } O_p(1)\,,$$

by virtue of the fact that for any z, $\mid \hat{\psi}_n^0(z) - \psi(z_0) \mid \leq \mid \hat{\psi}_n(z) - \psi(z_0) \mid$.

We will need the following two lemmas

Lemma 5.1 *For any $\epsilon > 0$, there exists $C_0 > 0, R_0 > 0$ such that for any $C \geq C_0$ and $R \leq R_0$, we have*

$$P\left[sup_{z \in I_n} \int \dot{l}(x, \psi(z_0))\, 1(u \in [z, z_0))\, d\,\mathbb{P}_n(x, u) \geq 0\right] \leq \epsilon/2\,,$$

for n sufficiently large; here $I_n = (z_0 - R, z_0 - C\, n^{-1/3}]$.

Lemma 5.2 *For any $\epsilon > 0$, there exists $C_1 > 0, R_1 > 0$ such that for any $C \geq C_1$ and $R \leq R_1$, we have*

$$P\left[inf_{z\in I_n} \int \dot{l}(x, \psi(z_0 - 2\,C\,n^{-1/3}))\, 1(u \in [z_0 - 2\,C\,n^{-1/3}\, z))\, d\,\mathbb{P}_n(x,u) \leq 0\right] \leq \epsilon/2\,,$$

for n sufficiently large; here, $I_n = (z_0 - C\,n^{-1/3}, z_0 + R)$.

The proofs of these lemmas will be provided later. We will use the above lemmas along with the Fenchel conditions that characterize the M.L.E $\hat{\psi}_n$ to establish the desired local consistency. So, let $\epsilon > 0$ and choose $C > M$ and R satisfying the conditions of both Lemma 5.1 and Lemma 5.2. We seek to show that there is an $N_1 > 0$ such that for all $n \geq N_1$, we have:

$$P\left[\hat{\psi}_n(z_0 - C\,n^{-1/3}) \geq \psi(z_0)\right] \leq \epsilon\,.$$

Define $L_n = \{\hat{\psi}_n(z_0 - C\,n^{-1/3}) \geq \psi(z_0))$. Also, let U_n denote the event that $\hat{\psi}_n$ has a jump point in the interval $(z_0 - R, z_0 - C\,n^{-1/3}]$. To show that L_n eventually has probability no bigger than ϵ, we write it as the disjoint union of $L_n \cap U_n$ and $L_n \cap U_n^c$. Next, by using the strong consistency of $\hat{\psi}_n$ for ψ in a neighborhood of z_0, and the fact that $\psi(z_0 - R) < \psi(z_0 - C\,n^{-1/3})$ eventually , we conclude that eventually, with probability more than $1 - \epsilon/2$, $\hat{\psi}_n(z_0 - C\,n^{-1/3}) > \hat{\psi}_n(z_0 - R)$, whence $\hat{\psi}_n$ must have (at least one) a jump in $(z_0 - R, z_0 - C\,n^{-1/3}]$. Thus, for all $n \geq N_1$, U_n has probability at least as large as $1 - \epsilon/2$, whence U_n^c has probability no bigger than $\epsilon/2$. Thus $L_n \cap U_n^c$ has probability no bigger than $\epsilon/2$. Next, consider $L_n \cap U_n$ for $n \geq N_1$. Fix an ω in this set and let $\tau_n \equiv \tau_n(\omega)$ denote the last jump–point of $\hat{\psi}_n$ on the set $(z_0 - R, z_0 - C\,n^{-1/3}]$. From the Fenchel conditions that characterize the MLE, we have:

$$\int \dot{l}(x, \hat{\psi}_n(z))\, 1(\tau_n \leq z < z_0)\, d\,\mathbb{P}_n(x,z) \geq 0\,.$$

Notice that $\dot{l}(x, \cdot)$ is monotone decreasing in the second co-ordinate (when the first is held fixed); now $z \geq \tau_n \Rightarrow \hat{\psi}_n(z) \geq \hat{\psi}_n(\tau_n) = \hat{\psi}_n(z_0 - C\,n^{-1/3}) \geq \psi(z_0)$, the last inequality being a consequence of the fact that $\omega \in L_n$. Therefore it follows that for any fixed x, $\dot{l}(x, \hat{\psi}_n(z)) \leq \dot{l}(x, \psi(z_0))$, whenever $z \geq \tau_n$. Hence

$$0 \leq \int \dot{l}(x, \hat{\psi}_n(z))\, 1(\tau_n \leq z < z_0)\, d\,\mathbb{P}_n(x,z) \leq \int \dot{l}(x, \psi(z_0))\, 1(\tau_n \leq z < z_0)\, d\,\mathbb{P}_n(x,z)\,.$$

Since $\tau_n \in (z_0 - R, z_0 - C\,n^{-1/3}]$ this implies that

$$\sup_{z\in I_n} \int \dot{l}(x, \psi(z_0))\, 1(z \leq u < z_0)\, d\,\mathbb{P}_n(x,u) \geq 0\,.$$

But by Lemma 5.1, this happens with probability at most $\epsilon/2$ for n large enough. Hence $L_n \cap U_n$ is eventually contained in a set that has probability at most $\epsilon/2$. We conclude that L_n eventually has probability no larger than ϵ.

Next, we seek to show that the event $\tilde{L}_n = \{\hat{\psi}_n(z_0 - C\,n^{-1/3}) \le \psi(z_0 - 2\,C\,n^{-1/3})\}$ eventually has probability at most ϵ. Define $\tilde{U}_n$ to be the event that $\hat{\psi}_n$ has a jump in the interval $\tilde{I}_n = (z_0 - C\,n^{-1/3}, z_0 + R)$. Again, by the consistency of $\hat{\psi}_n$ in a neighborhood of z_0 we can ensure that eventually $\tilde{U}_n$ has probability at least $1 - \epsilon/2$. It followsthat eventually $\tilde{L}_n \cap \tilde{U}_n^c$ has probability at most $\epsilon/2$. Next, consider $\tilde{L}_n \cap \tilde{U}_n$. If $\tilde{\tau}_n$ is the first jump point of $\hat{\psi}_n$ on the interval $\tilde{I}_n$, then, using the Fenchel conditions once again, we get:

$$\int \dot{l}(x, \hat{\psi}_n(z))\,1(z_0 - 2\,C\,n^{-1/3} \le z < \tilde{\tau}_n)\,d\,\mathbb{P}_n(x,z) \le 0\,.$$

Now, for $z \in [z_0 - 2\,C\,n^{-1/3}, \tilde{\tau}_n)$, we get $\hat{\psi}_n(z) \le \hat{\psi}_n(\tilde{\tau}_n -) = \hat{\psi}_n(z_0 - C\,n^{-1/3}) \le \psi(z_0 - 2\,C\,n^{-1/3})$ since we are on the set $\tilde{L}_n$. Therefore $\dot{l}(x, \hat{\psi}_n(z)) \ge \dot{l}(x, \psi(z_0 - 2\,C\,n^{-1/3}))$. It follows that

$$\int \dot{l}(x, \psi(z_0 - 2\,C\,n^{-1/3}))\,1(z_0 - 2\,C\,n^{-1/3} \le z < \tilde{\tau}_n)\,d\,\mathbb{P}_n(x,z) \le 0\,.$$

Since $\tilde{\tau}_n \in \tilde{I}_n$, this implies that

$$\inf_{z \in \tilde{I}_n} \int \dot{l}(x, \psi(z_0 - 2\,C\,n^{-1/3}))\,1(z_0 - 2\,C\,n^{-1/3} \le u < z)\,d\,\mathbb{P}_n(x,u) \le 0\,.$$

But this happens eventually, with probability at most $\epsilon/2$, by Lemma 5.2. Thus, we conclude that $\tilde{L}_n \cap \tilde{U}_n$ eventually has probability at most $\epsilon/2$; consequently $\tilde{L}_n$ eventually has probability less than ϵ.

It follows that eventually $L_n^c \cap \tilde{L}_n^c$ has probability at least $1 - 2\,\epsilon$. But this is precisely the event $\{\psi(z_0 - 2\,C\,n^{-1/3}) < \hat{\psi}_n(z_0 - C\,n^{-1/3}) < \psi(z_0)\}$. Similarly, we can derive that the event $\{\psi(z_0) < \hat{\psi}_n(z_0 + C\,n^{-1/3}) < \psi(z_0 + 2\,C\,n^{-1/3})\}$ eventually has probability more than $1 - 2\,\epsilon$. Hence, eventually, with probability more than $1 - 4\,\epsilon$,

$$n^{1/3}(\psi(z_0 - 2\,C\,n^{-1/3}) - \psi(z_0)) \le n^{1/3}\,(\hat{\psi}_n(z_0 - C\,n^{-1/3}) - \psi(z_0)) \le 0$$

and

$$0 \le n^{1/3}\,(\hat{\psi}_n(z_0 + C\,n^{-1/3}) - \psi(z_0)) \le n^{1/3}(\psi(z_0 + 2\,C\,n^{-1/3}) - \psi(z_0))\,.$$

It follows easily from the above displays, the differentiability of ψ at z_0, and the fact that $C > M$, that eventually, with probability more than $1 - 4\,\epsilon$,

$$\sup_{|h| \le M} n^{1/3}\ \mid \hat{\psi}_n(z_0 + h\,n^{-1/3}) - \psi(z_0) \mid \equiv \sup_{h \in \{-M, M\}} n^{1/3}\ \mid \hat{\psi}_n(z_0 + h\,n^{-1/3}) - \psi(z_0) \mid$$

is eventually bounded by a positive constant. This completes the proof of the lemma. □

Proof of Lemma 5.1: To this end, define:

$$g_z(u) = \dot{l}(x, \psi(z_0))\,1(z \le u < z_0)\ \ , z \le z_0\,.$$

Set $\mathcal{G}_r = \{g_z : z \in (z_0 - r, z_0)\}$, for $r > 0$. We have

$$
\begin{aligned}
P\, g_z &= P\left[\dot{l}(x,\psi(z_0))\, 1(z \le u < z_0)\right] \\
&= \int_{[z,z_0)} \left[\int_{\mathcal{X}} \dot{l}(x,\psi(z_0))\, p(x,\psi(u))\, d\,\mu(x)\right] p_Z(u)\, du \\
&= \int_{[z,z_0)} \left[\int_{\mathcal{X}} \left(\dot{l}(x,\psi(u)) + \ddot{l}(x,\psi(u))\, (\psi(z_0)-\psi(u))\right) p(x,\psi(u))\, d\mu(x)\right] p_Z(u)\, du + r(z)\,,
\end{aligned}
$$

where

$$
r(z) = \frac{1}{2}\int_{[z,z_0)} (\psi(z_0)-\psi(u))^2 \left\{\int_{\mathcal{X}} l'''(x,\psi^*_{u,z_0})\, p(x,\psi(u))\, d\,\mu(x)\right\} p_Z(u)\, du\,.
$$

In the above display, ψ^*_{u,z_0} is a point between $\psi(z_0)$ and $\psi(u)$, and hence between $\psi(z)$ and $\psi(z_0)$. For all z sufficiently close to z_0 we can therefore ensure that $\mid l'''(x,\psi^*_{u,z_0}) \mid$ is bounded by $B(x)$ (for all $u \in [z, z_0)$), by (A.5). It follows that for z sufficiently close to z_0,

$$
\begin{aligned}
\mid r(z) \mid &\le \frac{1}{2}\int_{[z,z_0)} (\psi(z_0)-\psi(u))^2\, E_{\psi(u)}(B(X))\, p_Z(u)\, du \\
&\le \text{constant} \int_{[z,z_0)} (z_0-u)^2 \\
&= O(\mid z - z_0 \mid^3)\,,
\end{aligned}
$$

on using (A.5) again in conjunction with the boundedness of the derivative of ψ in a neighborhood of z_0. Recalling that

$$
\int_{\mathcal{X}} \dot{l}(x,\psi(u))\, p(x,\psi(u))\, du = 0 \;\text{ and }\; \int_{\mathcal{X}} \ddot{l}(x,\psi(u))\, p(x,\psi(u))\, du = -I(\psi(u))\,,
$$

we get

$$
\begin{aligned}
P\, g_z &= -\int_{[z,z_0)} (\psi(z_0)-\psi(u))\, D(u)\, du + O(\mid z - z_0 \mid^3) \\
&= -\int_{[z,z_0)} (\psi(z_0)-\psi(u))\, D(z_0)\, du - \int_{[z,z_0)} (\psi(z_0)-\psi(u))\, (D(u)-D(z_0))\, du + O(\mid z - z_0 \mid^3) \\
&= -\int_{[z,z_0)} (\psi(z_0)-\psi(u))\, D(z_0)\, du + O(\mid z - z_0 \mid^3)\,,
\end{aligned}
$$

for all z sufficiently close to z_0; in the display above $D(u) = I(\psi(u))\, p_Z(u)$. Now,

$$
\psi(u) = \psi(z_0) + \psi'(z_0)\,(u - z_0) + o(\mid u - z_0 \mid)\,.
$$

Using this, we get:

$$
\begin{aligned}
P\, g_z &= D(z_0)\, \psi'(z_0) \int_{[z,z_0)} (u - z_0)\, du - D(z_0) \int_{[z,z_0)} o(\mid u - z_0 \mid)\, du + O(\mid z - z_0 \mid^3) \\
&= -\frac{I(\psi(z_0))\, p_Z(z_0)\, \psi'(z_0)}{2}\, (z - z_0)^2 + o(\mid z - z_0 \mid^2)\,.
\end{aligned}
$$

Thus,

$$(z-z_0)^{-2}\,P\,g_z = -\frac{I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)}{2} + o(1)\,;$$

therefore by choosing z sufficiently close to z_0, say $z \in (z_0 - R_0, z_0)$ for R_0 sufficiently small, we can ensure that

$$P\,g_z \leq -\frac{I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)}{4}\,(z-z_0)^2\,.$$

Now, consider the class of functions $\mathcal{G}_r = \{g_z : z \in (z_0 - r, z_0)\}$, for $r \leq R_0$. For each r, we can construct the natural envelope function for $\mathcal{G}_r$ (the pointwise supremum over the class), given by: $G_r(x,u) \equiv \mid \dot{l}(x,\psi(z_0)) \mid 1(z_0 - r \leq u \leq z_0)$. Using the boundedness of $E_\theta\,(\mid \dot{l}(x,\psi(z_0) \mid)$ in a neighborhood of $\theta_0 \equiv \psi(z_0)$ (that depends only on R_0) and the boundedness of p_Z in a neighborhood of z_0 (that depends only on R_0), we conclude readily that $P\,G_r^2 \leq \textit{constant} \times r$ for a constant that depends only on R_0. Also, note that the classes of functions $\{\mathcal{G}_r : r \leq R_0\}$ are " uniformly manageable" in the sense of Kim and Pollard (1990), since each class is formed by taking the product of a fixed function with the indicators of a class of intervals (and therefore has finite VC–dimension, not depending on r). We can now invoke Theorem 4.1 of Kim and Pollard (1990) to conclude that for every $\delta > 0$, there exist a sequence of random variables $\{M_n\}$ that is $O_p(1)$ such that

$$\mid (\mathbb{P}_n - P)\,g_z \mid \leq (z_0 - z)^2 + n^{-2/3}\,M_n^2 \;\; \forall z \in (z_0 - R_0, z_0)\,.$$

Now,

$$\mathbb{P}_n\,g_z = (\mathbb{P}_n - P)\,g_z + P\,g_z \leq \mid (\mathbb{P}_n - P)\,g_z \mid + P\,g_z\,.$$

Then, for all $z \in (z_0 - R_0, z_0 - C\,n^{-1/3}]$, it follows from the result of Kim and Pollard (1990) above that,

$$\mathbb{P}_n\,g_z \leq \delta\,(z_0 - z)^2 + n^{-2/3}\,M_n^2 - \frac{I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)}{4}\,(z-z_0)^2 = -\frac{I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)}{8}\,(z-z_0)^2 + n^{-2/3}\,M_n^2$$

if we choose $\delta = I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)/8$. Now, for $z \in (z_0 - R_0, z_0 - C\,n^{-1/3}]$, we have $(z - z_0)^2 \geq C^2\,n^{-2/3}$, whence

$$\mathbb{P}_n\,g_z \leq n^{-2/3}\left[M_n^2 - \frac{I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)}{8}\,C^2\right]\,.$$

The right side is independent of z; now, given $\epsilon > 0$ we can ensure that for all sufficiently large n (say $n > N$), $M_n^2 < (I(\psi(z_0))\,p_Z(z_0)\,\psi'(z_0)/8)\,C_0^2$, with probability larger than $1 - \epsilon/2$, by choosing C_0 large enough (and the inequality continues to hold for any $C > C_0$, whenever $n > N$), whence it follows that for every $C > C_0$

$$P\left(\sup_{z \in (z_0 - R_0, z_0 - C\,n^{-1/3}]} \mathbb{P}_n\,g_z \leq 0\right) \geq 1 - \frac{\epsilon}{2}\,,$$

for all $n > N(C)$ (where N_C may need to be larger than N, so that $z_0 - R_0 < z_0 - C\,n^{-1/3}$). Hence, for all $n > N(C)$,

$$P\left(\sup_{z \in (z_0 - R_0, z_0 - C\,n^{-1/3}]} \mathbb{P}_n\,g_z > 0\right) \leq \frac{\epsilon}{2}\,;$$

since for $R < R_0$, $z_0 - R > z_0 - R_0$, it follows that

$$P\left(\sup_{z \in (z_0 - R, z_0 - C\, n^{-1/3}]} \mathbb{P}_n\, g_z > 0\right) \leq \frac{\epsilon}{2}\,,$$

for n sufficiently large. This establishes the lemma. □

The proof of Lemma 5.2 is similar and therefore skipped. We next prove Lemma 2.7.

Proof of Lemma 2.7: We only show that $\tau_{L,M,n}$, the first jump point of $\hat{\psi}_n$ to the left of $z_0 - M\, n^{-1/3}$ satisfies $n^{1/3}\,(\tau_{L,M,n} - z_0)$ is $O_p(1)$. The proof for $\tau_{R,M,n}$ is similar. We first need the following lemma, whose proof is outlined later.

Lemma 5.3 *Given $\epsilon > 0$ and $\tilde{C} > 0$, there exists $\tilde{D} > 0$ and $\tilde{R} > 0$ such that for all sufficiently large n,*

$$P\left[sup_{z \in I_n} \int \dot{l}(x, \psi(z_0 - 2\,\tilde{C}\, n^{-1/3}))\, 1(z \leq u < z_0 - 2\,\tilde{C}\, n^{-1/3})\, d\,\mathbb{P}_n(x,u) \geq 0\right] \leq \epsilon/2\,.$$

Here $I_n = (z_0 - \tilde{R}, z_0 - (2\,\tilde{C} + \tilde{D})\, n^{-1/3}]$.

Now, set $\tilde{C} = C$, where C is chosen as in the proof of Lemma 2.2; we will show that $L_n \equiv \{\hat{\psi}_n(z_0 - (2\,C + \tilde{D})n^{-1/3}) < \psi(z_0 - 2\,C\, n^{-1/3})\}$ eventually has probability larger than $1-\epsilon$. From the proof of Lemma 2.2 we know that the event $\{\hat{\psi}_n(z_0 - C\, n^{-1/3}) > \psi(z_0 - 2\,C\, n^{-1/3})\}$ has probability eventually larger than $1 - \epsilon$. Hence, the event

$$\{\hat{\psi}_n(z_0 - (2\,C + \tilde{D})\, n^{-1/3}) < \psi(z_0 - 2\,C\, n^{-1/3}) < \hat{\psi}_n(z_0 - C\, n^{-1/3})\}$$

eventually has probability larger than $1 - 2\,\epsilon$. Since $C > M$, it follows that eventually, with probability larger than $1 - 2\,\epsilon$, $\hat{\psi}_n(z_0 - M\, n^{-1/3}) \geq \hat{\psi}_n(z_0 - C\, n^{-1/3}) > \hat{\psi}_n(z_0 - (2\,C + \tilde{D})\, n^{-1/3})$; this implies that eventually, with probability larger than $1 - 2\,\epsilon$, $\hat{\psi}_n$ must have at least one jump in the interval $(z_0 - (2\,C + \tilde{D})\, n^{-1/3}, z_0 - M\, n^{-1/3})$, which implies that $\mid n^{1/3}(\tau_{L,M,n} - z_0) \leq 2\,C + \tilde{D} \mid$ with probability more than $1 - 2\,\epsilon$ for all sufficiently large n.

It remains to show that L_n has probability eventually larger than $1 - \epsilon$. We argue in a manner similar to the proof of Lemma 2.2. Let U_n denote the event that $\hat{\psi}_n$ has a jump point in $(z_0 - \tilde{R}, z_0 - (2\,C + \tilde{D})n^{-1/3}]$. Once again, by the consistency of $\hat{\psi}_n$ for ψ, it is easy to see that U_n eventually has probability larger than $1 - \epsilon/2$, whence $L_n \cap U_n^c$ eventually has probability no larger than $\epsilon/2$. To show that $L_n \cap U_n$ eventually has probability less than $\epsilon/2$, consider $\omega \in L_n \cap U_n$. Let $\hat{\tau}_n \equiv \hat{\tau}_{n,\omega}$ denote the last jump point of $\hat{\psi}_n$ on the set $(z_0 - \tilde{R}, z_0 - (2\,C + \tilde{D})\, n^{-1/3}]$. From the Fenchel conditions characterizing $\hat{\psi}_n$, we have:

$$\int \dot{l}(x, \hat{\psi}_n(z))\, 1\,[\hat{\tau}_n \leq z < z_0 - 2\,C\, n^{-1/3})\, d\,\mathbb{P}_n(x,z) \geq 0\,.$$

Now $z \geq \hat{\tau}_n$ implies that $\hat{\psi}_n(z) \geq \hat{\psi}_n(\hat{\tau}_n) = \hat{\psi}_n(z_0 - (2\,C + \tilde{D})\,n^{-1/3})$, so that $\dot{l}(x, \hat{\psi}_n(z)) \leq \dot{l}(x, \hat{\psi}_n(z_0 - 2\,C\,n^{-1/3}))$. This implies that

$$\int \dot{l}(x, \hat{\psi}_n(z_0 - 2\,C\,n^{-1/3}))\,1\,[\hat{\tau}_n \leq z < z_0 - 2\,C\,n^{-1/3})\,d\,\mathbb{P}_n(x,z) \geq 0\,.$$

Since $\hat{\tau}_n \in I_n$ with I_n as defined in Lemma 5.3, it follows that

$$\sup_{z \in I_n} \int \dot{l}(x, \hat{\psi}_n(z_0 - 2\,C\,n^{-1/3}))\,1\,[z \leq u < z_0 - 2\,C\,n^{-1/3})\,d\,\mathbb{P}_n(x,u) \geq 0\,;$$

but this event has probability eventually no bigger than $\epsilon/2$. This shows that $L_n \cap U_n$ eventually has probability less than $\epsilon/2$ and completes the proof that L_n eventually has probability no larger than ϵ. □

Proof of Lemma 5.3: The proof of this lemma is very similar to that of Lemma 5.1, and is outlined below. For each $r > 0$, define a class of functions $\mathcal{G}_{r,n}$ as follows: $\mathcal{G}_{r,n} = \{g_{n,z}(x,u) : z \in (z_0 - r, z_0 - 2\,C\,n^{-1/3})\}$. Here,

$$g_{n,z}(x,u) = \dot{l}(x, \psi(z_0 - 2\,C\,n^{-1/3}))\,1[z \leq u < z_0 - 2\,C\,n^{-1/3})]\,.$$

Setting $z_n \equiv z_0 - 2\,C\,n^{-1/3}$, and proceeding in a manner similar to the proof of Lemma 5.1, we obtain:

$$P\,g_{n,z} = -\frac{I(\psi(z_n))\,\psi'(z_n)\,p(z_n)}{2}(z - z_n)^2 + o((z - z_n)^2)\,.$$

Note than z_n converges to z_0 with increasing n; hence by choosing z to be sufficiently close to z_0, say $z \in (z_0 - \tilde{R}, z_n)$ for $\tilde{R}$ sufficiently small, we can ensure that for all sufficiently large n,

$$P\,g_{n,z} \leq -\frac{I(\psi(z_0))\,\psi'(z_0)\,p(z_0)}{8}(z - z_n)^2\,.$$

As in the proof of Lemma 5.1, we can construct envelopes G_r for the classes $\mathcal{G}_{r,n}$ (the envelopes do not depend on r), such that $P\,G_r^2 \leq \mathit{constant} \times r$, for a constant that only depends on $\tilde{R}$; also, the functions $\mathcal{G}_{r,n}$ are "uniformly manageable" for these envelopes in the sense of Kim and Pollard (1990), as in the proof of Lemma 5.1. Hence, the proof of Lemma 4.1 of Kim and Pollard (1990) goes through without any modification; for every $\delta > 0$, we can therefore produce an $O_p(1)$ sequence of random variables M_n such that

$$|\,(\mathbb{P}_n - P)\,g_{n,z}\,| \leq \delta\,(z - z_n)^2 + n^{-2/3}\,M_n^2\;\forall z \in (z_0 - \tilde{R}, z_n)\,.$$

It now follows that for all $z \in (z_0 - \tilde{R}, z_0 - (2\,C + \tilde{D})\,n^{-1/3}]$,

$$\mathbb{P}_n\,g_{n,z} \leq \delta\,(z_n - z)^2 + n^{-2/3}\,M_n^2 - \frac{I(\psi(z_0))\,\psi'(z_0)\,p(z_0)}{8}(z - z_n)^2\,.$$

Now, choose $\delta = I(\psi(z_0))\,\psi'(z_0)\,p(z_0)/16$ and note that for $z \in (z_0 - \tilde{R}, z_0 - (2\,C + \tilde{D})\,n^{-1/3}]$, $(z - z_n)^2 \geq \tilde{D}^2\,n^{-2/3}$. It then follows that

$$\mathbb{P}_n\,g_{n,z} \leq n^{-2/3}\left[M_n^2 - \frac{I(\psi(z_0))\,\psi'(z_0)\,p(z_0)}{16}\,\tilde{D}^2\right]\,,$$

for all $z \in (z_0 - \tilde{R}, z_0 - (2\,C + \tilde{D})\, n^{-1/3}]$, for n sufficiently large. The right side of the above display does not depend on z, and by choosing $\tilde{D}$ sufficiently large, we can ensure that the right side of the above display is eventually less than 0, with probability larger than $1 - \epsilon/2$, whence it follows that

$$P\,\left(\sup_{z \in (z_0 - \tilde{R}, z_0 - (2\,C + \tilde{D})\, n^{-1/3}]} \mathbb{P}_n\, g_{n,z} > 0\right) \leq \frac{\epsilon}{2}\,,$$

eventually. □

Proof of Lemma 2.3: It suffices to show that $\tilde{B}_{n,\Psi}(h)$ converges to the process $a\,W(h) + b\,h^2$ 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

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

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:

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

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:

$$\begin{aligned} P\, f_{n,s}\, f_{n,t} &= \frac{E\,\left[n^{1/3}\,\left(\dot{\phi}(X,\psi(Z)) - (\psi(Z) - \psi(z_0))\,\ddot{\phi}(X,\psi(Z))\right)^2\, 1(Z \in (z_0, z_0 + (s \wedge t)\, n^{-1/3}])\right]}{(I(\psi(z_0))\, p_Z(z_0))^2} \\ &= n^{1/3} \int_{z_0}^{z_0 + (s \wedge t)\, n^{-1/3}} H(z)\, p_Z(z)\, dz\,, \end{aligned}$$

where

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

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

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

As $\theta \to \theta_0 \equiv \psi(z_0)$, the first term converges to $I(\theta_0)$ by (A.4) and the second term converges to 0, since $f_3(\theta,\theta)$ is bounded in a neighborhood of θ_0 (by A.6). The third term also converges to 0, because, by the Cauchy–Schwarz inequality

$$\mid\; E_\theta\,(\dot{\phi}(X,\theta)\,\ddot{\phi}(X,\theta)) \mid \leq \sqrt{I(\theta)\,f_3(\theta,\theta)}\,,$$

which is bounded in a neighborhood of θ_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 ψ at z_0 the result follows.

We conclude that

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

on observing that $H(z_0) = E_{\psi(z_0)}\,(\dot{\phi}(X,\psi(z_0)))^2 = I(\psi(z_0))$. It is easily shown that $P\,f_{n,s}$ and $P\,f_{n,t}$ both converge to 0 as $n\to\infty$, 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 $K(s,t) = [I(\psi(z_0))\,p_Z(z_0)]^{-1}\,(\mid s\mid \wedge \mid t\mid)$, for $s,t<0$. But this is the covariance kernel of the Gaussian process $a\,W(h)$ 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 $a\,W(h)$. We next show that $\sqrt{n}\,P\,f_{n,h} \to (\psi'(z_0)/2)\,h^2$ 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 $X_{a,b}(h) \equiv a\,W(h) + b\,h^2$ 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,

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

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

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

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

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

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 $\sqrt{n}\,P\,f_{n,h} \to (\psi'(z_0)/2)\,h^2$ uniformly over $0 \le h \le K$.

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

$$\begin{aligned} f_{n,h} &\le C_0\,n^{1/6}\,\left[\left|\dot\phi(X,\psi(Z)) - \ddot\phi(X,\psi(Z))\,(\psi(Z)-\psi(z_0))\right|1\,\left(Z \in [z_0 - K\,n^{-1/3}, z_0 + K\,n^{-1/3}]\right)\right] \\ &\le \tilde C_0\,n^{1/6}\,\left[\left(\mid \dot\phi(X,\psi(Z)) \mid + \mid \ddot\phi(X,\psi(Z)) \mid\right)\,1\,\left(Z \in [z_0 - K\,n^{-1/3}, z_0 + K\,n^{-1/3}]\right)\right]\,, \end{aligned}$$

where $C_0 \equiv (I(\psi(z_0))\,p_Z(z_0))^{-1}$ and $\tilde C_0$ are constants. Thus,

$$F_n = \tilde C_0\,n^{1/6}\,\left[\left(\mid \dot\phi(X,\psi(Z)) \mid + \mid \ddot\phi(X,\psi(Z)) \mid\right)\,1\,\left(Z \in [z_0 - K\,n^{-1/3}, z_0 + K\,n^{-1/3}]\right)\right]$$

can be chosen as the envelope. Now,

$$F_n^2 \le 2\,\tilde C_0^2\,n^{1/3}\,\left(\dot\phi(X,\psi(Z))^2 + \ddot\phi(X,\psi(Z))^2\right)1\,\left(Z \in [z_0 - K\,n^{-1/3}, z_0 + K\,n^{-1/3}]\right)\,;$$

the upper bound on the right side of the above display is integrable by virtue of the continuity of ψ at z_0 and Assumptions (A.4) and (A.6).

Next to show that

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

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

$$\begin{aligned} &n^{1/3}\,E\,\left[\left(\dot\phi(X,\psi(Z))^2 + \ddot\phi(X,\psi(Z))^2\right)\,1\,\left(Z \in [z_0 - K\,n^{-1/3}, z_0 + K\,n^{-1/3}]\right)\right. \\ &\qquad \left.\times\left(1\,(\mid \dot\phi(X,\psi(Z)) \mid > \eta\,n^{1/3}/2) + 1\,(\mid \ddot\phi(X,\psi(Z)) \mid > \eta\,n^{1/3}/2)\right)\right] \to 0\,. \end{aligned}$$

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

$$\begin{aligned} &n^{1/3}\,E\,\left[\left(\dot\phi(X,\psi(Z))^2 + \ddot\phi(X,\psi(Z))^2\right)\,1\,(Z \in [z_0 - K\,n^{-1/3}, z_0 + K\,n^{-1/3}])\right. \\ &\qquad \left.\times\left(1\,(\mid \dot\phi(X,\psi(Z)) \mid > M) + 1\,(\mid \ddot\phi(X,\psi(Z)) \mid > M)\right)\right] \to 0\,. \end{aligned}$$

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 ψ, 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

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

if $\delta_n \to 0$. 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:

$$\begin{aligned} P\,(f_{n,s} - f_{n,t})^2 &= C_0^2\, n^{1/3}\, E\,\Big[(\dot{\phi}(X, \psi(Z)) - \ddot{\phi}(X, \psi(Z))\,(\psi(Z) - \psi(z_0)))^2 \\ &\qquad\qquad 1\,\{Z \in [z_0 + t\,n^{-1/3}, z_0 + s\,n^{-1/3}]\}\Big] \\ &\leq 2\,C_0^2\, n^{1/3} \int_{z_0 + t\,n^{-1/3}}^{z_0 + s\,n^{-1/3}} \left(I(\psi(z)) + (\psi(z) - \psi(z_0))^2\, f_3(\psi(z), \psi(z))\right)\, p_Z(z)\, dz\,. \end{aligned}$$

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

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

where κ' is an upper bound for $p_Z(z)$ in a pre-fixed neighborhood of z_0. Note that the last expression in the above display converges to 0 as $n \to \infty$.

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

$$\sup_Q \int_0^{\delta_n} \sqrt{\log\, N(\epsilon\, \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q))}\, d\,\epsilon \to 0 \;\; \forall \delta_n \to 0\,, \tag{5.19}$$

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

$$\mathcal{F}_{\delta_0} = \left\{\Big[(\psi(Z) - \psi(z_0))\,\ddot{\phi}(X, \psi(Z)) - \dot{\phi}(X, \psi(Z))\Big]\; 1\,(Z \in [z_0, z_0 + \delta])\; : \delta \leq \delta_0\right\}\,.$$

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, $\mathcal{F}_{n,1} \subset n^{1/6} \times \mathcal{F}_{\delta_0}$; it follows that each $\mathcal{F}_{n,1}$ is a VC class with VC–dimension bounded above by V_0 (≥ 2). It is well known (see, for example, Theorem 2.6.7 of Van der Vaart and Wellner (1996)) that

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

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

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

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 $N(\epsilon \, \|F_n\|_{Q,2}, \mathcal{F}_{n,2}, L_2(Q))$. It follows that for a sufficiently large integer $s' \geq 1$ we can ensure that

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

for some universal constant $K^\star$. To check (5.19) 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. $\Box$

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,

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

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

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

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

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

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

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

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

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:

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

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

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

which, with arbitrarily high probability is eventually bounded by a constant C (by Lemma 2.2). Also, since for any such z, $\hat{\psi}_n^{\star}(z)$ 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 $\mid\phi'''(X,\hat{\psi}_n^{\star}(Z))\mid$ is eventually bounded by $B(X)$ (by Assumption (A.5)). It follows that with arbitrarily high probability, the random function ξ_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,

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

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. □

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 on the right side of the above display 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 (2.9). This observation can be written in terms of the processes X_n and Y_n in the following way:

$$\{\tilde{B}_n > M\} \quad \subset \quad \{Y_n(M) = 0\} \cup \{X_n(0) = X_n(M)\}\,. \tag{5.20}$$

Now

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

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

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

so that

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

and the right hand side is again less than $\epsilon/4$ for M sufficiently large. It now follows from (5.20) that

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

eventually. □

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.

Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. *Scandinavian Journal of Stat.* **32**, 405–424.

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. (1996) Inverse Problems in Statistics *Proceedings of the St. Flour Summer School in Probability. Lecture Notes in Math.* **1648** 67-164. Springer, Berlin.

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. (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.

Sen, B. and Banerjee, M. (2005) A pseudo-likelihood method for analyzing interval–censored data. **Technical Report 422**, Department of Statistics, University of Michigan.

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.

Van der Vaart, A. and Wellner, J.A. (1999). Preservation theorems for Glivenko-Cantelli and Uniform Glivenko-Cantelli classes. *High Dimensional Probability II*, **2000**, 113 - 132. Birkhauser, Boston.

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

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

Wilks, S.S (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **19** , 60–62.