Gaussian White Noise Models: Some Results for Monotone Functions

BY JON A. WELLNER*

University of Washington

Gaussian white noise models have become increasingly popular as a canonical type of model in which to address certain statistical problems. We briefly review some statistical problems formulated in terms of Gaussian "white noise", and pursue a particular group of problems connected with the estimation of monotone functions. These new results are related to the recent development of likelihood ratio tests for monotone functions studied by [2]. We conclude with some open problems connected with multivariate interval censoring.

1. Introduction. This paper briefly reviews some of the recent research involving white noise models, and then develops some new results for statistical inference about monotone functions in the presence of white noise. The themes developed here differ substantially from the talk (on *Semiparametric Models with Sum Tangent Spaces*) which I presented at the Rochester meeting held in the Fall of 1999 in honor of Jack Hall's 70th birthday. The subject of that talk was more directly connected with my joint work with Jack in the late 70's and early 80's on semiparametric models. But one thing I learned from Jack Hall during my time at Rochester was not to become too fixed on any one problem or point of view, and that often a research problem can only be thoroughly understood by coming at it from several different perspectives or standpoints.

Jack Hall had an enormous impact on my development as a young statistician. Jack's continued interest in research and enthusiasm for good problems has been an inspiration.

In Section 2, we briefly review a slice of the past and current research work on "white noise models". In Section 3, we present some results on estimation of a monotone function observed "in white noise", and study a canonical version of the problem which arises repeatedly in the asymptotic distribution theory for nonparametric estimators of monotone functions. Section 3 carries through an analogous estimation problem in which some additional knowledge of the monotone function is available, namely its value at one point. This arises naturally when addressing the problem of finding a likelihood ratio test of the hypothesis $H : f(t_0) = \theta_0$ where f is monotone. The resulting likelihood ratio test statistic is introduced and studied in Section 5. Section 6 raises some further questions and problems. In particular we pose a problem concerning estimation of a monotone function of two variables subject to white noise on the plane (Brownian sheet) with a connection to multivariate interval censoring.

*Research supported in part by National Science Foundation grant DMS-95-32039 and NIAID grant 2R01 AI291968-04

2. Gaussian white noise models; some recent results. The following type of "white noise model" has been widely used as a unifying context and testing ground for nonparametric statistics: suppose that we observe $X(t)$ for $t \in K \subset \mathbb{R}^d$ where, symbolically,

$$dX(t) = f(t)dt + \sigma dW(t)\,; \tag{2.1}$$

here f is an "unknown function" in some class $\mathcal{F}$ of functions defined on the subset K of $\mathbb{R}^d$, W is standard Brownian motion (or Brownian sheet when $d > 1$), and $\sigma > 0$ is the standard deviation parameter controlling the relationship of the "noise", $\sigma dW(t)$, to the "signal", $f(t)$.

This type of model apparently goes back at least to [27]. A rigorous study of various problems got underway in the mid-1970's and early 1980's with the work of Kutoyants [28], Ibragimov and Khasminskii [21], and Ingster [24]. See [23] and [22], pages 199-213, and the discussion on page 393 for these and other references.

Pinsker [31] found the L_2–minimax constant for a Sobolev class of functions $\mathcal{F}$. Pinsker's result has been extended to other norms and problems by Korostelev [25], Donoho [9], and Korostelev and Nussbaum [26].

More recently such white noise models have been used as test problems for adaptive estimation (see e.g. [29] and [13]), adaptive testing (as in [35]), and model selection (see e.g. [4]).

A variety of inverse problems formulated in terms of the white noise model (2.1) have been studied: see e.g. [10] and [8]. Testing of qualititative hypotheses (such as monotonicity of f) has been considered in a white noise framework by Dümbgen and Spokoiny [12].

Various authors have emphasized the unifications possible by reducing complex problems to a white noise model of the form (2.1); see e.g. [11], [10], [5], [30], and [6].

From this brief review, it is clear that the literature on "white noise models" is vast and growing rapidly. We will not attempt to give a complete review here. Rather, we will develop some results concerning the estimation of a monotone function f in white noise. Here, as in many other statistical problems, there are two distinct roles for the Gaussian model:

- As a "continuous-time" model of interest in its own right.
- As a "canonical limiting-problem" appearing in connection with many other discrete-time models involving nonparametric estimation of a monotone function: e.g. [32], [7], and [15].

In the second version, the "canonical limiting problem", the unknown function f is replaced by a "canonical monotone function," namely $f_{can}(t) = 2t$. We will consider both versions of the problem in sections 3 and 4; a connection between the two will appear in subsection 3.3.

Estimation of a convex function f in Gaussian white noise is considered from the perspective of the "canonical limiting problem" in [17] where the "canonical convex function" is $f_{can} = 12t^2$; see [18] for a study of the asymptotic distribution theory of nonparametric estimators of a convex function.

3. Monotone function estimation in Gaussian white noise: general monotone f.

3.1. *General Monotone f on $[-c,c]$.* Consider the problem of estimating a monotone function f on the interval $[-c,c]$ in Gaussian white noise:

$$dX(t) = f(t)dt + \sigma dW(t) \qquad t \in [-c,c]\,. \tag{3.1}$$

Let P_f denote the law of the process X on $C[-c,c]$ when f is the mean (or intensity of drift) function; we denote the "true mean function" by f_0. Then by the Cameron-Martin-Girsanov theorem (see e.g. [33], page 81), the Radon-Nikodym derivative (likelihood ratio) dP_f/dP_0 is given by

$$\frac{dP_f}{dP_0} = \exp\left(\int_{-c}^{c} f(t)dX(t) - \tfrac{1}{2}\int_{-c}^{c} f^2(t)dt\right). \tag{3.2}$$

Thus the maximum likelihood estimator $\widehat{f}_c$ of f maximizes

$$\int_{-c}^{c} f(t)dX(t) - \tfrac{1}{2}\int_{-c}^{c} f^2(t)dt \tag{3.3}$$

over the class of monotone functions $f : [-c,c] \to \mathbb{R}$; equivalently, $\widehat{f}_c \equiv \widehat{f}$ minimizes

$$\phi(f) \equiv \tfrac{1}{2}\int_{-c}^{c} f^2(t)dt - \int_{-c}^{c} f(t)dX(t) \tag{3.4}$$

over the class of monotone functions f. Note that these are the first two terms of the "heuristic least squares problem" of minimizing

$$\tfrac{1}{2}\int_{-c}^{c}\left(f(t) - \dot{X}(t)\right)^2 dt = \tfrac{1}{2}\int_{-c}^{c}\left(f(t) - (f_0(t) + \sigma\dot{W}(t))\right)^2 dt \tag{3.5}$$

over the class of monotone functions. (As usual with Gaussian problems, maximum likelihood and least squares are equivalent.)

However, the problem of minimizing (3.4) over all monotone functions f on $[-c,c]$ is not well-defined, since this set of functions is not compact. A more sensible formulation of the problem is to look at the problem of minimizing (3.4), under the side restriction

$$\sup_{t\in[-c,c]} |f(t)| \le K\,, \tag{3.6}$$

ensuring that the minimization problem is well-defined for each c, since the set of functions that we allow is compact if we use (for example) the topology, induced by the supremum distance on the set of monotone functions on $[-c,c]$.

THEOREM 3.1. *Suppose that the monotone function $\widehat{f} : [-c,c] \to \mathbb{R}$ satisfies*

$$\|\widehat{f}\|_c \le K \tag{3.7}$$

where $\|\cdot\|_c$ denotes the supremum norm for functions on $[-c,c]$, and where $K > 0$ is a constant.

Let $\widehat{F}$ be an integral of $\widehat{f}$ (so that $\widehat{F}' = \widehat{f}$), and suppose that the two (Lagrange) parameters λ_1 and λ_2, given by

$$\lambda_1 = \int_{\{u:\widehat{f}(u)=-K\}} d\{\widehat{F}(u) - X(u)\} \tag{3.8}$$

and

$$\lambda_2 = -\int_{\{u:\widehat{f}(u)=K\}} d\{\widehat{F}(u) - X(u)\}, \tag{3.9}$$

are non-negative. (Alternatively, take λ_1 and λ_2 to be the solution of (3.10) and (3.12) below: then

$$\lambda_1 = \frac{1}{2}\left\{\frac{1}{K}\int_{-c}^{c} \widehat{f}(u)d(\widehat{F} - X)(u) + \int_{-c}^{c} d(\widehat{F} - X)(u)\right\}$$

and

$$\lambda_2 = \frac{1}{2}\left\{\frac{1}{K}\int_{-c}^{c} \widehat{f}(u)d(\widehat{F} - X)(u) - \int_{-c}^{c} d(\widehat{F} - X)(u)\right\},$$

if these are non-negative.) Then $\widehat{f}$ minimizes (3.4) over all monotone functions $f : [-c, c] \to \mathbb{R}$, such that $\|f\|_c \le K$, if the following conditions are satisfied:

$$-K(\lambda_1 + \lambda_2) - \int_{-c}^{c} \widehat{f}(u)\, d\{\widehat{F}(u) - X(u)\} = 0, \tag{3.10}$$

$$\lambda_2 + \int_{t}^{c} d\{\widehat{F}(u) - X(u)\} \ge 0, \ \textit{for all } t \in (-c, c], \tag{3.11}$$

and

$$\lambda_1 - \lambda_2 = \int_{-c}^{c} d\{\widehat{F}(u)) - X(u)\}. \tag{3.12}$$

Proof. For monotone functions $f : [-c, c] \to \mathbb{R}$, define $\phi(f)$ by (3.4), and let the function $\psi_{\lambda_1,\lambda_2}$ be defined by

$$\psi_{\lambda_1,\lambda_2}(f) = \phi(f) + \lambda_1\{-K - f(-c)\} + \lambda_2\{f(c) - K\}$$

where we define $f(-c)$ by $f(-c) = \lim_{u\downarrow -c} f(u)$. Then we have, for λ_1 and λ_2, defined by (3.8) and (3.9),

$$\psi_{\lambda_1,\lambda_2}(\widehat{f}) = \phi(\widehat{f}).$$

To see this, note that, by the definitions of λ_1 and λ_2, λ_1 can only be different from zero if $f(-c) = -K$, and likewise λ_2 can only be different from zero if $f(c) = K$. But (3.10) to (3.12) are exactly the Fenchel conditions for minimizing $\psi_{\lambda_1,\lambda_2}(f)$ over all monotone functions f. Hence we get, for all monotone functions f on $[-c, c]$ such that $|f| \le K$:

$$\phi(\widehat{f}) = \psi_{\lambda_1,\lambda_2}(\widehat{f}) \le \psi_{\lambda_1,\lambda_2}(f) \le \phi(f).$$

Hence $\widehat{f}$ minimizes $\phi(f)$ over all such functions f.

Now we show that (3.10) to (3.12) are in fact the Fenchel conditions. If we perturb the solution $\widehat{f}$ by a monotone function h we find that $\widehat{f}$ satisfies

$$\begin{aligned}(3.13)\qquad 0 &\le \frac{d}{d\epsilon}\psi_{\lambda_1,\lambda_2}(\widehat{f}+\epsilon h)|_{\epsilon=0}\\ &= \int_{-c}^{c} h(u)\widehat{f}(u)du - \int_{-c}^{c} h(u)dX(u) - \lambda_1 h(-c) + \lambda_2 h(c).\end{aligned}$$

If the functions $\widehat{f}+\epsilon h$ are monotone for $|\epsilon| \le \epsilon_0$ for some $\epsilon_0 > 0$, then (3.13) holds with equality. Now we get (3.10) by choosing $h = \widehat{f}$ (and noting that equality then holds in (3.13)); (3.12) follows by choosing $h = 1_{[-c,c]}$; and (3.11) follows by choosing $h = 1_{[t,c]}$, $t > -c$. □

Of course Theorem 3.1 holds both when the true drift function f_0 involved in the process X is the "canonical drift function" $f_{can}(t) \equiv 2t$, and also in the family of cases in which X is given by $X(t) \equiv X_{a,\sigma}(t) = \sigma W(t) + at^2$ for some $a > 0$. In these latter special cases we will extend the processes $\widehat{F}_{a,\sigma,c}$ characterized by Theorem 3.1 on the interval $[-c, c]$, to the whole line $\mathbb{R}$.

3.2. *Extension of the solution for f_0 from $[-c, c]$ to $\mathbb{R}$.* Let $X(t) \equiv X_{a,\sigma}(t) = \sigma W(t) + at^2$ where $W(t)$ is standard two-sided Brownian motion starting from 0. Suppose now that we have "observed" $X_{a,\sigma}$ on the whole line $\mathbb{R}$, and use $X_{a,\sigma}$ to estimate the true monotone function $f(t) = 2at$. Thus we are taking $f(t) = 2at \equiv af_{can}(t)$ for $t \in \mathbb{R}$, where $f_{can}(t) \equiv 2t$ is called the "canonical" monotone function. As we will see in the following subsection, the resulting slope process determines the limiting behavior of the estimator $\widehat{f}_\sigma$ derived in Section 3.1 as $\sigma \searrow 0$.

THEOREM 3.2. *(Canonical Solution Extended to $\mathbb{R}$.) For each $a > 0$, $\sigma > 0$, there exists an almost surely uniquely defined random continuous function $\widehat{F} \equiv \widehat{F}_{a,\sigma}$ satisfying the following conditions:*
(i) The function $\widehat{F}$ is everywhere below the function $X \equiv X_{a,\sigma}$:

$$(3.14)\qquad \widehat{F}(t) \le X(t), \ \textit{for each}\ t \in \mathbb{R}.$$

(ii) $\widehat{F}$ has a monotone derivative $\widehat{f}$.
(iii) The function $\widehat{F}$ satisfies

$$(3.15)\qquad \int_{\mathbb{R}} \{X(t) - \widehat{F}(t)\}\, d\widehat{f}(t) = 0.$$

In fact, $\widehat{F}$ is the *greatest convex minorant* of X, and in particular $\widehat{f}_{1,1}(0) = \widehat{F}'_{1,1}(0)$ is the random variable which describes the limiting distribution in a wide variety of monotone estimation problems; see [15], [16], and [20] where the distribution of $\mathbb{S}(0) \equiv \widehat{f}_{1,1}(0)$ is computed. Theorem 3.2 can be proved (but more easily) by the same methods used to prove Theorem 2.1 in [17]. The basic idea is that when $c \to \infty$ (and $K = K_c \to \infty$, the effects of the constraints at the endpoints $\pm c$

wash out, and the resulting characterizing equations come from (3.10) - (3.12) with $\lambda_1 = \lambda_2 = 0$ and $c = \infty$.

Note that condition (iii), in the presence of (i), means that the (increasing) function $\widehat{F}' = \widehat{f}$ cannot change (i.e. increase) in a region where (i) is satisfied with strict inequality; i.e. $\widehat{F}(t) = X(t)$ at the points t of increase of $\widehat{f}$.

Now we describe the scaling properties of the processes $\widehat{F}_{a,\sigma}$ and $\widehat{f}_{a,\sigma}$. We take $X_{1,1}$ to be the standard (or canonical) version of the family of processes $\{X_{a,\sigma} : a > 0, \sigma > 0\}$. Similarly, the canonial drift function is $f_{can}(t) = 2t$ (so that its integral in $F_{can}(t) = t^2$). Let $\widehat{F}_{a,\sigma}$ be the greatest convex minorant process corresponding to $X_{a,\sigma}$, let $\widehat{F}_{1,1}$ be the greatest convex minorant process corresponding to $X_{1,1}$, and let $\widehat{f}_{a,\sigma}$ and $\widehat{f}_{1,1} \equiv \mathbb{S}$ be the corresponding slope (left derivative) processes obtained by taking the left derivative of $\widehat{F}_{a,\sigma}$ and $\widehat{F}_{1,1}$ respectively.

PROPOSITION 3.1. (Scaling of the processes $X_{a,\sigma}$ and the envelope processes $\widehat{F}_{a,\sigma}$.)

$$X_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} \sigma(\sigma/a)^{1/3} X_{1,1}((a/\sigma)^{2/3} t) \tag{3.16}$$

as processes for $t \in \mathbb{R}$, *and hence also*

$$\widehat{F}_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} \sigma(\sigma/a)^{1/3} \widehat{F}_{1,1}((a/\sigma)^{2/3} t) \tag{3.17}$$

and

$$\widehat{f}_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} \sigma(a/\sigma)^{1/3} \widehat{f}_{1,1}((a/\sigma)^{2/3} t) = \sigma(a/\sigma)^{1/3} \mathbb{S}((a/\sigma)^{2/3} t) \tag{3.18}$$

as processes for $t \in \mathbb{R}$.

COROLLARY 3.1. *For the greatest convex minorant and slope processes* $\widehat{F}_{a,\sigma}$ *and* $\widehat{f}_{a,\sigma}$ *at* $t = 0$,

$$(\widehat{F}_{a,\sigma}(0), \widehat{f}_{a,\sigma}(0)) \stackrel{\mathcal{D}}{=} (\sigma(\sigma/a)^{1/3} \widehat{F}_{1,1}(0), \sigma(a/\sigma)^{1/3} \widehat{f}_{1,1}(0)) . \tag{3.19}$$

COROLLARY 3.2. (Finite interval scaling.)

$$\sigma^{-4/3} a^{1/3} X_{a,\sigma}((\sigma/a)^{2/3} t) \stackrel{\mathcal{D}}{=} X_{1,1}(t), \qquad t \in [-c, c] , \tag{3.20}$$

and hence observation of $\{X_{1,1}(t) : t \in [-c, c]\}$ *is equivalent to observation of* $\{X_{a,\sigma}(t) : t \in [-1, 1]\}$, *if* $c = (a/\sigma)^{2/3}$.

Remark: Note that this makes some intuitive sense; σ represents the "noise level" or standard deviation of the noise and the variance of our "estimators" $\widehat{f}_{a,\sigma}(0)$, should converge to zero as $\sigma \to 0$. Similarly, $a =$ twice the slope of the function $2at$ at zero; the function gets easier to estimate at this point as the slope goes to zero, and the proposition makes this precise. Note that the scaling in (3.19) is consistent with the finite-sample convergence results of [19] with the identification $\sigma = n^{-1/2}$.

Proofs. Starting with the proof of Proposition 3.1, we will find constants k_1, k_2 so that

$$k_1 X_{a,\sigma}(k_2 t) \stackrel{\mathcal{D}}{=} X_{1,1}(t) . \tag{3.21}$$

Since $\alpha^{-1/2}W(\alpha u) \stackrel{\mathcal{D}}{=} W(u)$ for each $\alpha > 0$,

$$X_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} at^2 + \sigma\alpha^{-1/2}W(\alpha t)\,. \tag{3.22}$$

Now by (3.22)

$$k_1 X_{a,\sigma}(k_2 t) \stackrel{\mathcal{D}}{=} k_1 a(k_2 t)^2 + k_1\sigma\alpha^{-1/2}W(\alpha k_2 t) \tag{3.23}$$

$$= t^2 + W(t) \tag{3.24}$$

if we choose k_1, k_2, α so that

$$ak_1k_2^2 = 1\,, \qquad \alpha k_2 = 1\,, \qquad \text{and} \qquad \sigma\alpha^{-1/2}k_1 = 1\,. \tag{3.25}$$

This yields $\alpha = 1/k_2$, and hence (from the last equality in the last display)

$$\sigma k_1 k_2^{1/2} = 1\,.$$

This in turn implies that

$$\frac{a}{\sigma}k_2^{3/2} = 1 \quad \text{or} \quad k_2 = (\sigma/a)^{2/3}\,.$$

This yields $k_1 = (1/\sigma)(a/\sigma)^{1/3}$. Expressing (3.21) as

$$X_{a,\sigma}(k_2 t) \stackrel{\mathcal{D}}{=} k_1^{-1}X_{1,1}(t/k_2)$$

with $k_1^{-1} = \sigma(\sigma/a)^{1/3}$ and $1/k_2 = (a/\sigma)^{2/3}$ yields the first claim of the proposition. The second claim follows from immediately from (3.16) and the definitions of $\widehat{F}_{a,\sigma}$ and $\widehat{F}_{1,1}$.

Corollary 3.1 follows from (3.19) and straightforward differentiation.

To prove Corollary 3.2, note that (3.16) is equivalent to

$$\sigma^{-1}(a/\sigma)^{1/3}X_{a,\sigma}((\sigma/a)^{2/3}t) \stackrel{\mathcal{D}}{=} X_{1,1}(t)\,.$$

Hence observation of $X_{1,1}$ on the interval $[-c, c]$ is equivalent to observation of $\sigma^{-4/3}a^{1/3}X_{a,\sigma}(t)$ for $t \in [-1, 1]$ if $c = (a/\sigma)^{2/3}$. $\square$

3.3. *Small σ Limits for the general monotone f problem.* Now suppose that we observe $X_\sigma(t) \equiv \sigma W(t) + F_0(t)$ for $t \in [-1, 1]$, and use the maximum likelihood estimator $\widehat{f}_\sigma$ of f_0 characterized by Theorem 3.1. Our goal here is to show that when $f_0'(t_0) > 0$ we have

$$r_\sigma(\widehat{f}_\sigma(t_0) - f_0(t_0)) \to_d \mathbb{Z} \qquad \text{as} \qquad \sigma \to 0$$

for some normalizing function r_σ and non-degenerate limiting variable $\mathbb{Z}$. In fact the right choice of r_σ is $r_\sigma = \sigma^{-2/3}$ and the limiting variable $\mathbb{Z}$ is determined by the slope process $g_{1,1} \equiv \mathbb{S}$ characterized by Theorem 3.2.

THEOREM 3.3. *Suppose that we observe $\{X_\sigma(t) : t \in [-1,1]\}$. Suppose that $t_0 \in (-1,1)$, $f_0'(t_0) > 0$, and f_0' is continuous at t_0. Then for any $K > 0$ the MLE $\hat{f}_\sigma$ satisfies:*

$$\sigma^{-2/3}(\widehat{f}_\sigma(t_0 + \sigma^{2/3}t) - f_0(t_0)) \to_d a^{1/3}\,\mathbb{S}(a^{2/3}t) \tag{3.26}$$

in the sense of convergence of all finite-dimensional distributions for $t \in [-K,K]$ where $a \equiv \frac{1}{2}f_0'(t_0)$. In particular,

$$\sigma^{-2/3}(\widehat{f}_\sigma(t_0) - f_0(t_0)) \to_d (\tfrac{1}{2}f_0'(t_0))^{1/3}\,\mathbb{S}(0)\,. \tag{3.27}$$

Theorem 3.3 is perhaps a bit more understandable if we reformulate the result in terms of the case of a sequence $\sigma \equiv \sigma_n \equiv 1/\sqrt{n}$. Then for observation of $X_n(t) \equiv F_0(t) + n^{-1/2}W(t)$ for $t \in [-1,1]$, Theorem 3.3 can be restated as follows:

THEOREM 3.4. *Suppose that we observe $\{X_n(t) : t \in [-1,1]\}$. Suppose that $t_0 \in (-1,1)$, $f_0'(t_0) > 0$, and f_0' is continuous at t_0. Then for any $K > 0$ the MLE $\widehat{f}_n \equiv \widehat{f}_{\sigma_n}$ satisfies:*

$$n^{1/3}(\widehat{f}_n(t_0 + n^{-1/3}t) - f_0(t_0)) \to_d a^{1/3}\,\mathbb{S}(a^{2/3}t) \tag{3.28}$$

in the sense of convergence of all finite-dimensional distributions for $t \in [-K,K]$ where $a = \frac{1}{2}f_0'(t_0)$. In particular,

$$n^{1/3}(\hat{f}_n(t_0) - f_0(t_0)) \to_d (\tfrac{1}{2}f_0'(t_0))^{1/3}\,\mathbb{S}(0)\,. \tag{3.29}$$

Proof. We will sketch the proof of Theorem 3.4; the proof of Theorem 3.3 is completely analogous. The first step basically consists of reduction to the case $t_0 = 0$ and $f_0(t_0) = 0$. Consider the new processes

$$\begin{aligned}\tilde{X}_n(t) &\equiv X_n(t_0+t) - X_n(t_0) - tf_0(t_0)\\ &= n^{-1/2}(W(t_0+t) - W(t_0)) + F_0(t_0+t) - F_0(t_0) - tf_0(t_0)\end{aligned}$$

for $t \in [-1-t_0, 1-t_0]$ so that

$$d\tilde{X}_n(t) =_d n^{-1/2}dW(t) + f_0(t_0+t) - f_0(t_0) \equiv n^{-1/2}dW(t) + \tilde{f}_0(t)$$

where $\tilde{f}_0(0) = 0$.

Now define a local process $X_n^{loc}(t)$, $t \in [n^{1/3}(-1-t_0), n^{1/3}(1-t_0)] \equiv [\alpha_n, \beta_n]$, by

$$\begin{aligned}X_n^{loc}(t) &= n^{2/3}\tilde{X}_n(n^{-1/3}t)\\ &= n^{2/3}(X_n(t_0+n^{-1/3}t) - X_n(t_0) + F_0(t_0+n^{-1/3}t) - F_0(t_0) - n^{-1/3}tf_0(t_0))\\ &= n^{2/3}\left(\frac{1}{\sqrt{n}}(W(t_0+n^{-1/3}t) - W(t_0)) + n^{-2/3}t^2\tfrac{1}{2}f_0'(\tilde{t}_n)\right)\\ &\stackrel{\mathcal{D}}{=} W(t) + \tfrac{1}{2}f_0'(\tilde{t}_n)t^2 \qquad \text{by Brownian scaling}\\ &\Rightarrow W(t) + \tfrac{1}{2}f_0'(t_0)t^2 \qquad \text{in} \quad l^\infty[-K,K]\\ &\equiv W(t) + at^2\end{aligned}$$

where $|\tilde{t}_n - t_0| \le n^{-1/3}|t|$ and $a \equiv \frac{1}{2} f_0'(t_0)$.

Now the greatest convex minorant $\widehat{F}_n$ of X_n on $[-1,1]$ corresponds to the greatest convex minorant $\tilde{F}_n^{loc}$ of $\tilde{X}_n^{loc}$ on $[\alpha_n, \beta_n]$ and the relationship between $\tilde{F}_n^{loc}$ and $\widehat{F}_n$ is simply

$$\begin{aligned} \tilde{F}_n^{loc}(t) &= n^{2/3}(\widehat{F}_n(t_0 + n^{-1/3}t) - \widehat{F}_n(t_0)) - n^{1/3} t f_0(t_0) \\ &\Rightarrow \widehat{F}_{a,1}(t) \stackrel{\mathcal{D}}{=} a^{-1/3} \widehat{F}_{1,1}(a^{2/3}t) \end{aligned}$$

by Proposition 3.1. The corresponding slope process is

$$\begin{aligned} \tilde{f}_n^{loc}(t) &= n^{1/3}(\widehat{f}_n(t_0 + n^{-1/3}t) - f_0(t_0)) \\ &\to_d \widehat{f}_{a,1}(t) \stackrel{\mathcal{D}}{=} a^{1/3} \widehat{f}_{1,1}(a^{2/3}t) = a^{1/3} \mathbb{S}(a^{2/3}t) \end{aligned}$$

where the last convergence in law is in the sense of all finite-dimensional distributions for the process indexed by $t \in [-K, K]$. □

4. Monotone function estimation in Gaussian white noise: constrained estimation. Now we want to consider the problem of estimating f in the model (3.1), with the additional knowledge that $f(t_0) = \theta_0$, a fixed number. This optimization arises naturally in connection with likelihood ratio tests of the hypothesis $f(t_0) = \theta_0$. Without loss of generality we may suppose that $t_0 = 0$. Furthermore, note that the problem of minimizing (3.5) over the class of monotone functions g with $g(0) = \theta_0$ (together with restrictions at the endpoints $\pm c$ to make the problem well-defined) separates naturally into the two problems:
(R) minimize

$$\phi_R(f) \equiv \tfrac{1}{2} \int_0^c f^2(t)dt - \int_0^c f(t)dX(t) \tag{4.1}$$

subject to $f(0) = \theta_0$ and f monotone; and
(L) minimize

$$\phi_L(f) \equiv \tfrac{1}{2} \int_{-c}^0 f^2(t)dt - \int_{-c}^0 f(t)dX(t) \tag{4.2}$$

subject to $f(0) = \theta_0$ and f monotone. These two problems are really identical, so it suffices to deal with the problem to the right of zero, problem (R).

4.1. *General Monotone f on $[-c, c]$ with $f(0) = 0$.* Now we consider the constrained problem (with constraint at 0 and at $\pm c$). To this end, we first reformulate the problem as an isotonic regression problem. We focus on the problem to the right of 0; the corresponding problem to the left of zero is analogous.

THEOREM 4.1. *Suppose that the monotone function $\widehat{f}_0 : [0, c] \to \mathbb{R}$ satisfies*

$$||\widehat{f}_0||_c \le K \tag{4.3}$$

where $\|\cdot\|_c$ *denotes the supremum norm for functions on* $[0,c]$, *and where* $K>0$. *Suppose that the two (Lagrange) parameters* λ_1 *and* λ_2, *given by*

$$\lambda_1=\int_{\{u:\hat f_0(u)=\theta_0\}} d\{\widehat F_0(u)-X(u)\}, \tag{4.4}$$

and

$$\lambda_2=-\int_{\{u:\hat f_0(u)=K\}} d\{\widehat F_0(u)-X(u)\} \tag{4.5}$$

are non-negative. (Alternatively, take λ_1 *and* λ_2 *to be the solution of (4.6) and (4.8) below: then*

$$\lambda_1=\frac{\left\{\int_0^c \hat f_0(u)d(\widehat F_0-X)(u)-K\int_0^c d(\widehat F_0-X)(u)\right\}}{\theta_0-K}$$

and

$$\lambda_2=\frac{\left\{-\int_0^c \hat f_0(u)d(\widehat F_0-X)(u)+\theta_0\int_0^c d(\widehat F_0-X)(u)\right\}}{K-\theta_0},$$

if these are non-negative.) Then $\hat f_0$ *minimizes (4.1) over monotone functions* $f:[0,c]\to\mathbb{R}$, *such that* $\|f\|_c\le K$ *and* $f(0)=\theta_0$, *if the following conditions are satisfied:*

$$\theta_0\lambda_1-K\lambda_2-\int_0^c \hat f_0(u)\,d\{\widehat F_0(u)-X(u)\}=0, \tag{4.6}$$

$$\lambda_2+\int_t^c d\{\widehat F_0(u)-X(u)\}\ge 0, \text{ for all } t\in(0,c], \tag{4.7}$$

and

$$\lambda_1-\lambda_2=\int_0^c d\{\widehat F_0(u))-X(u)\}\,. \tag{4.8}$$

Proof. For monotone functions $f:[0,c]\to\mathbb{R}$, let $\phi_R(f)$ be defined by (4.1), and let the function $\psi_{\lambda_1,\lambda_2}$ be defined by

$$\psi_{\lambda_1,\lambda_2}(f)=\phi_R(f)+\lambda_1\{\theta_0-f(0)\}+\lambda_2\{f(c)-K\}$$

where we define $f(0)$ by $f(0)=\lim_{u\downarrow 0}f(u)$. Then we have, for λ_1 and λ_2, defined by (4.4) and (4.5),

$$\psi_{\lambda_1,\lambda_2}(\hat f_0)=\phi(\hat f_0).$$

To see this, note that, by the definitions of λ_1 and λ_2, λ_1 can only be different from zero if $f_0(0)=\theta_0$, and likewise λ_2 can only be different from zero if $f_0(c)=K$. But (4.6) to (4.8) are exactly the Fenchel conditions for minimizing $\psi_{\lambda_1,\lambda_2}(f)$ over all monotone functions f. Hence we get, for all monotone functions f on $[0,c]$ such that $|f|\le K$ and $f(0)\ge\theta_0$:

$$\phi(\hat f_0)=\psi_{\lambda_1,\lambda_2}(\hat f_0)\le\psi_{\lambda_1,\lambda_2}(f)\le\phi(f).$$

Hence $\widehat{f}_0$ minimizes $\phi_R(f)$ over all such functions f.

Now we show that (4.6) to (4.8) are in fact the Fenchel conditions. If we perturb the solution $\widehat{f}_0$ by a monotone function h, we find that $\widehat{f}_0$ satisfies

$$
\begin{aligned}
(4.9) \qquad 0 &\le \frac{d}{d\epsilon}\psi_{\lambda_1,\lambda_2}(\widehat{f}_0 + \epsilon h)|_{\epsilon=0} \\
&= \int_0^c h(u)\widehat{f}_0(u)du - \int_0^c h(u)dX(u) - \lambda_1 h(0) + \lambda_2 h(c).
\end{aligned}
$$

If the functions $\widehat{f}_0 + \epsilon h$ are monotone for $|\epsilon| \le \epsilon_0$ for some $\epsilon_0 > 0$, then (4.9) holds with equality. Now we get (4.6) by choosing $h = \widehat{f}_0$ (and noting that equality then holds in (4.9)); (4.8) follows by choosing $h = 1_{[0,c]}$; and (4.7) follows by choosing $h = 1_{[t,c]}$, $t > 0$. □

4.2. *Extension of the solution $\widehat{f}_0$ from $[-c, c]$ to $\mathbb{R}$.* Now suppose that $f_0(t) = f_{can}(t) \equiv 2t$, and we let $c \to \infty$ (and $K = K_c \equiv 5c \to \infty$, $\lambda_2 \to 0$): Then the conditions (4.6) - (4.8) of Theorem 4.1 become:

$$
(4.10) \qquad \theta_0\lambda_1 - \int_0^\infty \widehat{f}_0(u)\, d\{\widehat{F}_0(u) - X(u)\} = 0,
$$

$$
(4.11) \qquad \int_t^\infty d\{\widehat{F}_0(u) - X(u)\} \ge 0, \text{ for all } t \in (0, \infty),
$$

and

$$
(4.12) \qquad \lambda_1 = \int_0^\infty d\{\widehat{F}_0(u)) - X(u)\}\,.
$$

Replacing (4.12) in (4.10) we find that

$$
\int_0^\infty \widehat{f}_0(u)\, d\{\widehat{F}_0(u) - X(u)\} = \theta_0 \int_0^\infty d(\widehat{F}_0(u) - X(u))\,.
$$

This can be viewed as exactly the condition obtained by Banerjee [1] in a particular finite n situation; see also [2].

Let $X(t) = X_{1,1}(t) \equiv W(t) + t^2$ where $W(t)$ is standard two-sided Brownian motion starting from 0. For constrained estimation of a monotone function f in Gaussian white noise, the following theorem is basic.

Now consider estimation of a monotone function f in Gaussian white noise subject to the constraint that $f(0) = \theta_0$. By piecing together the solutions on the right and left as characterized in Section 4.1, we obtain the following result.

THEOREM 4.2. *There exists an almost surely uniquely defined random function $\widehat{F}_0 \equiv \widehat{F}_{\theta_0}$ satisfying the following conditions:*
(i) The function $\widehat{F}_0$ is everywhere below the function X:

$$
(4.13) \qquad \widehat{F}_0(t) \le X(t), \text{ for each } t \in \mathbb{R}.
$$

(ii) $\widehat{F}_0$ *has a monotone left derivative* $\widehat{f}_0$ *satisfying* $\widehat{f}_0(0) = \theta_0$.
(iii) The function $\widehat{F}_0$ *satisfies*

$$\int_{\mathbb{R}} \widehat{f}_0(t)d(\widehat{F}_0 - X)(t) = \theta_0 \int_{\mathbb{R}} d(\widehat{F}_0 - X)(t)\,. \tag{4.14}$$

In fact, $\widehat{F}_0$ also has a greatest convex minorant interpretation: For positive values of t, $\widehat{F}_0(t)$ is the greatest convex minorant of the process $\{X(t) : t > 0\}$ subject to having slope always greater than or equal to θ_0; similarly, for $t \le 0$, $\widehat{F}_0(t)$ is the greatest convex minorant of the process $\{X(t) : t \le 0\}$ subject to having slope always less than or equal to θ_0. Thus $\widehat{F}_0$ is continuous on the two sets $(0, \infty)$ and $(-\infty, 0)$, has a jump discontinuity at 0, but will always have left derivative $\widehat{f}_0(0) = \theta_0$ at 0. Note that $\widehat{F}$ and $\widehat{F}_0$ will be equal (and have equal derivatives) on the complement of a (random!) neighborhood of 0. Thus in forming the likelihood ratio, the only contribution will come from the interval containing 0 where the functions $\widehat{F}$ and $\widehat{F}_0$ differ.

When $\theta_0 = 0$, we obtain the following important corollary:

COROLLARY 4.1. *There exists an almost surely uniquely defined random function* $\widehat{F}_0$ *satisfying the following conditions:*
(i) The function $\widehat{F}_0$ *is everywhere below the function* X*:*

$$\widehat{F}_0(t) \le X(t), \textit{ for each } t \in \mathbb{R}. \tag{4.15}$$

(ii) $\widehat{F}_0$ *has a monotone left derivative* $\widehat{f}_0$ *satisfying* $\widehat{f}_0(0) = 0$.
(iii) The function $\widehat{F}_0$ *satisfies*

$$\int_{\mathbb{R}} \{X(t) - \widehat{F}_0(t)\}\, d\widehat{f}_0(t) = 0\,. \tag{4.16}$$

Clearly $\widehat{F}_0$ characterized by Corollary 4.1 also has a greatest convex minorant interpretation: For positive values of t, $\widehat{F}_0(t)$ is the greatest convex minorant of the process $\{X(t) : t > 0\}$ subject to having slope always greater than or equal to 0; similarly, for $t \le 0$, $\widehat{F}_0(t)$ is the greatest convex minorant of the process $\{X(t) : t \le 0\}$ subject to having slope always less than or equal to 0. Thus $\widehat{F}_0$ is continuous on the two sets $(0, \infty)$ and $(-\infty, 0)$, has a jump discontinuity at 0, but will always have left derivative $\widehat{f}_0(0) = 0$ at 0. Note that $\widehat{F}$ and $\widehat{F}_0$ will be equal (and have equal derivatives) on the complement of a (random!) neighborhood of 0. Thus in forming the likelihood ratio, the only contribution will come from the interval containing 0 where the functions $\widehat{F}$ and $\widehat{F}_0$ differ.

Theorem 4.2 can be proved by the same methods used to prove Theorem 2.1 in [17]. The basic idea is that when $c \to \infty$ (and $K = K_c \to \infty$, the effects of the constraints at the endpoint c washes out, and the resulting characterizing equations come from (4.6) - (4.8) with $\lambda_2 = 0$ and $c = \infty$.

Figures 1 - 3 illustrate Theorems 3.2 and 4.2.

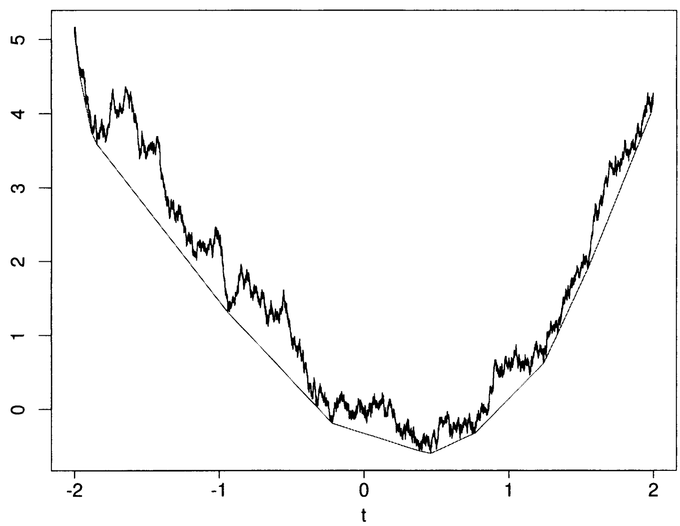

FIG. 1. *The Greatest Convex Minorant* $\widehat{F} \equiv \widehat{F}_{1,1}$ *and* $W(t) + t^2$.

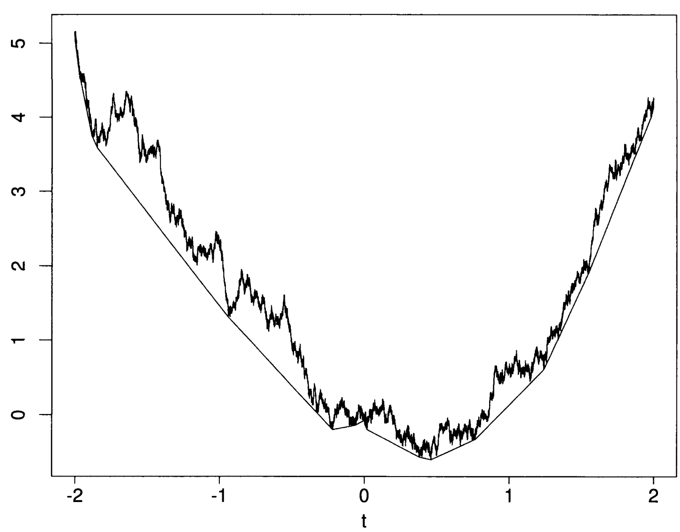

FIG. 2. *The one-sided convex minorants* $\tilde{F}_L$ *and* $\tilde{F}_R$ *and* $W(t) + t^2$.

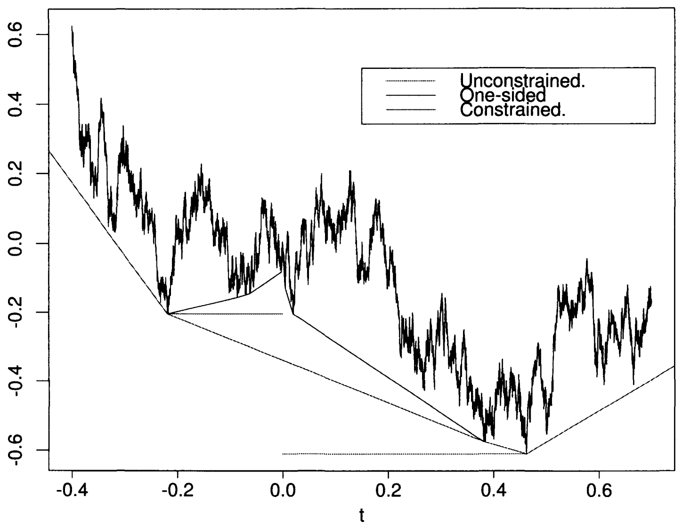

FIG. .3 *Close-up view of* $\widehat{F}_{1,1}$, $\tilde{F}_{L,R}$, $\widehat{F}^0_{1,1}$, *and* $W(t) + t^2$.

5. The Likelihood Ratio Statistic. We now consider the consequence of Theorems 3.2 and 4.2 for the likelihood ratio test of $H_0 : f(0) = 0$ versus $H_1 : f(0) \neq 0$.

Recall that by the Cameron-Martin-Girsanov theorem (see e.g. [33], page 81), the Radon-Nikodym derivative of P_f with respect to P_0 considered as laws of the process $\{X(t) \equiv W(t) + F(t) : t \in [-c, c]\}$, is given by

$$\frac{dP_f}{dP_0} = \exp\left(\int_{-c}^{c} f dX - \tfrac{1}{2} \int_{-c}^{c} f^2(t) dt \right) . \tag{5.1}$$

THEOREM 5.1. *For testing the null hypothesis* $H_0 : f(0) = 0$ *versus the alternative* $H_1 : f(0) \neq 0$, *based on observation of the process* $\{X(t) : t \in \mathbb{R}\}$, *the likelihood ratio statistic is*

$$2 \log \lambda = \int_D \left\{ \widehat{f}^2(t) - \widehat{f}_0^2(t) \right\} dt \equiv \mathbb{D} \tag{5.2}$$

where $D \equiv \{t \in R : \ \widehat{f}(t) \neq \widehat{f}_0(t)\}$.

Proof. Let $\mathcal{F}(c, K)$ denote the class of monotone functions on $[-c, c]$ with $c \leq K$, and let $\mathcal{F}_0(c, K)$ be the corresponding subclass of $\mathcal{F}(c, K)$ satisfying $f(0) = 0$. Then by (5.1) and Theorems 3.1 and 4.1 it follows immediately that

$$2 \log \lambda_c = 2 \log \left(\frac{\sup_{f \in \mathcal{F}(c,K)} dP_f / dP_0}{\sup_{f \in \mathcal{F}_0(c,K)} dP_f / dP_0} \right)$$

$$= 2\log\left(\frac{dP_{\widehat{f}}/dP_0}{dP_{\widehat{f}_0}/dP_0}\right)$$

$$= 2\left\{\int_{-c}^{c}\widehat{f}_c dX - \tfrac{1}{2}\int_{-c}^{c}\widehat{f}_c^2(t)dt - \int_{-c}^{c}\widehat{f}_{c,0}dX + \tfrac{1}{2}\int_{-c}^{c}\widehat{f}_{c,0}^2(t)dt\right\}$$

$$= 2\int_{-c}^{c}(\widehat{f}_c - \widehat{f}_{c,0})dX - \int_{-c}^{c}[\widehat{f}_c^2(t) - \widehat{f}_{c,0}^2(t)]dt\,. \tag{5.3}$$

Now consider taking the limit across (5.3) as $c \to \infty$ (and $K_c \equiv 5c \to \infty$). Then, with $2\log\lambda \equiv \lim_{c\to\infty} 2\log\lambda_c$, we find that

$$2\log\lambda = 2\int_D(\widehat{f} - \widehat{f}_0)dX - \int_D[\widehat{f}^2(t) - \widehat{f}_0^2(t)]dt \tag{5.4}$$

where the functions $\widehat{f}$ and $\widehat{f}_0$ are characterized in Theorem 3.2 and Corollary 4.1 respectively. But from part (iii) of Theorem 3.2 and Corollary 4.1,

$$\int_{\mathbb{R}}(X - \widehat{F})d\widehat{f} = 0 \qquad \text{and} \qquad \int_{\mathbb{R}}(X - \widehat{F}_0)d\widehat{f}_0 = 0\,. \tag{5.5}$$

Hence, via integration by parts,

$$\begin{aligned}\int_{\mathbb{R}}(\widehat{f} - \widehat{f}_0)dX &= \int_D(\widehat{f} - \widehat{f}_0)dX = -\int_D Xd(\widehat{f} - \widehat{f}_0)\\ &= -\int_D \widehat{F}d\widehat{f} + \int_D \widehat{F}_0 d\widehat{f}_0 \qquad \text{by (5.5)}\\ &= \int_D \widehat{f}d\widehat{F} - \int_D \widehat{f}_0 d\widehat{F}_0 \qquad \text{by integration by parts}\\ &= \int_D \widehat{f}^2(t)dt - \int_D \widehat{f}_0^2(t)dt\,. \qquad (5.6)\end{aligned}$$

Substitution of (5.6) in (5.4) yields the claim:

$$2\log\lambda = \int_D[\widehat{f}^2(t) - \widehat{f}_0^2(t)]dt\,.$$

□

The importance of Theorem 5.1 is that the limiting distributions of likelihood ratio statistics for tests concerning nonparametric estimation of monotone functions will be exactly the distribution of $\mathbb{D}$ given in (5.2). For example, consider estimation of a distribution function F based on current status (or case 1 interval censored) data. Suppose that (X_i, T_i) , $i = 1\ldots, n$, are i.i.d., where for each pair X_i and T_i are independent, $X_i \sim F$ and $T_i \sim G$ where F and G are distribution functions on $[0, \infty)$. For each pair we observe $Y_i = (T_i, \Delta_i)$ where $\Delta_i = 1\{X_i \le T_i\}$. The goal is to make inference about the monotone (increasing) function F. The nonparametric maximum likelihood estimator $\mathbb{F}_n$ of F is well known; see e.g. [19] where it is shown that if F and G have a densities f and g at t_0 with $f(t_0) > 0$, $g(t_0) > 0$, then

$$n^{1/3}(\mathbb{F}_n(t_0) - F(t_0)) \to_d \left(\frac{F(t_0)(1 - F(t_0))f(t_0)}{2g(t_0)}\right)^{1/3}\mathbb{S}(0)\,.$$

We are interested here in likelihood ratio tests of $H_0 : F(t_0) = \theta_0$ versus $H_1 : F(t_0) \neq \theta_0$ for $t_0 \in (0, \infty)$ and $\theta_0 \in (0, 1)$ fixed.

The log-likelihood ratio statistic for testing $H_0 : F(t_0) = \theta_0$ versus $H_1 : F(t_0) \neq \theta_0$ is

$$2 \log \lambda_n = 2n\mathbb{P}_n \left\{ \Delta \log \frac{\mathbb{F}_n}{\mathbb{F}_n^0}(T) + (1 - \Delta) \log \frac{1 - \mathbb{F}_n}{1 - \mathbb{F}_n^0}(T) \right\} \tag{5.7}$$

where $\mathbb{F}_n$ and $\mathbb{F}_n^0$ are the unconstrained and constrained maximum likelihood estimators of F respectively.

THEOREM 5.2. *Under the null hypothesis H_0, if F and G are differentiable at t_0 with strictly positive densities $f(t_0)$ and $g(t_0)$ respectively, then*

$$2 \log \lambda_n \to_d \mathbb{D} \tag{5.8}$$

where $\mathbb{D}$ is given in (5.2).

Theorem 5.2 is proved in [2]. Note that Theorem 5.2 says that $2 \log \lambda_n$ is asymptotically distribution free. This means that we can use the asymptotic distribution to obtain asymptotically valid confidence intervals for $F(t_0)$ by inverting the likelihood ratio test: letting $2 \log \lambda_n(\theta)$ denote the test statistic for testing $H_0 : F(t_0) = \theta$, and letting s_α be the upper αth percentage point of the distribution of S, an approximate $1 - \alpha$ confidence interval for $F(t_0)$ is given by

$$\{\theta : 2 \log \lambda_n(\theta) \leq s_\alpha\} .$$

These confidence bounds are explored in more detail in [1] and [3].

6. Some Open Problems. Questions:

1. Can we determine the distribution of $\mathbb{D}$ analytically using the methods of [14], [15], and [16]? The distribution has been estimated via Monte-Carlo methods in [2], but it would be very desirable to compute this distribution analytically.
2. Can we get asymptotically valid confidence bands for the whole monotone function f in the white-noise setting?
3. Does a limit theorem like that in Theorem 5.1 hold for the other problems listed as examples in [20]?
4. Does this approach to likelihood ratio tests and confidence intervals extend to the setting of convex functions treated in [17] and [18]?

A Bivariate Problem:
Suppose that we want to estimate a bivariate monotone function f in Gaussian white noise:

$$dX(\underline{t}) = f(\underline{t})dt_1 dt_2 + \sigma dW(\underline{t}), \qquad \underline{t} \in [-c, c] \times [-c, c] . \tag{6.1}$$

Here "monotonicity" of f will be meant in the sense that

$$\Delta_2(f)(\underline{s}, \underline{t}] \equiv f(t_1, t_2) - f(t_1, s_2) - f(s_1, t_2) + f(s_1, s_2) \geq 0$$

for all $\underline{s} = (s_1, s_2), \underline{t} = (t_1, t_2) \in [-c, c] \times [-c, c]$, and W can be taken to be a (quadruple) Brownian sheet (i.e four independent Brownian sheets, one on each of the four natural orthants contained in $[-c, c] \times [-c, c]$. It seems that a natural candidate for a "canonical monotone function" in this setting is the function $4t_1t_2$, so that

$$X(\underline{t}) = t_1^2 t_2^2 + W(\underline{t}) .$$

• What is the MLE of f (under some suitable constraints guaranteeing compactness) based on observation of $X(\underline{t})$, $\underline{t} \in [-c, c] \times [-c, c]$?
• What is the MLE of f based on observation of $X(\underline{t})$, $\underline{t} \in \mathbb{R}^2$?

This "white-noise model" is one that arises in connection with estimation of a bivariate distribution function based on bivariate interval censored data; see e.g. [34].

7. Acknowledgemetns The author owes a profound debt to Jack Hall for encouraging his early efforts at research. Much of the material in this paper has developed as a result of conversations and collaboration with Piet Groeneboom. The figures in Section 4 were provided by Moulinath Banerjee.

REFERENCES

[1] M. Banerjee. *Likelihood Ratio Inference in Regular and Nonregular Problems.* PhD thesis, Department of Statistics, University of Washington, 2000.

[2] M. Banerjee and J.A. Wellner. Likelihood ratio tests for monotone functions. *Annals of Statistics*, 29:1699 – 1731, 2001.

[3] M. Banerjee and J.A. Wellner. Pointwise confidence sets for the interval censoring model. Technical report, Department of Statistics, University of Washington, 2003. (in preparation).

[4] L. Birgé and P. Massart. Gaussian model selection. *Journal of the European Mathematical Society*, 3:203 – 268, 2001.

[5] L.D. Brown and M.G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, 24:2384–2398, 1996.

[6] L.D. Brown and C-H. Zhang. Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Annals of Statistics*, 26:279–287, 1998.

[7] H.D. Brunk. Estimation of isotonic regression. In M.L. Puri, editor, *Nonparametric Techniques in Statistical Inference.* Cambridge University Press, Cambridge, 1970.

[8] L. Cavalier and A.B. Tsybakov. Sharp adaptation for inverse problems with random noise. Preprint, 2000.

[9] D. Donoho. Asymptotic minimax risk (for sup-norm loss): solution via optimal recovery. *Probability Theory and Related Fields*, 99:145 – 170, 1994.

[10] D. Donoho and M. Low. Renormalization exponents and optimal pointwise rates of convergence. *Annals of Statistics*, 20:944–970, 1992.

[11] D.L. Donoho and R.C. Liu. Geometrizing rates of convergence, iii. *Annals of Statistics*, 19:668–701, 1991.

[12] L. Dümbgen and V.G. Spokoiny. Multiscale testing of qualitative hypotheses. *Annals of Statistics*, 29:124 – 152, 2001.

[13] S. Efromovich. On global and pointwise adaptive estimation. *Bernoulli*, 4:273 – 282, 1998.

[14] P. Groeneboom. The concave majorant of Brownian motion. *Annals of Probability*, 11:1016–1027, 1983.

[15] P. Groeneboom. Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 539 – 554. Wadsworth Advanced Books and Software, Monterey, CA, 1985.

[16] P. Groeneboom. Brownian motion with a parabolic drift and airy functions. *Probability Theory and Related Fields*, 81:79 – 109, 1988.

[17] P. Groeneboom, G. Jongbloed, and J.A. Wellner. A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Annals of Statistics*, 29:1620 - 1652, 2001.

[18] P. Groeneboom, G. Jongbloed, and J.A. Wellner. Estimation of convex functions: characterizations and asymptotic theory. *Annals of Statistics*, 29:1653 - 1698, 2001.

[19] P. Groeneboom and J.A. Wellner. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhauser, Boston, 1992.

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

[21] I.A. Ibragimov and R.S. Khasminskii. On the estimation of an infinite dimensional parameter in gaussian white noise. *Soviet Mathematics Doklady*, 236:1053–1055, 1977.

[22] I.A. Ibragimov and R.S. Khasminskii. *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York, 1981.

[23] I.A. Ibragimov and R.S. Khasminskii. On nonparametric estimation of the value of a linear functional in gaussian white noise. *Theory of Probability and its Applications*, 29:18 - 32, 1984.

[24] Yu.I. Ingster. Minimax nonparametric detection of signals in white gaussian noise. *Problems of Information Transmission*, 18:130–140, 1982.

[25] A.P. Korostelev. Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory of Probability and its Applications*, 38:775–782, 1993.

[26] A.P. Korostelev and M. Nussbaum. The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, 5:1099 - 1118, 1999.

[27] V. Kotel'nikov. *The Theory of Optimum Noise Immunity*. McGraw Hill, New York, 1959.

[28] Yu.A. Kutoyants. On a problem of testing hypotheses and asymptotic normality of stochastic integrals. *Theory of Probability and its Applications*, 20:376–384, 1975.

[29] O.V. Lepskii. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35:454–466, 1990.

[30] M. Nussbaum. Asymptotic equivalence of density estimation and white noise. *Annals of Statistics*, 24:2399 - 2430, 1996.

[31] M.S. Pinsker. Optimal filtering of square integrable signals in gaussian white noise. *Problems of Information Transmission*, 16:120 - 133, 1980.

[32] B.L.S. Prakasa Rao. Estimation of a unimodal density. *Sankhȳa Series A*, 31:23–36, 1969.

[33] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 2. Wiley, New York, 1987.

[34] S. Song. *Estimation with bivariate interval-censored data*. PhD thesis, University of Washington, Department of Statistics, 2001.

[35] V.G. Spokoiny. Adaptive hypothesis testing using wavelets. *Annals of Statistics*, 24:2477–2498, 1996.

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
jaw@stat.washington.edu