The Annals of Statistics
1992, Vol. 20, No. 1, 545–554

RENORMALIZATION AND WHITE NOISE APPROXIMATION FOR NONPARAMETRIC FUNCTIONAL ESTIMATION PROBLEMS

By Mark G. Low

University of Pennsylvania and University of California, Berkeley

White noise models often renormalize exactly yielding optimal rates of convergence for pointwise nonparametric functional estimation problems. Similar rescaling ideas lead to a sequence of experiments appropriate for pointwise density estimation problems.

1. Introduction. The objective of this paper is to introduce two conceptual approaches to nonparametric functional estimation problems, renormalization and white noise approximation. Renormalization takes on its simplest form for pointwise estimation problems when our observations arise from the following white noise model:

$$dX_t = f(t)\,dt + \frac{1}{\sqrt{n}}\,dW_t, \qquad f \in \mathbf{F}, \tag{1}$$

where W_t is Brownian motion.

In Section 2, we show how invariance ideas can often clarify rates of convergence results as $n \to \infty$ for a variety of parameter spaces $\mathbf{F}$. Extensions and generalisations of these results can be found in Donoho and Low (1990a) and Low (1991).

Millar [1979] has illustrated the power of looking at nonparametric problems from the viewpoint of Le Cam's theory of experiments, with the parameter space indexed by an infinite-dimensional Hilbert space. This highly successful approach led to a simple unified theory for estimating distribution functions under, for example, sup norm loss. The power of Hilbert space parametrizations has not, however, been exploited in pointwise estimation problems arising, for example, from density estimation, nonparametric regression or estimation of a variable intensity function from a Poisson process. In fact, it is only recently that Le Cam's theory of experiments has even been brought to bear on the problem of density estimation, and then only through a sequence of one-dimensional parametrizations. Donoho and Liu (1991) however did pick these one-dimensional parametrizations in an optimal way and showed the power of Le Cam's methodology. Romano (1988) has also applied Le Cam's theory to sequences of one-dimensional experiments naturally arising from estimation of the mode.

Received January 1990; revised May 1991.

AMS 1980 *subject classifications*. Primary 62G07; secondary 62C20.

Key words and phrases. Nonparametric functional estimation, renormalization, white noise approximation, density estimation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

The invariance ideas of Section 2 lead naturally to the introduction, in Section 3, of a new sequence of experiments appropriate for pointwise density estimation problems. The original inspiration for the consideration of this sequence of experiments was Donoho and Liu (1991). However it can also be looked upon as a generalisation of some sequences given by Millar (1979). The limiting experiment for this new sequence of experiments is the white noise model considered in Section 2. The main theorem in Section 3 gives a precise statement of a local asymptotic equivalence of density estimation and white noise models. The area of white noise approximation is developed further in Brown and Low (1990) and Donoho and Low (1990).

2. In this section we focus on the following white noise model:

$$dY_t = f(t)\,dt + \frac{1}{\sqrt{n}}\,dW_t, \qquad t \in D. \tag{2}$$

We shall always assume $D = (-\infty, \infty)$ or $D = [0, \infty)$. (2) induces a statistical experiment (for each n) when we let $f \in \mathbf{F} \subseteq L_2(D)$. We focus on pointwise estimation problems.

For these problems it is convenient to introduce a second sequence of statistical experiments generated by

$$dY_t^n = \frac{f(\beta_n t)}{\alpha_n}\,dt + \frac{1}{\sqrt{n}}\,dW_t, \qquad t \in D, \tag{3}$$

where $\alpha_n^2 \beta_n = n$ and $f \in \mathbf{F}$. We shall sometimes write $Y^n(t)$ for Y_t^n when the resulting expression is easier to read.

The importance of this second sequence of experiments will be clear from the lemma given below and the remarks following it. Its proof is clear and so is left to the reader.

LEMMA 1. *Suppose $Y^n(t)$ has a distribution given by* (3). *Then*

$$Z(t) = \alpha_n \beta_n Y^n(t/\beta_n) \tag{4}$$

follows a distribution given by

$$dZ(t) = f(t)\,dt + dW_t. \tag{5}$$

Similarly, if $Z(t)$ has a distribution given by (5), *then $Y^n(t)$ defined by* (4) *has a distribution given by* (3).

REMARK. The lemma establishes a precise equivalence between every pair of experiments in the sequence of experiments given in (3).

To connect the lemma with the more interesting sequence of experiments given by (2), we need to take a more decision-theoretic viewpoint and introduce loss functions. In fact we will allow the loss function to depend on n subject to the following condition.

ASSUMPTION A. We restrict attention to a sequence of loss functions l_n and a fixed loss function l such that (i) l_n: $L_2[D] \times R \to R^+$, l: $L_2[D] \times R \to R^+$; (ii) there is a function g: $R \times R \to R$ and a function h: $R \times R \to R$ such that if $\alpha_n^2 \beta_n = n$, then

$$l_n\left(\frac{f(\beta_n t)}{\alpha_n}, h(\alpha_n, \beta_n)a\right) = g(\alpha_n, \beta_n) l(f, a).$$

Now let δ be an estimator δ: $L_2[D) \to R$. $E_g^m l_n(f, \delta(X(t)))$ is then to be interpreted as taking the expectation under the model

$$dX_t = g(t)\,dt + \frac{1}{\sqrt{m}}\,dW_t \tag{6}$$

of the random function $l_n(f, \delta)$. When $g = f$ and $m = n$, this is the risk of the estimator δ with loss function l_n under model (2).

THEOREM 1. *Let $(T_n f)(t) = (f(\beta_n t)/\alpha_n)$, where $\alpha_n^2 \beta_n = n$ and for each estimator δ, let δ_n be defined by*

$$h(\alpha_n, \beta_n)\delta(Z(t)) = \delta_n\left(\frac{Z(\beta_n t)}{\alpha_n \beta_n}\right). \tag{7}$$

Then

$$E_{T_n f}^n l_n(T_n f, \delta_n) = g(\alpha_n, \beta_n) E_f^1 l(f, \delta), \tag{8}$$

$$\sup_{f_n \in T_n \mathbf{F}} E_{f_n}^n l_n(f_n, \delta_n) = g(\alpha_n, \beta_n) \sup_{f \in \mathbf{F}} E_f^1 l(f, \delta), \tag{9}$$

$$\inf_{\delta_n} \sup_{f_n \in T_n \mathbf{F}} E_{f_n}^n l_n(f_n, \delta_n) = g(\alpha_n, \beta_n) \inf_{\delta} \sup_{f \in \mathbf{F}} E_f^1 l(f, \delta). \tag{10}$$

COROLLARY 1. *If $T_n \mathbf{F} = \mathbf{F}$, then*

$$\inf_{\delta_n} \sup_{f \in \mathbf{F}} E_f^n l_n(f, \delta_n) = g(\alpha_n, \beta_n) \inf_{\delta} \sup_{f \in \mathbf{F}} E_f^1 l(f, \delta). \tag{11}$$

REMARK. Corollary 1, of course, follows immediately from equation (10) of Theorem 1. It yields $g(\alpha_n, \beta_n)$ as an optimal rate of convergence as long as $\inf_\delta \sup_{f \in \mathbf{F}} E_f^1 l(f, \delta) < \infty$. Equation (7) also defines a sequence δ_n which attains this rate as long as $\sup_{f \in \mathbf{F}} E_f^1 l(f, \delta) < \infty$.

APPLICATIONS. We now give a few examples of how to apply the above theorem. Similar results will be given later for density estimation problems.

EXAMPLE 1. Write $f^k(x)$ for the kth derivative of f. Let $\mathbf{F}(k, M) = \{f \in L_2(0, \infty)$: $|f^k(x) \le M \ \forall\ x\}$ and let

$$l_n \equiv l \quad \text{satisfy} \quad l(f, a) = |f^j(0) - a|^q, \quad \text{where } 0 \le j < k.$$

Furthermore, take $\alpha_n = n^{k/2k+1}$, $\beta_n = n^{1/2k+1}$. Then $\alpha_n^2\beta_n = n$, $T_n\mathbf{F} = \mathbf{F}$ and

$$l\left(\frac{f(\beta_n t)}{\alpha_n}, \frac{\beta_n^j}{\alpha_n}a\right) = \frac{\beta_n^{qj}}{\alpha_n^q} l(f, a).$$

The assumptions of Corollary 1 then clearly hold and yield

$$\inf_{\delta_n} \sup_{f \in \mathbf{F}(k,M)} E_f^n |f^j(0) - \delta_n|^q = \frac{1}{n^{q(k-j)/(2k+1)}} \inf_{\delta} \sup_{f \in \mathbf{F}(k,M)} E_f^1 |f^j(0) - \delta|^q.$$

Now let $\alpha_1 = M^{-1/(2k+1)}$, $\beta_1 = M^{2/2k+1}$. Then $T_1\mathbf{F}(k,1) = \mathbf{F}(k,M)$ and equation (10) of Theorem 1 yields

$$\inf_{\delta} \sup_{f \in \mathbf{F}(k,M)} E_f^1 |f^j(0) - \delta|^q = M^{(q(2j+1))/(2k+1)} \inf_{\delta} \sup_{f \in \mathbf{F}(k,1)} E_f^1 |f^j(0) - \delta|^q.$$

EXAMPLE 2. As a simple case of an application with a varying l_n, take the parameter space to be

$$\mathbf{F}(1, M), \qquad l(f, a) = (f(1) - a)^2, \qquad \alpha_n = n^{1/3}, \qquad \beta_n = n^{1/3}$$

and $l_n(f, a) = (f(n^{-1/3}) - a)^2$. Then $g(\alpha_n, \beta_n) = (1/\alpha_n)$ and it follows from Corollary 1 that

$$\inf_{\delta_n} \sup_{f \in \mathbf{F}(1,M)} E_f^n (f(n^{-1/3}) - \delta_n)^2 + \frac{1}{n^{2/3}} \inf_{\delta} \sup_{f \in \mathbf{F}(1,M)} E_f^1 (f(1) - \delta)^2.$$

EXAMPLE 3. Let $\mathbf{G}(M) = \{f \in L^2[-\infty, \infty): \int f'^2 \le M\}$. Take $l_n \equiv l$ such that $l(f, a) = |f(0) - a|^q$. Let $\alpha_n = n^{1/4}$, $\beta_n = n^{1/2}$. Then $T_n\mathbf{G}(M) = \mathbf{G}(M)$ and Corollary 1 yields

$$\inf_{\delta_n} \sup_{f \in \mathbf{G}(M)} E_f^n |f(0) - \delta_n|^q = \frac{1}{n^{q/4}} \inf_{\delta} \sup_{f \in \mathbf{G}(M)} E_f^1 |f(0) - \delta|^q.$$

3. A sequence of experiments. The main theorem in this section is given in terms of the convergence of a sequence of experiments to the white noise model of Section 2. In particular, if we let $\mathbf{H} = \{h: R \to R, \int h^2 < \infty, \sup_x |h(x)| < \infty\}$, then the limiting experiment is given by

$$dX_t = h(t)\,dt + dW_t, \qquad -\infty < t < \infty, h \in \mathbf{H}. \tag{12}$$

This is a representation of the standard Gaussian shift experiment on $\mathbf{H}$ which we shall write as $G = \{G_h: h \in \mathbf{H}\}$. In other words, the theorem states that the distributions of the likelihood ratios for the sequence of experiments converges weakly to the distributions of the likelihood ratios for the limiting experiment. A more detailed and complete description for the reader unfamiliar with this idea can be found either in Le Cam (1986) or Millar (1979). The reader should also recall that under G_0 the distribution of $\log(dG_h/dG_0)$ is $N(-(\int h^2(x)\,dx)/2, \int h^2(x)\,dx)$.

We now introduce our sequence of experiments. First fix a probability density f_0 on R such that f_0 is continuous at 0, $f_0(0) > 0$, and $\sup_x f_0(x) < \infty$. Corresponding to f_0 and $\{\alpha_n\}_{n=1}^{\infty}$, $\{\beta_n\}_{n=1}^{\infty}$, any nondecreasing sequence of positive numbers satisfying $\lim_{n\to\infty} \alpha_n = \infty$ and

$$\frac{\alpha_n^2 \beta_n}{f_0(0)n} \to 1 \tag{13}$$

will be the following sequence of experiments. For $h \in \mathbf{H}$, let

$$h_n = \int \frac{h(\beta_n x)}{\alpha_n} f_0(x)\, dx. \tag{14}$$

h_n is finite since h and f_0 are square integrable. Furthermore, the conditions we imposed on h and f_0 imply that $h_n = O((\alpha_n \beta_n)^{-1})$. If

$$1 + \frac{h(\beta_n x)}{\alpha_n} - h_n \geq 0 \quad \text{for all } x, \tag{15}$$

define

$$f_n(h;x) = \left(1 + \frac{h(\beta_n x)}{\alpha_n} - h_n\right) f_0(x). \tag{16}$$

Otherwise, let

$$f_n(h;x) = f_0(x). \tag{17}$$

Defining $f_n(h;x)$ by (17) when (15) is not satisfied is only a technical condition. Its only purpose is to make the sequence of experiments given below to have parameter space $\mathbf{H}$ for each n. Note that since $\sup_x |h(x)| < \infty$ and α_n increases to infinity, for any given h, $f_n(h;x)$ is defined by (16) for all sufficiently large n. Finally, define P_h^n to be the probability on R^n having density

$$\prod_{i=1}^{n} f_n(h; x_i). \tag{18}$$

The collection $\{P_h^n \colon h \in \mathbf{H}\}$ now defines an experiment for each n.

THEOREM 2. *The sequence of experiments* $\{P_h^n \colon h \in \mathbf{H}\}$ *constructed above converges weakly to the standard Gaussian experiment* $\{G_h \colon h \in \mathbf{H}\}$.

The importance of this theorem is contained in the following corollary which is just a statement of the Hajek–Le Cam minimax theorem in the present context combined with Lindae's theorem [see page 92 of Le Cam (1986)].

First, we need to establish some notation which we shall use throughout the rest of this paper. Write $\delta_n = \delta_n(X_1, \dots, X_n)$ to be any decision procedure based on n independent observations from a density $f_n(h)$, where $h \in \mathbf{H}$. Also by $E_h l(h, \delta_n)$, we mean the risk of the estimator (in estimating some scalar functional of h) when the density is $f_n(h)$. We write δ (no subscript) to be any decision procedure based on one observation from the Gaussian shift problem given in (12) and $E_h^G l(h, \delta)$ for the associated risk in estimating h.

COROLLARY. *Let* $\mathbf{K} \subset \mathbf{H}$. *If l is any loss function* $l\colon \mathbf{K} \times R \to R$, *lower semicontinuous in the second argument, then*

$$\liminf_{n\to\infty} \inf_{\delta_n} \sup_{h\in\mathbf{K}} E_h l(h, \delta_n) \geq \inf_{\delta} \sup_{h\in\mathbf{K}} E_h^G l(h, \delta). \tag{19}$$

If $\mathbf{K}$ *is compact in* $L_2(R)$ *under the usual norm* $\|f\|^2 = \int f^2$ *and the loss function is bounded, the inequality in* (19) *can be replaced by an equality.*

REMARK. Ibragimov and Khas'minskii (1991) define a concept of local asymptotic normality with norming factors A_ε, where $\{A_\varepsilon\}$ is a family of linear operators. The above results essentially fit into their framework if we define A_ε by $(A_\varepsilon h)(t) = h(\beta_\varepsilon t)/\alpha_\varepsilon$ for suitable choices of $\alpha_\varepsilon, \beta_\varepsilon$.

APPLICATIONS. We now give two simple applications of Theorem 2 similar to those found in Section 2. The results given here are not substantially new. Our main purposes is to exemplify the method and to illustrate by an example how α_n, β_n and $\mathbf{K}$ can be appropriately chosen. The rates given in our first example can also be found in Farrell (1972) and Stone (1980). In our second example we improve on some lower bounds given by Wahba (1975) for estimating a density function at a point under Sobolev constraints. Ibragimov and Hasminskii (1984) have previously shown that this rate holds in the white noise model. We should also mention that Millar (1979) has exploited Theorem 1, with $\alpha_n = (f_0(0))^{1/2} n^{1/2}$, $\beta_n = 1$ to obtain lower bounds for estimating distribution functions.

EXAMPLE 4. Suppose we observe $X_1, \dots, X_n$ i.i.d. with density $f \in \mathbf{F}$ and we want to estimate f at 0, where we use as a measure of loss l_p defined by

$$l_p(f, a) = |f(0) - a|^p. \tag{20}$$

There are two major obstacles to applying Theorem 2 in this context: (i) The loss function l_p is defined on the functions $f_n(h;\ )$ whereas the loss function in the corollary is defined on h. (ii) As mentioned above, we need to be able to choose α_n, β_n and $\mathbf{K}$ appropriately.

To answer these questions we consider particular classes of $\mathbf{F}$. Write $f^k(x)$ for the kth derivative of f. Let $\mathbf{F}(a, k, M) = \{f\colon R \to R^+;\ f(0) \leq a,\ \int f = 1,\ \sup_x |f^k(x)| \leq M\}$.

First fix some $f_0 \in \mathbf{F}(a, k, M)$ such that (i) $f_0(0) = b < a$; (ii) $|f_0^k(x)| < M$ for all x; (iii) for some $\varepsilon > 0$, $f_0(x) = f_0(0)$ for $|x| \leq \varepsilon$.

Conditions 1 and 2 make sure that f_0 is an interior point of the set $\mathbf{F}(a, k, M)$. Condition 3 facilitates the construction of the perturbations given below.

Let

$$\mathbf{K}(c) = \{h: |h^k(x)| \le 1, h(x) = 0 \text{ for } |x| \ge c\}.$$

We impose the condition that $h(x) = 0$ for $|x| \ge c$ primarily to make $\mathbf{K}(c)$ compact and hence insure strong convergence.

Now let

$$\alpha_n = M^{-1/(2k+1)}(f_0(0))^{(k+1)/(2k+1)} n^{k/(2k+1)},$$

$$\beta_n = M^{2/(2k+1)}(f_0(0))^{-1/(2k+1)} n^{1/(2k+1)}.$$

A few simple calculations which we leave to the reader [partly made easy by requiring $h(x) = 0$ for $|x| \ge c$], show that for some N, $f_n(h; x) \in \mathbf{F}(a, k, M)$ for all h when $n \ge N$. Note also that $\alpha_n^2 \beta_n = f_0(0)n$. These same calculations should also give the reader a good idea of why we imposed (i) and (ii) on f_0.

The corollary to Theorem 2 then yields

$$\liminf_{n\to\infty} \inf_{\delta_n} \sup_{h \in \mathbf{K}(c)} E_h l_p(h, \delta_n) \ge \inf_{\delta} \sup_{h \in \mathbf{K}(c)} E_h^G l_p(h, \delta). \tag{21}$$

We shall now connect equation (21) to the problem of estimating $f_n(h;\ )$ instead of h. Note that

$$l_p\left(f_0(x)\left(1 - + \frac{h(\beta_n x)}{\alpha_n}\right), f_0(0)\left(1 + \frac{\delta}{\alpha_n}\right)\right) = \left(\frac{f_0(0)}{\alpha_n}\right)^p l_p(h, \delta).$$

Hence

$$\begin{aligned}\liminf_{n\to\infty} &\left(\frac{\alpha_n}{f_0(0)}\right)^p \inf_{\delta_n} \sup_{h\in\mathbf{K}(c)} E_h l_p\left(f_0(x)\left(1 + \frac{h(\beta_n x)}{\alpha_n}\right), \delta_n\right) \\ &= \liminf_{n\to\infty} \inf_{\delta_n} \sup_{h\in\mathbf{K}(c)} E_h l_p(h, \delta_n) \\ &\ge \inf_{\delta} \sup_{h\in\mathbf{K}(c)} E_h^G l_p(h, \delta).\end{aligned}$$

Furthermore since $h_n = O((\alpha_n \beta_n)^{-1})$, it follows that

$$\begin{aligned}\liminf_{n\to\infty} &\left(\frac{\alpha_n}{f_0(0)}\right)^p \inf_{\delta_n} \sup_{h\in\mathbf{K}(c)} E_h l_p\left(f_0(x)\left(1 + \frac{h(\beta_n x)}{\alpha_n} - h_n\right), \delta_n\right) \\ &\ge \inf_{\delta} \sup_{h\in\mathbf{K}(c)} E_h^G l_p(h, \delta).\end{aligned}$$

Now let $c \to \infty$ and note that $\alpha_n = M^{-1/(2k+1)}(f_0(0))^{(k+1)/(2k+1)} n^{k/(2k+1)}$ and we get

$$\begin{aligned}\lim_{c\to\infty} \liminf_{n\to\infty} &M^{-p/(2k+1)}(f_0(0))^{-pk/(2k+1)} n^{pk/(2k+1)} \inf_{\delta_n} \sup_{h\in\mathbf{K}(c)} E_h l_p(f_n(h), \delta_n) \\ &\ge \lim_{c\to\infty} \inf_{\delta} \sup_{h\in\mathbf{K}(c)} E_h^G l_p(h, \delta).\end{aligned}$$

Note that since for $h \in K(c)$, $f_n(h) \in \mathbf{F}(a, k, M)$ this last equation immediately yields on taking $\sup_{f \in \mathbf{F}} f(0)$

$$\liminf_{n\to\infty} M^{-p/(2k+1)} a^{-pk/(2k+1)} n^{pk/(2k+1)} \inf_{\delta_n} \sup_{f \in \mathbf{F}(a,k,M)} E_f l_p(f, \delta_n) \geq \lim_{c\to\infty} \inf_{\delta} \sup_{h \in \mathbf{K}(c)} E_h^G l_p(h, \delta). \tag{22}$$

EXAMPLE 5. Let $\mathbf{F} = \{f\colon R \to R,\ f \geq 0,\ \int f = 1,\ f \text{ absolutely continuous}, \int f'^2 \leq 1\}$. Wahba (1975) found a variety of sequences of estimators, say $\{\delta_n\}$, satisfying

$$0 < \limsup_{n\to\infty} n^{1/2} \sup_{f \in \mathbf{F}} E_f(f(0) - \delta_n)^2 < \infty. \tag{23}$$

Furthermore, Wahba (1975) showed that for any $\varepsilon > 0$,

$$\limsup_{n\to\infty} n^{(1/2)+\varepsilon} \inf_{\delta_n} \sup_{f \in \mathbf{F}} E_f(f(0) - \delta_n)^2 > 0. \tag{24}$$

We will now use Theorem 2 to show that the best asymptotic rate of convergence for a sequence of estimators is $n^{1/2}$. In other words,

$$\liminf_{n\to\infty} n^{1/2} \inf_{\delta_n} \sup_{f \in \mathbf{F}} E_f(f(0) - \delta_n)^2 > 0. \tag{25}$$

First, take

$$f_0(x) = \begin{cases} \frac{17}{8} + x, & -\frac{17}{8} \leq x \leq -\frac{15}{8}, \\ \frac{1}{4}, & -\frac{15}{8} \leq x \leq \frac{15}{8}, \\ \frac{17}{8} - x, & \frac{15}{8} \leq x \leq \frac{17}{8}, \\ 0, & |x| \geq \frac{17}{8}. \end{cases}$$

Then $\int f_0(x)\,dx = 1$ and $\int f'^2(x)\,dx = 1/2$. Let

$$g(x) = \begin{cases} \frac{1}{8} - |x|, & |x| \leq \frac{1}{8}, \\ 0, & |x| > \frac{1}{8}. \end{cases}$$

Then $\int g'^2(x)\,dx = 1/4$. Let $K = \{\theta g\colon 0 \leq \theta \leq 1\}$ and $\alpha_n = n^{1/4}$, $\beta_n = n^{1/2}/4$. Then $\alpha_n^2 \beta_n = f_0(0) n$ and for large n,

$$f_n(\theta g; x) \in \mathbf{F},$$

where $f_n(h; x)$ is defined by (16). Moreover, $f_n(\theta g; 0) = f_0(0)(1 + (\theta g(0)/\alpha_n))(1 + o(1))$. Hence

$$\liminf_{n\to\infty} \left(\frac{n^{1/4}}{f_0(0)} \right)^2 \inf_{\delta_n} \sup_{f \in \mathbf{F}} E_f(f(0) - \delta_n)^2 \geq \inf_{\delta} \sup_{0 \leq \theta \leq 1} E_\theta^G(\theta g(0) - \delta)^2 > 0. \tag{26}$$

4. Proofs.

PROOF OF THEOREM 1. Lemma 1 immediately yields

$$E^n_{T_n f} l_n(T_n f, \delta_n(X(t))) = E^1_f l_n\left(T_n f, \delta_n\left(\frac{X(\beta_n t)}{\alpha_n \beta_n}\right)\right). \quad (27)$$

Now by (7), $h(\alpha_n, \beta_n)\delta(X(t)) = \delta_n(X(\beta_n t)/\alpha_n \beta_n)$ and it follows that (27) is equal to

$$E^1_f l_n(T_n f, h(\alpha_n, \beta_n)\delta(X(t))). \quad (28)$$

By Assumption A,

$$l_n(T_n f, h(\alpha_n, \beta_n)\delta(X(t))) = g(\alpha_n, \beta_n) l(f, \delta(X(t))).$$

Hence (28) is equal to

$$g(\alpha_n, \beta_n) E^1_f l(f, \delta) \quad [\text{which is the same as } (8)]. \quad (29)$$

This establishes (8). (9) follows immediately upon taking sup's. Likewise, (10) follows on taking inf's.

PROOF OF THEOREM 2. Let X_i, $i = 1, \dots, n$ be i.i.d. each with density f_0. Let

$$Q_n = \sum \left(\frac{h(\beta_n X_i)}{\alpha_n} - h_n\right), \quad (30)$$

$$R_n = \frac{1}{2} \sum \left(\frac{h(\beta_n X_i)}{\alpha_n} - h_n\right)^2. \quad (31)$$

Simple calculations show that

$$\lim_{n \to \infty} EQ_n = 0, \quad (32)$$

$$\lim_{n \to \infty} \operatorname{var} Q_n = \int h^2(y)\, dy, \quad (33)$$

$$\lim_{n \to \infty} ER_n = \tfrac{1}{2} \int h^2(y)\, dy, \quad (34)$$

$$\lim_{n \to \infty} \operatorname{var} R_n = 0. \quad (35)$$

Finally note that since $h_n = O((\alpha_n \beta_n)^{-1})$,

$$E\left|\frac{h(\beta_n X_i)}{\alpha_n} - h_n\right|^j = O_p\left(\frac{1}{\alpha_n^j \beta_n}\right) = o(n^{-1}) \quad \text{for } j \geq 3 \quad (36)$$

and so

$$\log \prod_1^n \frac{f_n(h; X_i)}{f_0(X_i)} - Q_n - R_n = o_p(1). \quad (37)$$

It then follows immediately from the asymptotic expansion given in (37) and the results in (32)–(35) that the experiments $\{P_h^n: h \in \mathbf{H}\}$ converge weakly to the standard Gaussian experiment $\{G_h: h \in \mathbf{H}\}$. □

Acknowledgments. This work is a combination of two previous papers. The author would like to thank the referees and the Associate Editor for all their comments on those original papers. This paper has also greatly benefited from many discussions that the author had with Lawrence D. Brown and David L. Donoho.

REFERENCES

BROWN, L. D. and LOW, M. G. (1990). Asymptotic equivalence of nonparametric regression and white noise. Technical report, Cornell Univ.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

DONOHO, D. L. and LOW, M. G. (1990a). Renormalization exponents and optimal pointwise rates of convergence. Technical report, Univ. California, Berkeley.

DONOHO, D. L. and LOW, M. G. (1990b). White noise approximation and minimax risk. Technical report, Univ. California, Berkeley.

FARRELL, R. H. (1972). On the best obtainable rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1984). On nonparametric estimation of values of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 19–32. (In Russian.)

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1991). Asymptotically normal families of distributions and effective estimation. *Ann. Statist.* **19** 1681–1724.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LOW, M. G. (1991). Renormalizing upper and lower bounds for integrated risk in the white noise model. Technical report, Univ. California, Berkeley.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233–252.

ROMANO, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16** 629–647.

STONE, C. J. (1980). Optimum rate of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

WAHBA, G. (1975). Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302