Asymptotic Equivalence of Nonparametric Regression and White Noise

Lawrence D. Brown
University of Pennsylvania

Mark G. Low
University of Pennsylvania

Asymptotic Equivalence of Nonparametric Regression and White Noise

Abstract

The principal result is that, under conditions, to any nonparametric regression problem there corresponds an asymptotically equivalent sequence of white noise with drift problems, and conversely. This asymptotic equivalence is in a global and uniform sense. Any normalized risk function attainable in one problem is asymptotically attainable in the other, with the difference in normalized risks converging to zero uniformly over the entire parameter space. The results are constructive. A recipe is provided for producing these asymptotically equivalent procedures. Some implications and generalizations of the principal result are also discussed.

Keywords

risk equivalence, local asymptotic minimaxity, linear estimators

Disciplines

Statistics and Probability

The Annals of Statistics
1996, Vol. 24, No. 6, 2384–2398

ASYMPTOTIC EQUIVALENCE OF NONPARAMETRIC REGRESSION AND WHITE NOISE

By Lawrence D. Brown[1] and Mark G. Low[2]

University of Pennsylvania

The principal result is that, under conditions, to any nonparametric regression problem there corresponds an asymptotically equivalent sequence of white noise with drift problems, and conversely. This asymptotic equivalence is in a global and uniform sense. Any normalized risk function attainable in one problem is asymptotically attainable in the other, with the difference in normalized risks converging to zero uniformly over the entire parameter space. The results are constructive. A recipe is provided for producing these asymptotically equivalent procedures. Some implications and generalizations of the principal result are also discussed.

Introduction. The principal result of this paper is that to any nonparametric regression problem there corresponds a white noise with drift problem which is asymptotically equivalent. The impact of this asymptotic equivalence is that any asymptotic solution to one of these problems will automatically yield a corresponding solution to the other. In addition, there is an explicit recipe for this correspondence. For example, the optimal rates of convergence will be equal as will suitably normalized local and global asymptotic risks, and knowledge of a minimax procedure or of a linear minimax procedure in one problem automatically yields the corresponding procedure in the other and so forth.

In particular, many classical functional estimation problems which have previously been treated separately fall into this framework. These problems include:

1. Estimating the whole function, considered in the white-noise model by Pinsker (1980) and in the regression context by Nussbaum (1985) and Speckman (1985). See also Donoho, Liu and MacGibbon (1990) and Golubev and Nussbaum (1990).
2. Estimating a point functional such as $f(x_0)$. For white noise, see Ibragimov and Hasminskii (1984) and for regression, see Ibragimov and Hasminskii (1982). See also Donoho and Liu [(1991), pages 677ff and 688] and Donoho and Low (1992).

Received January 1992; revised April 1993.
[1] Supported in part by NSF Grants DMS-91-07842, 93-10228, 95-96094.
[2] Supported in part by NSF Grant DMS-91-07894.
AMS 1991 *subject classifications*. Primary 62G07; secondary 62G20, 62M05.
Key words and phrases. Risk equivalence, local asymptotic minimaxity, linear estimators.

3. Estimating a nonlinear functional such as $\int f^2(x)\,dx$, treated for white noise in Fan (1991b) and in earlier references cited there and for both white noise and regression in Donoho and Nussbaum (1990).
4. Estimating the whole function based on indirect observations as in Fan (1991a).

The equivalence theory also covers many adaptive situations. If there is a particular sequence of estimators which is asymptotically minimax over a collection of parameter spaces in the white-noise case, then there is a corresponding sequence based on the regression model which is also minimax over each of these parameter spaces. Such a sequence of adaptive estimators for estimating the whole function has been found by Efroimovich and Pinsker (1984) in the white-noise case and by Golubev (1987) for nonparametric regression. See also Golubev (1991). For the problem of estimating a linear functional in both regression and white noise, see Lepskii (1991).

Hall and Johnstone (1992) discuss examples of estimating optimal, possibly random, bandwidths in the regression and white-noise contexts. These problems are also covered by the general equivalence theory developed in this paper. See Remark 4.5.

The equivalence theory can provide an additional technical advantage. Some proofs, for example those involving rates of convergence, may be much simpler in the white-noise model. Thus one may use the white-noise model to figure out the optimal rate via homogeneity and then the same rate holds in nonparametric regression. See, for example, Section 7 of Donoho and Low (1992). Analogous equivalence results should be valid for some other nonparametric problems. Indeed, Nussbaum (1993) has very recently proved one such result for nonparametric density estimation.

The first part of the paper contains necessary background. This includes descriptions of the nonparametric regression and white-noise problems, the definition of asymptotic equivalence and a discussion of some of its general consequences.

The second part of the paper contains the main equivalence theorems. Two cases are treated separately. In one case, the independent variables are deterministically fixed; in the other, they are a random sample from a specified distribution.

Part 1. Background.

1. Nonparametric regression. The nonparametric regression model to be treated in this paper is as follows: Let $I \subseteq \mathbb{R}$ be a possibly infinite interval. Let $f(\cdot)$: $I \to \mathbb{R}$ and $\sigma^2(\cdot)$: $I \to (0, \infty)$ be two measurable functions and let H: $\mathbb{R} \to [0,1]$ be an increasing c.d.f. The variables (X_{ni}, Y_{ni}), $i = 1, \dots, n$, are observed. In the *deterministic* X variant the independent variables are given by a deterministic scheme which will be given by

$$x_{ni} = H^{-1}(i/(n+1)), \qquad i = 1, \dots, n, \tag{1.1}$$

except where otherwise noted. In the *random* X variant the X_i are independent random variables.

$$X_{ni} \sim H \qquad \text{i.i.d. } i = 1, \dots, n. \tag{1.2}$$

In either case the conditional distribution of Y given X is described via

$$Y_{ni} = f(x_{ni}) + \sigma(x_{ni})\varepsilon_{ni}, \qquad \varepsilon_{ni} \sim N(0,1) \text{ ind., } i = 1, \dots, n. \tag{1.3}$$

The parameter space Θ consists of a possibly large set of choices of f. The c.d.f. H and the function $\sigma(\cdot)$ are assumed fixed and known prior to experimentation.

Here are some examples which are subject to later results of this paper.

EXAMPLE 1.1. Θ is given in terms of a Lipshitz condition as

$$\Theta^{(1)}_{\alpha,B} = \left\{ f: \left| f(x+\Delta) - f(x) - \sum_{i=1}^{[\alpha]-1} \left(f^{(i)}(x)/i!\right)\Delta^i \right| \right.$$
$$\left. \le B|\Delta|^\alpha \text{ and } \sup_{x\in I}|f(x)| \le B \right\} \quad \text{if } \alpha \ge 1, \tag{1.4}$$
$$\Theta^{(1)}_{\alpha,B} = \{ f: |f(x+\Delta) - f(x)| \le B|\Delta|^\alpha \} \quad \text{if } 0 < \alpha < 1.$$

[(1.4) holds for all x, $x + \Delta \in I$.] Later we shall require $\alpha > 1/2$, as explained in Remark 4.7.

EXAMPLE 1.2. Θ is given by a Sobolev type condition such as

$$\Theta^{(2)}_{\alpha,B} = \left\{ f: \int \left(f^{(\alpha)}(t)\right)^2 dt \le B \text{ and } \sup_{x\in I}|f(x)| \le B \right\}$$

for $\alpha = 1, \dots,$ where $f^{(\alpha)}$ denotes the αth derivative of f, assumed to exist in the sense that $f^{(\alpha-1)}$ is absolutely continuous. Note that when $\int_I dt = 1$, then $\Theta^{(1)}_{\alpha,B} \subset \Theta^{(2)}_{\alpha,B}$.

2. White noise. Assume with no loss of generality that $0 \in I$. Let $\{B_t: t \in I\}$ denote Brownian motion on I conditional on $B_0 = 0$, for example, $B_{\pm t} \sim N(0, t)$ and B_{t_1}, $t_1 > 0$, is independent of B_{t_2}, $t_2 < 0$. Fix $n = 1, \dots$. Let $\{Z^{(n)}_t\}$ denote the Gaussian process whose white-noise version is represented as $dZ^{(n)}_t = \mu(t)\,dt + \lambda(t)\,dB_t/\sqrt{n}$.

The parameter space in this model is, as before, the set of possible mean functions. Consequently, the statistical white-noise problem for given n is to observe the process $\{Z^{(n)}_t\}$ defined above for some $\mu \in \Theta$. In this way a sequence of problems is defined for $n = 1, 2, \dots,$ each of which has the same parameter space Θ.

Suppose $\{Y_t^{(n)}\colon t \in J\}$ and $\{Z_t^{(n)}\colon t \in J\}$ are two Gaussian processes, like those above, with respective mean functions $\mu(t)$ and $\nu(t)$ and with identical variance functions $\lambda^2(t)/n$. Let $g_Y^{(n)}$ and $g_Z^{(n)}$ denote their respective probability densities with respect to any dominating measure ξ. Let

$$L_1 = \int |g_Y^{(n)}(\omega) - g_Z^{(n)}(\omega)| \xi(d\omega). \tag{2.1}$$

[Where convenient we also use either the notation $L_1(Y_t^{(n)}, Z_t^{(n)})$ or $L_1(g_Y^{(n)}, g_Z^{(n)})$.] Standard calculations yield

$$L_1 = 2(1 - 2\Phi(-D/2)) \tag{2.2}$$

where

$$D^2 = \int \Big[n(\mu(t) - \nu(t))^2 / \lambda^2(t) \Big] \, dt.$$

Consequently, $L_1 = O(D)$.

REMARK 2.1. Let $\{Z_t^{(n)}\}$ be as above. Assume H is absolutely continuous with $dH/dt = h$ and assume $h > 0$ (a.e.) on I. Define the Gaussian process

$$V_\tau^{(n)} = Z_{H^{-1}(\tau)}^{(n)}, \qquad \tau \in [0,1].$$

Then $V_\tau^{(n)}$ has mean $\mu^*(\tau)$ and variance function $\lambda^{*2}(\tau)/n$ given by

$$\mu^*(\tau) = \frac{\mu(H^{-1}(\tau))}{h(H^{-1}(\tau))}, \qquad \lambda^{*2}(\tau) = \frac{\lambda^2(H^{-1}(\tau))}{h(H^{-1}(\tau))}. \tag{2.3}$$

Hence there is no significant loss of generality in assuming $I = [0,1]$ and H is uniform on I, although to do so will affect the definition of Θ, in the manner suggested by (2.3).

3. Statistical equivalence. Consider two statistical problems, $\mathscr{P}^{(1)}$ and $\mathscr{P}^{(2)}$, with sample spaces $\mathscr{X}^{(i)}$, $i = 1,2$ (and suitable σ-fields), respectively, but with the same parameter space Θ. Denote the respective families of distributions by $\{G_\Theta^{(i)}\colon \theta \in \Theta\}$. The following paragraph describes Le Cam's metric for the distance between two such experiments [see, e.g., Le Cam (1986) or Le Cam and Yang (1990)].

Let $\mathscr{A}$ be any (measurable) action space and let $L\colon \Theta \times \mathscr{A} \to [0,\infty)$ denote a loss function. Let $\|L\| = \sup\{L(\Theta, a)\colon \theta \in \Theta,\ a \in \mathscr{A}\}$. $\delta^{(i)}$ will be the generic symbol for a (randomized) decision procedure in the ith problem. The risk from using procedure $\delta^{(i)}$ when L is the loss function and θ is the true value of the parameter is denoted by $R^{(i)}(\delta^{(i)}, L, \theta)$. Le Cam's metric is

$$\begin{aligned} &\Delta(\mathscr{P}^{(1)}, \mathscr{P}^{(2)}) \\ &\quad = \max\Big[\inf_{\delta^{(1)}} \sup_{\delta^{(2)}} \sup_{\theta} \sup_{L:\,\|L\|=1} |R^{(1)}(\delta^{(1)}, L, \theta) - R^{(2)}(\delta^{(2)}, L, \theta)|, \\ &\qquad\qquad \inf_{\delta^{(2)}} \sup_{\delta^{(1)}} \sup_{\theta} \sup_{L:\,\|L\|=1} |R^{(1)}(\delta^{(1)}, L, \theta) - R^{(2)}(\delta^{(2)}, L, \theta)| \Big]. \end{aligned} \tag{3.1}$$

(Thus, if $\Delta < \varepsilon$, this means that for every procedure $\delta^{(i)}$ in problem i there is a procedure $\delta^{(j)}$ in problem j, $j \neq i$, with risk differing by at most ε, uniformly over all L such that $\|L\| = 1$ and $\theta \in \Theta$.) Two sequences of problems $\{\mathscr{P}_n^{(1)}: n = 1, \dots\}$ and $\{\mathscr{P}_n^{(2)}: n = 1, \dots\}$ are *asymptotically equivalent* if $\Delta(\mathscr{P}_n^{(1)}, \mathscr{P}_n^{(2)}) \to 0$ as $n \to \infty$. In this case, for any sequence of procedures $\delta_n^{(1)}$ in problems $\mathscr{P}_n^{(1)}$, $n = 1, \dots,$ there is a sequence $\delta_n^{(2)}$ in problems $\mathscr{P}_n^{(2)}$ for which

$$\lim_{n \to \infty} \sup_{\Theta} \sup_{L: \|L\| = 1} |R_n^{(1)}(\delta_n^{(1)}, L, \Theta) - R_n^{(2)}(\delta_n^{(2)} L, \Theta)| = 0.$$

Such sequences of procedures are said to be *asymptotically equivalent*.

In our proofs of statistical equivalence the key step is to arrange matters so that $\mathscr{X}^{(1)} = \mathscr{X}^{(2)}$. Then define

$$L_1(\mathscr{P}^{(1)}, \mathscr{P}^{(2)}) = \sup_{\theta \in \Theta} \int |g_\Theta^{(1)}(x) - g_\Theta^{(2)}(x)| \xi(dx), \tag{3.2}$$

where ξ dominates $G_\Theta^{(i)}$ and $g_\Theta^{(i)} = dG_\Theta^{(i)}/d\xi$, $i = 1, 2$. The following well-known fact can then be used to establish asymptotic equivalence.

THEOREM 3.1.

$$|R^{(1)}(\delta, L, \Theta) - R^{(2)}(\delta, L, \Theta)| \le L_1(\mathscr{P}^{(1)}, \mathscr{P}^{(2)}) \|L\|. \tag{3.3}$$

Consequently,

$$\Delta(\mathscr{P}^{(1)}, \mathscr{P}^{(2)}) \le L_1(\mathscr{P}^{(1)}, \mathscr{P}^{(2)}). \tag{3.4}$$

Also,

$$\Delta(\mathscr{P}_n^{(1)}, \mathscr{P}_n^{(2)}) \to 0 \text{ if } L_1(\mathscr{P}_n^{(1)}, \mathscr{P}_n^{(2)}) \to 0.$$

PROOF. (3.3) is just a restatement of the standard inequality

$$\left| \int h(x)(g^{(1)}(x) - g^{(2)}(x)) \xi(dx) \right| \le (\sup |h(x)|) \left(\int |g^{(1)}(x) - g^{(2)}(x)| \xi(dx) \right). \qquad \square$$

Two other techniques are also used repeatedly: one is reduction by sufficiency; the other involves relations between the L_1 norm and the Hellinger metric. The use of sufficiency is based on the following.

LEMMA 3.2. *Let $\mathscr{X}$ and its σ-field be a Polish space with its associated Borel field. Let $\mathscr{P}^{(1)}$ denote an experiment with sample space $\mathscr{X}$. Let $S: \mathscr{X} \to \mathscr{Y}$ be a sufficient statistic and let $\mathscr{P}^{(2)}$ denote the experiment in which $Y = S(X)$ is observed. Then $\Delta(\mathscr{P}^{(1)}, \mathscr{P}^{(2)}) = 0$. (A Polish space is one which is locally compact, metrizable and second countable. All the measurable spaces encountered in this paper possess these properties.)*

PROOF. The lemma follows from the fact that under the hypothesis there exists a measurable map $\delta(dx|y)$ such that $\int \delta(B|S(x)) G_\Theta(dx) = G_\Theta(B)$. See Le Cam (1986). $\square$

The Hellinger metric $H(G^{(1)}, G^{(2)})$ is defined by

$$H^2(G^{(1)}, G^{(2)}) = \int \left(g^{(1)}(x)^{1/2} - g^{(2)}(x)^{1/2} \right)^2 \xi(dx).$$

It provides a bound for L_1 since

$$L_1(G^{(1)}, G^{(2)}) \le 2H(G^{(1)}, G^{(2)}). \tag{3.5}$$

Another useful fact is that if the $G^{(i)}$ are product measures, $G^{(i)} = \prod_{j=1}^m G_j^{(i)}$, then

$$H^2(G^{(1)}, G^{(2)}) = 2\left[1 - \prod_{j=1}^m \left[1 - \frac{H^2\left(G_j^{(1)}, G_j^{(2)}\right)}{2}\right]\right] \tag{3.6}$$

[Le Cam (1986)]. Finally, direct calculation yields

$$\begin{aligned} &H^2\left(N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)\right) \\ &\quad = 2\left[1 - \left[\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}\right]^{1/2} \exp\left[-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)}\right]\right]. \end{aligned} \tag{3.7}$$

Part 2. Equivalence results.

4. Deterministic *X*. Let $I = [\alpha, \beta]$, $-\infty < \alpha < \beta < \infty$. Let $\sigma^2(\cdot) > 0$ be a given absolutely continuous function on I such that

$$\left| \frac{\partial}{\partial t} \ln \sigma(t) \right| \le C_1, \qquad t \in I. \tag{4.1}$$

for some $C_1 < \infty$. Suppose

$$\sup\{|f(t)|: t \in I, f \in \Theta\} = B < \infty \tag{4.2}$$

and also (4.5), below. Assume H is absolutely continuous on I and

$$H'(t) = h(t) > 0 \quad \text{a.e. on } I. \tag{4.3}$$

With x_{ni} as in (1.1) define the step function

$$\bar{f}_n(t) = \begin{cases} f(x_{ni}), & \xi_{i-1} \le t < \xi_i,\ i, \dots, n, \\ f(x_{nn}), & t = \beta, \end{cases} \tag{4.4}$$

where $\xi_i = H^{-1}(i/n)$. (The dependence of ξ_i on n is suppressed for convenience.) Assume

$$\lim_{n\to\infty} \sup_{f \in \Theta} n \int_\alpha^\beta \left(f(t) - \bar{f}_n(t) \right)^2 h(t)\, dt = 0. \tag{4.5}$$

REMARK 4.1. Equation (4.5) is a uniform smoothness condition on f. It is satisfied in a wide variety of examples. In particular it is satisfied in Examples 1.1 and 1.2 when $\alpha > 1/2$. (See Remark 4.7 for discussion of the case $\alpha \le 1/2$.)

For example, to verify (4.5) in Example 1.1 for the case $\alpha > 1/2$, $I = [0,1]$, H uniform and $\sigma(\cdot) = 1$ we note that $|f(t) - \bar{f}_n(t)| \le B|t - i/(n+1)|^{\alpha}$, where $(i-1)/n \le t < i/n$. Hence

$$n\int_0^1 \left(f(t) - \bar{f}_n(t)\right)^2 dt \le nB^2 \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} \left| t - \frac{i}{n+1} \right|^{2\alpha} dt$$
$$\le n^2 B^2 \int_0^{1/n} t^{2\alpha}\, dt = \frac{n^2 B^2}{2\alpha+1} \left(\frac{1}{n}\right)^{2\alpha+1} \to 0$$

since $2\alpha + 1 > 2$. Hence (4.5) is satisfied.

THEOREM 4.1. *Under assumptions* (4.1)–(4.5) *the deterministic X nonparametric regression model is asymptotically equivalent to the white-noise model* (2.1) *having* $\mu(t) = f(t)$, $\lambda^2(t) = \sigma^2(t)/h(t)$.

PROOF. Let $\bar{Z}_t^{(n)}$ be the white-noise model described by

$$d\bar{Z}_t^{(n)} = \bar{f}_n(t)\, dt + \frac{\lambda(t)}{\sqrt{n}}\, dB_t.$$

Note that $\sup\{\lambda(t)\colon t \in I\}/\inf\{(\lambda(t)\colon t \in I\} < \infty$ because of (4.1) and (4.3). Hence (4.5), (2.2) and Theorem 3.1 show that

$$\Delta\left(\{Z_t^{(n)}\}, \{\bar{Z}_t^{(n)}\}\right) \to 0.$$

Define

$$K_i^2 = \frac{n\sigma^2(x_{ni})}{\int_{\xi_{i-1}}^{\xi_i} dt/(\lambda^2(t))} \quad \text{and} \quad S_i^{(n)} = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)}. \tag{4.6}$$

The variables $\{S_i^{(n)}\colon i = 1,\dots,n\}$ are sufficient for $\{\bar{Z}_t^{(n)}\colon t \in I\}$. Hence

$$\Delta\left(\{\bar{Z}_t^{(n)}\}, \{S_i^{(n)}\}\right) = 0$$

by Lemma 3.2.

The variables $S_i^{(n)}$, $i = 1,\dots,n$, are independent with

$$V\left(S_i^{(n)}\right) = \frac{K_i^2}{n} \int_{\xi_{i-1}}^{\xi_i} \frac{dt}{\lambda^2(t)} = \sigma^2(x_{ni}),$$

$$\begin{aligned} E\left(S_i^{(n)}\right) &= K_i \int_{\xi_{i-1}}^{\xi_i} \frac{\bar{f}_n(t)}{\lambda^2(t)}\, dt \\ &= \sqrt{n}\,\sigma(x_{ni}) f(x_{ni}) \left[\int_{\xi_{i-1}}^{\xi_i} \frac{h(t)}{\sigma^2(t)}\, dt\right]^{1/2} \\ &= f(x_{ni}) \left[\frac{\sigma(x_{ni})}{\sigma(\bar{\xi}_{ni})}\right], \end{aligned} \tag{4.7}$$

where $\xi_{i-1} \le \bar{\xi}_{ni} \le \xi_i$. The existence of $\bar{\xi}_{ni}$ follows from the mean value theorem since $\int_{\xi_{i-1}}^{\xi_i} h(t)\,dt = 1/n$. Assumption (4.1) then yields

$$E(S_i^{(n)}) = f(x_{ni})(1 + O(\xi_i - \xi_{i-1}))$$

uniformly in $f \in \Theta$, i and n. [Note also that the $O(\cdot)$ term is zero if $\sigma^2(t) = \gamma^2$, a constant.]

The above, together with (3.6) and (3.7), yields

$$\begin{aligned} &H^2(\{S_i^{(n)}\}, \{x_{ni}, Y_{ni}\}) \\ &\quad = O\Big(\sum f^2(x_{ni})(\xi_i - \xi_{i-1})^2\Big) \\ &\quad \le B^2(b-a)O(\sup\{(\xi_i - \xi_{i-1}) : i = 1, \dots, n\}) \\ &\quad = o(1) \quad (\text{uniformly over } f \in \Theta) \end{aligned} \tag{4.8}$$

by (4.2) and (4.3). Hence $\lim_{n\to\infty} \Delta(\{S_i^{(n)}\}, \{x_{ni}, Y_{ni}\}) = 0$ by (3.5) and Theorem 1. It follows from the preceding facts that $\lim_{n\to\infty} \Delta(\{Z_t^{(n)}\}, \{x_{ni}, Y_{ni}\}) = 0$. □

TECHNICAL NOTE. The conditions (4.1)–(4.3) are clearly stronger than necessary. They are used in order to yield the conclusion $H^2 = o(1)$ (uniformly over $f \in \Theta$) which appears in (4.8). For example, if $\sigma^2(\cdot) = \gamma^2$, a constant, $I = [0, 1]$ and $h(t) = 1$ on $[0, 1]$, then (4.3) holds and we may also drop conditions (4.1) and (4.2).

The above theorem yields a prescription for producing, from a sequence of procedures in one problem, an asymptotically equivalent sequence in the other. The following corollary gives a precise recipe. The corollary applies to either nonrandomized or randomized procedures.

COROLLARY 4.1. *Let $\{\delta_n\}$ be a sequence of procedures in the regression model of Theorem* 4.1. *Define γ_n in the corresponding white-noise problem by*

$$\gamma_n(Z_t^{(n)}) = \delta_n(S_i^{(n)}) \qquad \textit{where } S_i^{(n)} = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{dZ_t^{(n)}}{\lambda^2(t)} \tag{4.9}$$

as in (4.6). *Then $\{\gamma_n\}$ is asymptotically equivalent to $\{\delta_n\}$.*

Conversely, suppose $\{\gamma_n\}$ is a given sequence of procedures in the white-noise problem. Then $\{\delta_n\}$ is an asymptotically equivalent sequence in the asymptotically equivalent regression problem, where δ_n is the randomized procedure described for any measurable $A \in \mathscr{A}_n$ as follows:

$$\delta_n(A|\{s_i^{(n)}\}) = E\left(\gamma_n(A|\{\bar{Z}_t^{(n)}\}) \Big| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)}\,dt = s_i^{(n)},\ i = 1, \dots, n\right). \tag{4.10}$$

PROOF. The corollary is implicit in the proof of Theorem 5.1. Note that δ_n in (4.10) is well defined since it is independent of the drift parameter $\bar{f}^{(n)}(t)$ of the $\{\bar{Z}_t^{(n)}\}$ process because for each n, $\{S_i^{(n)}\}$ is sufficient for $\{\bar{Z}_t^{(n)}\}$. □

The construction (4.10) is not entirely felicitous for two reasons: (1) δ_n will, in general, be randomized even when γ_n is not; (2) the conditional expectation may be hard to evaluate. The following two remarks show that these difficulties can sometimes be partially alleviated.

REMARK 4.2. Jensen's inequality can sometimes be used to improve (4.10). Here is a precise statement. Suppose each $\mathscr{A}_n$ is a closed convex subset of a separable Banach space, and suppose each $L_n(\theta, \cdot)$ is a convex lower semicontinuous function satisfying $\lim_{\|a\| \to \infty} L_n(\theta, a) = \infty$. Define δ_n' as the nonrandomized procedure taking the values

$$\delta_n'\big(\{s_i^{(n)}\}\big) = \int a\, \delta_n\big(da | \{s_i^{(n)}\}\big). \tag{4.11}$$

Then $\{\delta_n'\}$ is asymptotically at least as good as $\{\delta_n\}$.

Even when L_n is not convex it may be that

$$\sup_\theta |R(\delta_n', L_n \wedge B, \theta) - R(\delta_n, L_n \wedge B, \theta)| \to 0. \tag{4.12}$$

in which case $\{\delta_n'\}$ and $\{\delta_n\}$ are asymptotically equivalent sequences for $\{L_n\}$. See Remark 4.3 for one such situation.

A (real-valued) linear estimator in the white-noise problem is a nonrandomized estimator which can be expressed in the form

$$\begin{gathered}\gamma_n\big(Z_t^{(n)}\big) = \int \rho_n(t)\, dZ_t^{(n)}, \text{ with} \\ \int \rho_n^2(t)\lambda^2(t)\, dt < \infty.\end{gathered} \tag{4.13}$$

It follows that $\gamma_n(Z_t^{(n)})$ has a normal distribution with mean $\int \rho_n(t)\mu(t)\, dt$ and variance $\int \rho_n^2(t)\lambda^2(t)\, dt/n$.

A linear estimator in the regression problem is, similarly, one which can be expressed in the form

$$\delta_n\big(s_i^{(n)}\big) = \frac{1}{n}\sum_{i=1}^{n} \rho_{ni} s_i^{(n)}. \tag{4.14}$$

As above, such estimators are normally distributed.

REMARK 4.3. The equivalence transformation is especially convenient for linear estimators. In order to demonstrate this, it is easier to write (4.14) in an alternate form. As before, for convenience, specialize to the case $I = [0, 1]$, H is uniform, and also assume $\lambda^2(\cdot) = \sigma^2(\cdot) = \gamma^2$. Let $e_n(t)$ denote the normalized linearized cumulative sums of $\{s_i^{(n)}\}$; that is,

$$e_n(t) = \frac{1}{n}\left\{\sum_{k=1}^{n-1} s_i^{(n)} + \frac{t - \xi_{i-1}}{\xi_i - \xi_{i-1}}\big(s_m^{(n)} - s_{m-1}^{(n)}\big)\right\},$$
$$\xi_{i-1} \le t < \xi_i,\ i = 1, \dots, n.$$

Then the general linear estimator (4.14) can be rewritten in the form

$$\delta_n\big(s_i^{(n)}\big) = \int_0^1 \tilde{\rho}_n(t)\, de_n(t). \tag{4.15}$$

Note that

$$e_n(t) = E\left[\bar{Z}_t^{(n)} \Big| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)}\, dt = s_i^{(n)},\ i = 1,\dots,n\right].$$

[Here $\xi_i = i/n$, $\lambda^2(t) = \gamma^2$ and $K_i = n\gamma$.] Hence, suppose a linear estimator (4.13) is given in the white-noise problem. Then the estimator δ_n' of (4.11) is given by (4.15) with $\tilde{\rho}_n(t) = \rho(t)$.

The preceding results apply to regression problems on a bounded interval. They can be extended without much additional complication to also apply on unbounded intervals if a somewhat different scheme for determining the X values is introduced, as follows:

Let $I = (a, b)$ and $I_n = [\alpha_n, \beta_n]$, $-\infty \le a < \alpha_n < \beta_n < b \le \infty$ and let H_n be an increasing c.d.f. on $[\alpha_n, \beta_n]$. Choose the independent variables according to

$$x_{ni} = H_n^{-1}(i/(n+1)). \tag{4.16}$$

The constants ξ_i (which also depend on n) are defined as before, but with H_n in place of H, and $\bar{f}_n$ is defined by (4.4) with the additional convention

$$\bar{f}_n(t) = 0, \qquad t < \alpha_n,\ t > \beta_n.$$

The assumption (4.5) is then replaced by

$$\lim_{n\to\infty} \sup_{f\in\Theta} n \int_{\alpha_n}^{\beta_n} \Big[h_n(t)\big(f(t) - \bar{f}_n(t)\big)^2 / \sigma^2(t) \Big]\, dt = 0. \tag{4.17}$$

Here is a formal statement of the equivalence result. Its proof involves only minor modifications of the proof of Theorem 4.1, and will be omitted.

COROLLARY 4.2. *Define* $\{x_{ni}\}$ *as above. Assume* (4.1)–(4.2) *are satisfied on* $(-\infty, \infty)$ *and* H_n *satisfies* (4.3). *Assume* (4.17) *is satisfied and either* $\sigma^2(\cdot) = \gamma^2$, *a constant, or* $(\beta_n - \alpha_n)\max_{1\le i\le n}(\xi_i - \xi_{i-1}) = o(1)$. [*Here* $\xi_i = H_n^{-1}(1/n)$, $\xi_0 = \alpha_n$ *and* $\xi_n = \beta_n$. *As usual, the dependence of* ξ *on* n *is suppressed in the notation.*] *Then the deterministic X nonparametric regression model is asymptotically equivalent to the white-noise model* (2.1) *having* $\mu(t) = f(t)$ *and incremental variance function depending on* n *and given by* $\lambda_n^2(t) = \sigma^2(t)/h_n(t)$.

REMARK 4.4. There is an additional question which arises in the preceding situation: Suppose $-\infty < a < b < \infty$ and $H_n \to H$ with $H' = h$. Let $\{Z_t^{(n)}\}$ be the white noise with drift $f(t) = \mu(t)$ and with variance function $\lambda^2(t) = h(t)/\sigma^2(t)$. Is it true that $\Delta(\{x_{ni}, Y_{ni}\}, \{Z_t^{(n)}\}) \to 0$?

An affirmative answer to the preceding question appears to require some condition on the convergence of H_n to H. In some examples it is not hard to supply such a condition. Consider, for example, the situation in Example 1.1 with $\alpha \geq 1$. Assume (for simplicity of exposition) that $\sigma^2(t) = \gamma^2$. Make the other assumptions in Theorem 4.1 and assume also that

$$\sup_t |H_n(t) - H(t)| = \Delta_n = o(1/\sqrt{n}) \tag{4.18}$$

as $n \to \infty$. Then the answer is affirmative.

To establish the preceding claim, let $\{x_{ni}^*, Y_{ni}^*\}$ denote the deterministic X nonparametric regression based on H. Then

$$\begin{aligned}\Delta(\{x_{ni}, Y_{ni}\}, \{x_{ni}^*, Y_{ni}^*\}) &= \gamma^2 \sum \left[f\left(H_n^{-1}\left(\frac{i}{n+1}\right)\right) - f\left(H^{-1}\left(\frac{i}{n+1}\right)\right)\right]^2 \\ &\leq \gamma^2 n \Delta_n^2 B^2 \to 0.\end{aligned}$$

Then apply Theorem 4.1 to see that $\Delta(\{x_{ni}^*, Y_{ni}^*\}, \{Z_t^{(n)}\}) \to 0$.

REMARK 4.5. The preceding methodology can be used when the loss functions also depend on the observations. Thus, suppose the loss in the regression problem is $L_n^{(1)}: \Theta \times \mathscr{A} \times \{(x_{ni}, y_{ni})\} \to [0, B]$ and in the white-noise problem it is $L_n^{(2)}: \Theta \times \mathscr{A} \times \{Z_t^{(n)}\} \to [0, B]$. For simplicity assume $\sigma^2(\cdot) = \gamma^2$, a constant. Define $\{T_i^{(n)}\}$ from $\{Z_t(n)\}$ in the same manner as $\{S_i^{(n)}\}$ was defined from $\{\overline{Z}_t^{(n)}\}$ [see (4.6)]. Assume the two loss functions asymptotically agree in the sense that for each $f \in \Theta$,

$$E_f\left(|L_n^{(2)}(f, a, Z_t) - L_n^{(1)}(f, a, \{x_{ni}, T_i^{(n)}\})|\right) \to 0 \quad \text{uniformly in } a \in \mathscr{A}. \tag{4.19}$$

The proof of Theorem 4.1 can then easily be adapted to prove that corresponding procedures in the two problems have asymptotically equal risk functions under the respective losses $L_n^{(1)}$ and $L_n^{(2)}$.

The preceding observation can be used to prove equivalence in the problem of optimal bandwidth selection as formulated in Hall and Johnstone (1992). The condition (4.19) is relatively straightforward to check in their context. The resulting conclusion is that their asymptotic result, once proved in the white-noise setting, is then also valid in the nonparametric regression setting.

REMARK 4.6. Looking at Example 1.1 in the case where $\alpha \leq 1/2$ shows what can happen when the key regularity condition (4.5) fails. Let $I = [0, 1]$, $h(\cdot) = 1$ and $\sigma(\cdot) = 1$. In the white-noise problem

$$\sqrt{n}\left[Z_1^{(n)} - \int_0^1 \mu(t)\, dt\right] \to N(0, 1)$$

in distribution uniformly over $\mu \in \Theta^{(1)}_{(1/2,1)}$. [In fact the left side is exactly $N(0,1)$.] However, in the regression problem of Section 1 there cannot exist an estimator $d_n^{(2)}$ such that

$$\sqrt{n}\left[d_n^{(2)}(x_{ni}, Y_{ni}) - \int_0^1 f(t)\,dt\right] \to N(0,1) \tag{4.20}$$

uniformly over $f \in \Theta^{(1)}_{(1/2,1)}$. By virtue of Remark 3.1, this shows that the two problems are not asymptotically equivalent. [To verify the impossibility of (4.20), let $\varepsilon_n = \sqrt{|t - i(t)/(n+1)|}$, where $i(t)$ is defined by $|t - i(t)/(n+1)| = \min\{|t - j/(n+1)|: j = 1,\dots,n\}$. Then $f(t) = 0$ and $f(t) = \varepsilon_n(t)$ are both in $\Theta^{(1)}_{(1/2,1)}$, and (x_{ni}, Y_{ni}) has the identical distribution in both cases. However, $\sqrt{n}\int_0^1 \varepsilon_n(t)\,dt \sim \sqrt{2}/3$. Hence validity of (4.20) for $f(t) = 0$ implies its failure for $f(t) = \varepsilon_n(t)$.]

Although the two sequences of problems are not asymptotically equivalent in the strong sense of Theorem 4.1, nevertheless in many special cases they are asymptotically equally useful. For example, they will have the same local asymptotic minimax risks for estimating $f(x_0)$ under squared error loss or for estimating $f(t)$ under integrated squared error loss.

5. Random X.

Results analogous to those of the preceding section can be derived for the random X nonparametric regression problem, as defined in (1.2). In this case the nonparametric regression problem is asymptotically equivalent to a white-noise model in which the drift depends on the observed values of X, as well as the unknown parameters.

To describe this white-noise model on $I = [\alpha, \beta]$, $-\infty < \alpha < \beta < \infty$, let $x_{(1)}, \dots, x_{(n)}$ denote the ordered values of X and let $x_{(0)} = \alpha$, $x_{(n+1)} = \beta$ and let

$$\tilde{H}_n(t) = \frac{i-1}{n+1} + \frac{t - x_{(i-1)}}{(n+1)(x_{(i)} - x_{(i-1)})} \quad \text{if } x_{(i-1)} \le t \le x_{(i)},\ i = 1,\dots,n+1. \tag{5.1}$$

Now define $\tilde{\xi}_i = \tilde{H}_n^{-1}(i/n)$ and define $\tilde{f}_n$ by (4.4) with $\tilde{\xi}_i$ in place of ξ_i. Let $\tilde{h}_n(t)$ denote the left-hand derivative of $\tilde{H}_n$ at t. In place of (4.5) assume

$$\lim_{n\to\infty} \operatorname{prob} \sup_{f\in\Theta} n\int_\alpha^\beta \bigl(f(t) - \tilde{f}_n(t)\bigr)^2 \bigl(\tilde{h}_n(t)/\sigma^2(t)\bigr)\,dt = 0. \tag{5.2}$$

REMARK 5.1. Assumption (5.2) is usually not much harder to verify than is (4.5). For example, if $I = [0,1]$, H is the uniform distribution, $\sigma(t) = 1$ and $\alpha > 1/2$ in Example 1.1, then (5.2) follows since

$$\int \bigl(f(t) - \tilde{f}_n(t)\bigr)^2 \tilde{h}_n(t)\,dt = \int \Bigl(f\bigl(\tilde{H}_n^{-1}(v)\bigr) - \tilde{f}_n\bigl(\tilde{H}_n^{-1}(v)\bigr)\Bigr)^2\,dv$$

and

$$E\Big\{n\Big(f\big(\tilde{H}_n^{-1}(v)\big) - \tilde{f}_n\big(\tilde{H}_n^{-1}(v)\big)\Big)^2\Big\} \le nB^2E(x_{(j)} - x_{(j-1)})^{2\alpha} \to 0$$

uniformly in v, n, where $j = \min\{i\colon x_{(i)} > \tilde{H}_n^{-1}(v)\}$.

In the preceding situation define $\tilde{\lambda}_n^2(t) = \sigma^2(t)/\tilde{h}_n(t)$ and

$$dZ_t^{(X,n)} = f_n(t)\,dt + \frac{\tilde{\lambda}_n(t)}{\sqrt{n}}\,dB_t. \tag{5.3}$$

Note that the distribution of $\{Z_t^{(X,n)}\}$ depends on the ancillary observations X through the local variance function $\tilde{\lambda}_n^2$.

THEOREM 5.1. *Assume* σ, f, H *satisfy* (4.1)–(4.3) *and* (5.2). *Then the random X nonparametric regression model is asymptotically equivalent to the white-noise model* (5.3).

PROOF. Analogously to Theorem 4.1, let

$$d\tilde{Z}_t^{(X,n)} = \tilde{f}_n(t)\,dt + \frac{\tilde{\lambda}_n(t)}{\sqrt{n}}\,dB_t.$$

The remainder of the proof exactly follows the pattern of proof of Theorem 4.1 with $\tilde{Z}$ in place of Z and $\tilde{f}$ in place of f. Also (4.8) needs to be modified slightly to begin as

$$H^2\big(\{X_{ni}, S_i^{(n)}\}, \{X_{ni}, Y_{ni}\}\big) = O\Big(E\Big(\sum f^2(x_{ni})(\xi_i - \xi_{i-1})^2\Big)\Big) \tag{5.4}$$

and so forth, since $\{x_{ni}\}$ and $\{\xi_i\}$ are random. (An alternative proof could be based directly on Corollary 4.2.) □

REMARK 5.2. A conspicuous feature of the preceding result is that knowledge of H is not required to construct $Z_t^{(X,n)}$ nor to carry through the proof of the theorem. Thus, suppose it is only assumed that $H \in \mathscr{H}$, where

$$\lim_{n\to\infty} \operatorname{prob} \sup_{H\in\mathscr{H}} \sup_{f\in\Theta} n\int_\alpha^\beta \big(f(t) - \tilde{f}_n(t)\big)^2 \big(\tilde{h}_n(t)/\sigma^2(t)\big)\,dt = 0$$

and also that either $\sigma^2(t) = \gamma^2$ or (4.3) is modified to require the existence of an h_0 satisfying

$$\inf\{h(t)\colon h \in \mathscr{H}\} \ge h_0(t) > 0 \quad \text{a.e. on } I.$$

Then the equivalence assertion of Theorem 5.1 remains valid.

Note also that the construction in Corollary 4.2 of equivalent procedures together with the contents of Remarks 4.2–4.4 can easily be carried over to the current situation.

REMARK 5.3. Suppose H is known. An issue which then arises is whether this knowledge can serve in place of observation of the ancillary statistics $X_1, \ldots, X_n$. More precisely, let $X_{j_1} \le \cdots \le X_{j_n}$ denote the ordered values of $\{X_i: i = 1, \ldots, n\}$ and let $Y_{j_1}, \ldots, Y_{j_n}$ denote the corresponding values of $\{Y_i\}$. Then given knowledge of H, is the experiment $\{Y_{j_i}: i = 1, \ldots, n\}$ asymptotically equivalent to $\{(X_i, Y_i): i = 1, \ldots, n\}$? Because of Theorem 5.1, this is equivalent to asking whether the experiment $\{\tilde{Z}_t^{(X,n)}: t \in I\}$ is asymptotically equivalent to $\{(\tilde{H}_n, \tilde{Z}_t^{(X,n)}): t \in I\}$.

The general answer to this question is "No." Suppose $I = [0,1]$, $\sigma^2(t) = 1$ for $t \in I$ and $\Theta = \{f: f(t) = ct,\ c \in \mathbb{R}\}$. Then in the experiment $\{\tilde{Z}_t^{(X,n)}\}$ the UMVU and minimax estimator is $2\int t\, d\tilde{Z}_t^{(X,n)}$, which has variance

$$4n^{-1}\int t\tilde{\lambda}_n^2(t)\, dt + \operatorname{Var}\left(2\int t\, d\tilde{H}_n(t)\right) = 2n^{-1}(1 + 1/3). \tag{5.5}$$

Meanwhile, for the experiment $\{(\tilde{H}_n, \tilde{Z}_t^{X,n})\}$ the UMVU and minimax estimator is $2\int \tilde{H}_n(t)\, d\tilde{Z}_t^{(X,n)}$, which has variance $2n^{-1}$. This shows the two experiments are not asymptotically equivalent.

On the other hand, for many purposes the two experiments are equally useful asymptotically. Roughly speaking this happens when the rate of convergence of the estimator used is slower than the classical $\sqrt{n}$ rate. Heuristically, what happens in such problems is that the imprecision introduced by not observing $\tilde{H}_n$ is represented by a term analogous to the second one on the left of (5.5), which is of order $1/n$. When the estimator converges at slower than $\sqrt{n}$ rate, so that its variance converges more slowly than n^{-1}, this second term is asymptotically negligible. Such behavior can be deduced in examples like those cited in Donoho and Liu (1991) and Low (1992).

Acknowledgment. The authors thank Lucien Le Cam for helpful conversations.

REFERENCE

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

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk for hyperrectangles. *Ann. Statist.* **18** 1416–1437.

DONOHO, D. L. and LOW, M. C. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970.

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323.

EFROIMOVICH, S. Y. and PINSKER, M. S. (1984). A learning algorithm for nonparametric filtering. *Automat. i Telemekh.* **11** 58–65. (In Russian.)

FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

FAN, J. (1991b). On the estimation of quadratic functions. *Ann. Statist.* **19** 1273–1294.

GOLUBEV, G. K. (1987). Adaptive asymptotically minimax estimates of smooth signals. *Problemy Peredachi Informatsii* **23** 57–67. (In Russian.)

GOLUBEV, G. K. (1991). LAN in nonparametric functional estimation problems and lower bounds for quadratic risk. *Teoriya Veroyatnost. Primenen.* **36** 143–149. (In Russian.)

GOLBEV, K. and NUSSBAUM, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18** 758–778.

HALL, P. and JOHNSTONE, I. M. (1992). Empirical functionals and efficient smoothing parameter selection. *J. Roy. Statist. Soc. Ser. B* **54** 475–530.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1982). Bounds for the risk of nonparametric regression estimates. *Theory Probab. Appl.* **27** 84–99. (English translation.)

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 18–32. (English translation.)

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

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York.

LEPSKII, O. V. (1991). Asymptotically minimax adaptive estimation I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LOW, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545–554.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

NUSSBAUM, M. (1993). Asymptotic equivalence of density estimation and white noise. *Ann. Statist.* **24** 2399–2430.

PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 120–133. (English translation.)

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
3620 LOCUST WALK
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: lbrown@compstat.wharton.upenn.edu
lowm@compstat.wharton.upenn.edu