Direct asymptotic equivalence of nonparametric regression and the infinite dimensional location problem

Lawrence D. Brown *

Linda H. Zhao †

Department of Statistics

University of Pennsylvania

Abstract

We begin with a random design nonparametric regression having random predictors and Gaussian errors. We produce a convenient, easily implementable mapping of this problem into a Gaussian infinite dimensional location problem. Such an infinite dimensional problem can reflect a Fourier, or wavelet, or other orthogonal basis representation of the original regression situation. In this way it may be easier to analyze than the original regression formulation. There is considerable literature on doing this; beyond describing the situation we do not pursue here this issue of the analysis of such infinite dimensional models. For most of our results the random regressors in our theory may have either a known or unknown distribution.

The correspondence we produce between the regression and location problems is an asymptotic equivalence mapping. (We also explicitly describe the converse mapping from the location problem to the regression.) Thus any solution to a statistical problem in one formulation can be easily converted to a solution for the other formulation.

The basic mapping from the regression to location formulations involves a few steps. First, bin the regression observations and use the bin averages to compute an empirical infinite series transform. Then truncate this series appropriately. Add a small amount of prescribed Gaussian noise to the truncated series coefficients. Then use a subset of these to linearly predict the remaining tail coordinates of the infinite series. In many applications the latter two steps are not necessary even though they are needed for an explicit asymptotic equivalence mapping.

Key words: Nonparametric regression; Random design; Fourier series; Wavelet; Equivalence;

*Research supported in part by NSF grant DMS-9971751.

†Research supported in part by NSF grant DMS-9971848.

1 Introduction

Over the preceding decades nonparametric regression has become a standard, frequently employed statistical model. Early results for this model and its close relation, nonparametric density estimation, appear in Rosenblatt (1956), Parzen (1962), Nadaraya (1964) and Watson (1964). Within the last two decades a variety of inferential methods have been developed and implemented for this model. Various important extensions and generalizations have also been formulated. Monographs such as Härdle (1990), Wahba (1990), Fan and Gijbels (1996), Loader (1999) and Efromovich (1999) contain extensive treatments.

Many methods for these models are based on orthogonal series expansions or are closely related to such expansions. Examples include methods based on damped or truncated Fourier series (See, e.g., Efromovich and Pinsker (1984), Eubank and Speckman (1991).), smoothing splines (Wahba (1990)), and wavelets (Donoho and Johnstone (1995, 1998)). In many cases these methods rely on an asymptotic idealization of the problem in terms of an orthogonal series expansion. A recent monograph by Johnstone (2000) develops the theory given such an expansion.

The orthogonal series models are mathematically natural objects. They can often be easily manipulated to produce desirable nonparametric inferential procedures such as asymptotically minimax estimators and adaptive versions of such procedures. See Donoho (1994) for some general theory, and the references cited earlier for some specifics.

These infinite dimensional models have intuitive appeal and manipulative convenience. Hence it is valuable to know when nonparametric regression problems can automatically be reduced to such infinite dimensional location problems, and to provide a simple algorithm for doing so. Such an algorithm and the accompanying results are the focus of this paper.

The standard nonparametric regression model can be formulated in either a fixed design or a random design version. The difference relates to whether the predictor variables (X_i) are nonrandom or are independent random variables. See for example Antoniadis, Gregorie and McKeague (1994) or Brown and Low (1996). The equivalence formulation in Brown and Low (1996) is satisfactory for the fixed-design case, especially when the predictor variables are evenly spaced. As suggested there it is less satisfactory for the random-design case. In the present paper we concentrate only on the random-design model.

The goal is to describe the algorithm and then establish asymptotic equivalence in the strongest possible statistical sense between the nonparametric regression and its accompanying infinite dimensional location problem. This strong statistical sense is a global version of the metric for statistical experiments developed by Le Cam (1953, 1963, 1986). In our applications this is a consequence of uniform convergence in the L_1 metric of appropriate probability densities. Hence here it can be directly described and validated.

This type of asymptotic convergence is much stronger than other familiar notions such as convergence in distribution or convergence in asymptotic risk for a specific loss function. These weaker notions suffice for some but not all statistical applications. The more demanding type of convergence we pursue naturally requires somewhat more restrictive conditions and more care in its validation.

The statistician begins with a nonparametric regression problem involving observation of (X_i, Y_i), $i = 1, \ldots, n$, where Y_i are conditionally Gaussian with $E(Y_i|X_i) = f(X_i)$. Assume $X \in [0, 1]$, partly for convenience. Some sort of inference concerning f is desired, e.g. an estimate, test, or confidence band. At the same time it is known how to produce a satisfactory inferential solution to the related infinite dimensional location problem. This state of affairs often comes about because, as already noted, such infinite dimensional location problems are often easier to analyze than their regression counterparts.

The infinite dimensional location problem involves an index n and observation of Gaussian variables $Z_{n1}, Z_{n2}, \ldots$. Typically, $Z_{ni} \sim N(\theta_i, 1/n)$, independent, but as we shall see other Gaussian distributions may sometimes arise. The $\{\theta_i : i = 1, \ldots\}$ are related to f by

$$\theta_i = \int_0^1 f(t)\varphi_i(t)\, dt \tag{1}$$

where the $\{\varphi_i\}$ are the Fourier or other basis elements of the Hilbert space. Let $\delta_n(\mathbf{Z}_n)$ denote the satisfactory inferential solution given $\mathbf{Z}_n = \{Z_{ni} : i = 1, \ldots\}$. The algorithm described below in (i) - (iii) produces an empirical series $\widetilde{\mathbf{Z}}_n = \{\widetilde{Z}_{ni} : i = 1, \ldots\}$ as a function of the regression observations $\{X_i, Y_i\}$. The equivalence guarantees that for $f \in \mathscr{F}$ and any sequence of measurable subsets, $\{D_n\}$, of the inference space

$$|P_f(\delta_n(\mathbf{Z}_n) \in D_n) - P_f(\delta_n(\widetilde{\mathbf{Z}}_n) \in D_n)| \to 0 \tag{2}$$

uniformly for $f \in \mathscr{F}$. Consequently $\delta_n(\widetilde{\mathbf{Z}}_n)$ will asymptotically have the same desirable asymptotic properties for the nonparametric regression as did $\delta_n(\mathbf{Z}_n)$ in the series problems.

We propose a simple empirical series re-expression of the nonparametric regression problem in Section 4.1. This can be described in three steps:

(i) Construct a conditionally unbiased empirical version of the first $m(n)$ orthogonal series coefficients. Here $m(n) = o(n^{1-\epsilon})$ as in (19). The formula for this is (20), (21).

(ii) Add a very small amount of Gaussian noise. This is described in (35). As explained in remarks following Theorem 5.1 this step should be unnecessary in most application of theory even though it seems to be needed for formal validity of the theorems in which it appears.

(iii) Replace the remaining coefficients by non-informative normal variables having the appropriate covariance structure. This is explained in (26) for the case where X are uniform and in Section 5.3 and formula (100) for the more general case.

The converse asymptotic equivalence going from the series to the regression problem is described in Section 5.3.

An alternative to (iii) for the general case is to use procedures that depend only on the first $m(n)$ coordinates of the series expression. It follows as a consequence of Theorem 5.2 that no asymptotically useful information is lost because of such a restriction.

All these steps need to be done in the appropriate way. In particular, the counterexample in Section 3.1 shows it is not suitable to replace step (i) by using the most intuitive marginally unbiased, empirical orthogonal series coefficients.

In order to derive concrete results of the type we desire for nonparametric regression it is necessary to make some minimal assumptions about the space of allowable regression functions. For this purpose we introduce in Section 2.5 the familiar Hilbert space norm of functions of smoothness α. We assume the set $\mathscr{F}$ of allowable regression functions satisfies $\mathscr{F} \subset S_\alpha(B)$ where $S_\alpha(B)$ is a bounded ball in such a space.

In particular, space $S_\alpha(B)$ can be generated from Fourier series (or their close relatives) or from wavelet bases, yielding Besov-Hilbert spaces. For the case of Fourier type bases we can prove asymptotic equivalence when $\alpha > 1/2$. For the Besov setting we require $\alpha > 1$ for full equivalence; but see (3) below for description of an alternative formulation.

In the case of Besov bases we cannot prove (2) to hold for our empirical series when $1/2 < \alpha \leqslant 1$. However we can prove an only marginally weaker property that we feel should suffice for nearly all practical applications. We describe this property in Section 6 where we call it asymptotic relative sufficiency. It involves an index $n^*(n)$ with $n^*(n) < n$ but $n^*(n)/n \to 1$. Then

$$|P_f(\delta_{n^*}(\mathbf{Z}_{n^*}) \in D_n) - P_f(\delta_{n^*}(\widetilde{\mathbf{Z}}_n^o) \in D_n)| \to 0. \tag{3}$$

Here $\widetilde{\mathbf{Z}}_n^o = \widetilde{\mathbf{Z}}_n + \mathbf{V}_n^*$ where $\mathbf{V}_n^*$ is an independent mean 0 Gaussian infinite series having the covariance operator explicitly described in (78). (3) says that one can conveniently construct from $\{X_n, Y_n\}$ an inferential procedure that behaves like $\delta_{n^*}(\mathbf{Z}_{n^*})$. Since $n^*/n \to 1$ such a procedure will usually be operationally equivalent asymptotically to what is obtainable from $\delta_n(\mathbf{Z}_n)$.

For the case of smoothness $\alpha \leqslant 1/2$ equivalence of the form (2) cannot hold. This follows from Brown and Zhang (1998). This is also the case for asymptotic relative sufficiency.

As already noted, we consider the case where the X_i are a random sample from some distribution H on $[0, 1]$. Our results hold when H has a density $h \in S_\alpha(B), \alpha > 1/2$. Most of them for cases where H is known or unknown, except that somewhat more stringent smoothness assumptions are needed on h when H is not known.

Transformations from the regression problem to the series problem occur frequently in the literature. We remark here on two recent treatments that have some relation to our approach. These are the formulation in Efromovich (1999; equation (4.2.5)), referred to as E, and in Donoho and Johnstone (1999), referred to as DJ. Both formulation refer primarily to the problem of estimating the regression functions under specific losses; and so in this way are less general than our formulation. In both of these, the loss is unbounded; hence the results involve matching the normalized limiting risks in the regression and series problems. By contrast, our results apply directly only to risks under bounded normalized loss functions. This directly yields statements about asymptotic risks, but not about limiting risks. However standard techniques allow one to extend such results to unbounded losses, although these techniques require focusing on specific classes of losses or procedures. See Brown and Low (1996) for some further discussion.

E refers to a Fourier series type of formulations; DJ to a wavelet formulation. Both transformations are different from ours, though that in E is somewhat closely related. (His analog of our step (i) uses a rectangular kernel, rather than the fixed bins we use; it may be possible to adapt our results to such a kernel step, but we have found our formulation convenient for our proofs and it is also perhaps slightly easier to implement. As another difference E uses a logarithmic rate bandwidth for his kernel step rather than the slightly larger polynomial rate bins we use. It may be that our (otherwise more general) results could be extended to allow such narrower bin widths but we have been unable to do so.)

The approach in DJ is more specifically tailored to wavelets, and so appears to be less closely related to our formulation or that in E. It may be of interest to note that DJ contains a randomization step analogous to our (ii). E does not, but this is probably related to the specific nature of the estimation goals and procedures there. See our Remark 5.5 for further discussion. Neither E nor DJ need to include a step like our (iii) since they each use only suitably truncated procedures in their respective series problems. The wavelet approach in DJ apparently allows their results to apply to a broader range of function spaces than those in our Section 4.4.

2 Nonparametric formulations

2.1 Nonparametric regression

One observes (X_i, Y_i), $i = 1, \ldots, n$. The X_i are independent random variables on a bounded set. Assume $\mathscr{X} = [0, 1] \subset \mathscr{R}^1$. They have distribution function H on $\mathscr{X}$. Given $X_i = x_i$ the real variables Y_i can be written as

$$Y_i = f(x_i) + \epsilon_i, \tag{4}$$

with ϵ_i independent $N(0, \sigma^2)$ random variables. Throughout we assume H is absolutely continuous with bounded derivative h, and satisfies $\inf_{x\in\mathscr{X}} h(x) \geqslant \epsilon$ for some known value $\epsilon > 0$.

We assume σ^2 is known. Without further loss of generality we take $\sigma^2 = 1$. Generalization to the case where σ^2 is unknown and may depend on x_i are possible but for the sake of brevity will not be pursued here. The basic results should also extend readily to $\mathscr{X} = [0,\ 1]^d \subset \mathscr{R}^d$.

The model (4) is referred to as "nonparametric" because the regression function, f, is only assumed to lie in a suitable infinite dimensional class of functions $\mathscr{F}$. Assume $\mathscr{F} \subset \mathscr{L}_2 = \{f :$

$\int f^2(t)dt < \infty\}$. For most applications it is necessary to formulate some further restrictions on $\mathscr{F}$ in order to obtain suitable results. This is also true of the equivalence results to follow. These assumptions take the form of smoothness restrictions on $\mathscr{F}$ and will be discussed in Section 2.4.

2.2 Infinite dimensional location problem

In this problem one observes an infinite sequence of independent normal random variables $\mathbf{Z}_n = \{Z_{nj},\ j = 1, 2, \ldots\}$ with $Z_{nj} \sim N(\theta_j, \sigma^2/n)$. The parameter vector $\boldsymbol{\theta} = \{\theta_j,\ j = 1, \ldots\}$ is assumed to lie in a (large) subset $\boldsymbol{\Theta}$ of $\mathscr{R}^\infty$.

Let $||\boldsymbol{\theta}||^2$ denote the usual ℓ_2 norm of $\boldsymbol{\theta}$ whenever this is finite, i.e.

$$||\boldsymbol{\theta}||^2 = \sum_{j=1}^{\infty} \theta_j^2.$$

Let $\ell_2 = \{\boldsymbol{\theta} : ||\boldsymbol{\theta}||^2 < \infty\}$. We assume throughout that $\boldsymbol{\Theta} \subset \ell_2$, and will generally need to also make additional assumptions.

The appearance of n in the variance, σ^2/n, can be understood as follows. Consider n independent observations of infinite dimensional vectors $\{W_{kj} :\ j = 1, \ldots\},\ k = 1, \ldots, n$, with each $W_{kj} \sim N(\theta_j, \sigma^2)$, independently. Then $\mathbf{Z}_n = \{Z_{nj} = \frac{1}{n}\sum_{k=1}^{n} W_{kj},\ j = 1, \ldots\}$ is a sufficient statistic for the experiment involving observation of $W_{kj},\ j = 1, \ldots;\ k = 1, \ldots, n$. For this reason we refer to n as the sample size in this setting. With no loss of generality we set $\sigma^2 = 1$ throughout the remainder of the paper.

The preceding definitions are suitable for matching to the regression problem where H is the uniform distribution on $[0, 1]$. They will be generalized in Section 5.1.

As is well known this formulation is isomorphic to a problem involving white-noise with drift. See Brown and Low (1996).

For the purpose of (asymptotically) matching this location problem to the nonparametric regression problem start with a given orthonormal basis, $\{\varphi_j\}$ of $L_2[0, 1]$. Then define the functional connection

$$\theta_j = \int_0^1 f(t)\varphi_j(t)\,dt = \langle \varphi_j,\ f \rangle. \tag{5}$$

2.3 Wavelet Series

Wavelet series are orthogonal expansions in L_2. For an introduction to wavelets in a statistical context see Donoho, et.al. (1994, 1995, 1998) or Cai (1999). In the usual format a wavelet basis for $[0,1]$ would be a complete orthonormal set of functions in L_2 with additional structure. This set of functions for given $K \geq 0$ consists of $\{\varphi_{K,m},\ m = 0,\ldots,2^K-1;\ \psi_{k,m},\ k = K+1,\ldots,\ m = 0,\ldots,2^{k-1}-1\}$. (The functions $\varphi_{K,m}$ are often referred to as father wavelets and the $\psi_{k,m}$ as mother wavelets.) Note that $\varphi_{K,m}(t) = 2^{K/2}\varphi(t - m/2^K)$, and the $\psi_{k,m}$ are also expressible in terms of φ. Let $\theta_{K,m} = \langle \varphi_{K,m}, f\rangle$, $\theta_{k,m} = \langle \psi_{k,m}, f\rangle$, $k = K+1,\ldots$; $Z_{K,m} = \langle \varphi_{K,m}, dZ_n\rangle$ and $Z_{k,m} = \langle \psi_{k,m}, dZ_n\rangle$, $k = K+1,\ldots$. Then $Z_{k,m} \sim N(\theta_{k,m}, 1/n)$, $k = K,\ldots,m = 0,\ldots,(2^{k-1}) \vee (2^K-1)$, correspond to an infinite dimensional location problem with parameter space $\Theta \subset \{\boldsymbol{\theta} = \{\theta_{k,m}\}\}$.

The Haar basis has properties that make it convenient for certain of our results. In this basis

$$\varphi_{j,k}(x) = 2^{j/2} I_{[k/2^j,(k+1)/2^j)}(x)$$

and

$$\psi_{j,k} = (\varphi_{j,2k+1} - \varphi_{j,2k})/\sqrt{2},$$

$k = 0,\ldots,2^{j-1}-1$.

The above double indexing system is convenient for discussions involving wavelets, and we will sometimes use it in this context. However it is desirable for a unified treatment of all infinite dimensional location problems to have a unified notation for such problems. For purposes of our general treatment we will write such a model as $\mathbf{Z} = \{Z_j : j = 1,\ldots\}$ and $\boldsymbol{\theta} = \{\theta_j : j = 1,\ldots\}$, etc.. A wavelet problem with $\{\theta_{k,m}\}$, etc., can be transferred to this form by the simple device of writing $j = j_{k,m} = 2^k - 2^K + m + 1$ for $k = K,\ldots;\ m = 0,\ldots,(2^{k-1} \vee (2^K-1))$. Consequently our general theory applies to wavelet series as well as to orthogonal series in the more conventional single index form, such as Fourier series.

2.4 Sobolev and Besov Spaces

Let $\alpha > 0$ and let $\{c_j : j = 1,\ldots\}$ be a sequence of non-negative constants satisfying $c_j \asymp j^{2\alpha}$. (i.e. for some $\epsilon > 0$, $\epsilon j^{2\alpha} < c_j < \epsilon^{-1} j^{2\alpha}, j = 1, ...$) The corresponding Sobolev or Besov-Hilbert ball is

$$S_\alpha(B) = \{\boldsymbol{\theta} : \, ||\boldsymbol{\theta}||_\alpha^2 \triangleq \sum c_j \theta_j^2 < B\}. \tag{6}$$

The entire space is of course, $S_\alpha = S_\alpha(\infty)$.

If $\{\varphi_j\}$ is the Fourier cosine basis $\{\varphi_j(t) = (\sqrt{2})^{\text{sgn}(j-1)} \cos(\pi(j-1)t), \;\; j = 1, \ldots\}$ and α is an integer then S_α corresponds to bounded functions whose $(\alpha - 1)^{th}$ derivative is absolutely continuous and who satisfy

$$\pi^{-2\alpha} \int_0^1 f^2(t)dt + \int_0^1 (f^{(\alpha)}(t))^2 dt < \infty. \tag{7}$$

If we further take $c_j = (j-1)^{2\alpha}$ then the left side of (7) is exactly $\pi^{2\alpha}||\boldsymbol{\theta}||_\alpha^2$. The Sobolev space S_α, $\alpha > 0$, is a natural generalization. (One can use the usual sine-cosine basis to get a similar representation for periodic functions.)

If $\{\varphi_j\}$ corresponds to a wavelet basis, then S_α corresponds to a Besov-Hilbert space and the sets $S_\alpha(B)$ are balls within that space. For more about Besov spaces see the previously cited references about wavelets. In either case, sets such as $S_\alpha(B)$ are a convenient vehicle for controlling the smoothness properties of the corresponding functions. For convenience we will also use the notation $S_\alpha(B)$ to refer to the corresponding set of functions in situations like the above.

We will assume throughout that the basis elements satisfy

$$|\varphi_j(x)| \leqslant C\sqrt{j} \quad j = 1, \ldots \tag{8}$$

for some generic $C < \infty$. The usual Fourier and wavelet bases satisfy this assumption. In addition, for certain results, but not for others, we will need to make the stronger assumption that $|\varphi_j(x)|$ is bounded, as follows.

Assumption B: The functions $|\varphi_j|$ are uniformly bounded in both j and x.

This assumption is satisfied by the Fourier basis and other related bases. It is not satisfied by the usual wavelet bases.

The following standard example is included here and discussed later in more detail in order to emphasize how the series problem can be conveniently manipulated to construct attractive statistical estimators. See Donoho, et.al. (1995) and other wavelet references for more information.

Example 2.1: Suppose $(\{\varphi_{K,m}\}, \{\psi_{k,m}\})$ is a compactly supported differentiable wavelet basis and $\mathscr{F} \subset S_\alpha(B)$ for some $\alpha > 1/2$, $B < \infty$. Consider an estimator of $\boldsymbol{\theta}$ with coordinates given by

$$\check{\theta}_{k,m} = \begin{cases} Z_{k,m} \quad \text{if } k \le K_0 \\ \operatorname{sgn}(Z_{k,m})(|Z_{k,m}| - \lambda_n)_+ \quad \text{if } k > K_0 \text{ and } 2^k \le n \\ 0 \quad \text{if } 2^k > n \end{cases} \tag{9}$$

where $\lambda_n = \sqrt{2(\log n)/n}$. It is now well known that this estimator is within a factor of the order of $\log n$ of being asymptotically minimax under squared error, independent of α. (Donoho and Johnstone (1995).)

To close this discussion we want to emphasize an additional feature of the above procedure, namely that it is non-linear. Verification of the asymptotic minimax property involves non-linear features of the normal distribution. In particular, it uses the fact that $\lim_{n\to\infty} P(\sup_{1\le j\le n} |Z_j| > \sqrt{2(\log n)/n}\ ) = 0$ when $Z_j \sim N(0, 1/n)$, independent. (This implies that if $\boldsymbol{\theta} = 0$ then $\lim_{n\to\infty} P(\check{\boldsymbol{\theta}} = 0) = 1$.)

Estimators adaptively achieving the minimax rate or the rate and its constant can also conveniently be determined in this context. See for example Donoho and Johnstone (1995), Cai (1999), Zhang (2000) and Cai, Low and Zhao (2000). These estimators are more prominently non-linear than $\check{\boldsymbol{\theta}}$ in the preceding example.

Remark: We have assumed the $\{\varphi_j\}$ are orthogonal on $[0, 1]$. We take advantage of this fact to simplify certain arguments. However it is not necessary. It is fairly straightforward to generalize results to bases $\{\varphi_j\}$ that satisfy a condition

$$\int_0^1 \varphi_j(t)\varphi_k(t)g(t)\,dt = 0 \text{ if } j \ne k$$

for some g satisfying $\epsilon < g < 1/\epsilon$. Some other bases $\{\varphi_j\}$ may also be accommodated with some additional case.

3 Asymptotic Equivalence and Dominance

The objective of this paper is to derive a simple explicit asymptotic equivalence mapping taking the nonparametric regression problem to the infinite dimensional location problem. In this preliminary section we describe the nature of the desired mapping and provide some other necessary information related to asymptotic equivalence. We begin with a counter example intended to help motivate our objective. This example shows that an intuitively natural mapping fails to have the desired properties. The example also helps explain the features of the equivalence definition adopted later in this section.

3.1 Counterexample

Consider the nonparametric regression setting of Section 2.1. Assume H is the uniform distribution. An equivalence mapping would use $\{(X_i, Y_i)\}$ to produce an infinite vector $\hat{\mathbf{Z}}$, say, having asymptotically identical statistical properties to $\mathbf{Z}$ in the infinite dimensional location problems. In particular the coordinates of $\hat{\mathbf{Z}}$ should asymptotically satisfy

$$\sqrt{n}(\hat{Z}_j - \theta_j) \to N(0,1) \tag{10}$$

in distribution where $\theta_j = \langle f, \varphi_j \rangle$ and where the $\hat{Z}_j$ are also asymptotically independent in a suitable fashion.

An intuitive attempt to achieve this might involve defining the empirical coefficients

$$\hat{Z}_j = \frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) Y_i. \tag{11}$$

With this definition

$$E(\hat{Z}_j) = E(\varphi_j(X)Y) = \int \varphi_j(x) f(x)\, dx = \theta_j \tag{12}$$

exactly as desired. However,

$$\begin{aligned} n\mathrm{Var}(\hat{Z}_j) \quad &= E(\mathrm{Var}(\varphi_j(X)Y|X)) + \mathrm{Var}(E(\varphi_j(X)Y|X)) \\ &= \textstyle\int \varphi_j^2(x)\, dx + \int (\varphi_j(x) f(x) - \int \varphi_j(t) f(t)\, dt)^2\, dx \\ &= 1 + \Delta_j(f), \end{aligned} \tag{13}$$

say, where $\Delta_j(f) \geq 0$, and $\Delta_j(f) > 0$ unless $\varphi_j(x) f(x)$ is a constant. Hence (10) is not generally valid. A similar calculation would also show that the $\hat{Z}_j$ need not be asymptotically uncorrelated, contrary to what is desired. See also Efromovich (1999, p.128).

3.2 Asymptotic Equivalence

The following definition of asymptotic equivalence entails (10) but is stronger in important respects. Let $\hat{\mathbf{Z}}_n$ be a possibly randomized function of $\{(X_i, Y_i) : i = 1, \ldots, n\}$. In much of what follows we will for convenience suppress the dependence on n from the notation, and just write $\hat{\mathbf{Z}} = \hat{\mathbf{Z}}_n$. The distribution of $\{(X_i, Y_i)\}$ of course depends on $f \in \mathscr{F}$. We consider $\mathbf{Z} = \mathbf{Z}_n$ to be the infinite dimensional observation in the infinite dimensional location problem with $\boldsymbol{\theta} = \boldsymbol{\theta}(f)$ having coordinates $\theta_j = \langle \varphi_j, f \rangle, \ j = 1, \ldots$. Let $G_{\hat{\mathbf{Z}}}$ and $G_{\mathbf{Z}}$ denote the distributions of $\hat{\mathbf{Z}}$ and $\mathbf{Z}$ on $\mathbf{R}^\infty$. We say $\hat{\mathbf{Z}}$ is **asymptotically equivalent** to $\mathbf{Z}$ if

$$||G_{\hat{\mathbf{Z}}} - G_{\mathbf{Z}}||_{\text{Total Variation}} \longrightarrow 0 \tag{14}$$

uniformly for $f \in \mathscr{F}$.

A consequence of (14) is that if T_n is any sequence of random variables then $T_n(\hat{\mathbf{Z}})$ and $T_n(\mathbf{Z})$ strongly converge to each other in distribution, uniformly for $f \in \mathscr{F}$. For T_n real-valued (14) can be rewritten as

$$\lim_{n\to\infty} \sup_{T_n} \sup_{f\in\mathscr{F}} \sup_{b\in(-\infty,\infty)} |P_f(T(\hat{\mathbf{Z}}) < b) - P_{\boldsymbol{\theta}(f)}(T(\mathbf{Z}) < b)| = 0. \tag{15}$$

Note that this convergence is considerably stronger than (10). Suppose that the coordinates of $\hat{\mathbf{Z}}$ are independent, or are asymptotically independent in an appropriate sense, and (10) holds uniformly in $j = 1, \ldots$. Then a version of (15) follows in which the supremum over T_n is restricted to uniformly bounded linear functionals of $\mathbf{Z}$. However (15) would not necessarily follow for general, non-linear functionals. In view of the increasing importance of non-linear functionals, as suggested by Example 2.1, it is desirable to obtain equivalence in the stronger sense of (14) or (15).

Equivalence as defined here is an alternative version of the previously mentioned notion of asymptotic statistical equivalence as defined by Le Cam (1953, 1964, 1986). It is equivalent to the following statistical formulation. Let $\mathscr{D}$ denote the set of all decision procedures on an action space and let $\mathscr{L}$ denote the set of non-negative loss functions on that space bounded by 1. Then

$$\sup_{\delta\in\mathscr{D}} \sup_{L\in\mathscr{L}} \sup_{f\in\mathscr{F}} |E_f(L(\delta(\hat{\mathbf{Z}}), f)) - E_{\boldsymbol{\theta}(f)}(L(\delta(\mathbf{Z}), \boldsymbol{\theta}(f)))| \longrightarrow 0 \tag{16}$$

as $n \to \infty$.

In practice, it is convenient to verify (14) via probability densities. Let $g_{\mathbf{Z}_n}$, $g_{\hat{\mathbf{Z}}_n}$ denote the probability densities of $\mathbf{Z}_n$ and $\hat{\mathbf{Z}}_n$ with respect to some common dominating measure. (This dominating measure may depend on n.) These densities also depend on $f \in \mathscr{F}$. Then (14) holds if and only if

$$||g_{\hat{\mathbf{Z}}} - g_{\mathbf{Z}}||_{L_1} \longrightarrow 0 \text{ uniformly for } f \in \mathscr{F} \text{ as } n \to \infty. \tag{17}$$

We will use the notation $\{\hat{\mathbf{Z}}_n\} \approx \{\mathbf{Z}_n\}$ to describe such an equivalence.

3.3 Equivalence for Gaussian Experiments

Suppose $\mathbf{V}_n$ and $\mathbf{W}_n$ are d dimensional Gaussian random variables with means $\boldsymbol{\nu}_n, \boldsymbol{\omega}_n$, respectively, and covariance matrices $\boldsymbol{\Sigma}_n/n$ and $\mathbf{B}_n/n$, respectively. Assume $\boldsymbol{\Sigma}_n$ has eigenvalues uniformly bounded below by C^{-1} for some $C > 0$. Let $g_{\mathbf{V}_n}$ and $g_{\mathbf{W}_n}$ denote their densities. Then a direct calculation yields

$$||g_{\mathbf{V}_n} - g_{\mathbf{W}_n}||_{L_1}^2 \leq 8\{n||\boldsymbol{\nu}_n - \boldsymbol{\omega}_n||^2 C - \operatorname{tr}[(\mathbf{I} - \boldsymbol{\Sigma}_n^{-1}\mathbf{B}_n)^2]\}. \tag{18}$$

See Brown, Cai, Low and Zhang (1999). The same inequality holds for infinite dimensional Gaussian variables with means $\widetilde{\nu}_n$, $\widetilde{\omega}_n$ and covariance operators $\boldsymbol{\Sigma}_n/n$ and $\mathbf{B}_n/n$.

This simple inequality will later be applied to prove $\mathbf{Z}_n^{(d)}$ and $\hat{\mathbf{Z}}_n^{(d)}$ are asymptotically equivalent in the sense of (17), where $\mathbf{Z}_n^{(d)} = \{Z_{ni} : i = 1, \ldots, d\}$ and similarly for $\hat{\mathbf{Z}}_n^{(d)}$. (Here, d may depend on n.)

3.4 Notation; Asymptotic Equivalence and Sufficiency

Let $\{V_n\}$ and $\{W_n\}$ be the observations in two sequences of statistical problems having the same parameter space Θ. If there are functions ν_n such that $dist_\theta(W_n) = dist_\theta(\nu_n(V_n))$ for all $\theta \in \Theta$, then of course for each n V_n is sufficient for W_n. Sufficiency for each n also holds if ν_n is a randomized map. We will write $V_n :\longrightarrow W_n$ to denote the situation where V_n is sufficient for W_n for each n.

In Section 3.2 we introduced an asymptotic notion of equivalence. In this notion $\{V_n\} \approx \{W_n\}$ if their distribution satisfy (14) or, equivalently, if their densities satisfy (16). (For notational convenience we will also write this as $V_n \approx W_n$, and similarly for other similar notation.) These two situations can be combined, as follows. Suppose $V_n :\rightarrow W'_n$ and $W'_n \approx W_n$. Then we say $\{V_n\}$ is asymptotically sufficient for $\{W_n\}$ and write $\{V_n\} \rightsquigarrow \{W_n\}$.

In the sequel we will establish results about asymptotic equivalence and asymptotic sufficiency. In connection with these results the forms of the relevant (randomized) maps $\{\nu_n\}$ will be established alongside. In this way our results will enable explicit construction of procedures of the form $\delta_n(\nu_n(V_n))$ that have the same asymptotic performance characteristics (in the sense of (16)) as those of the form $\delta_n(W_n)$.

Our emphasis is thus on verifying (16) in a manner that allows explicit construction. See Example 2.1 (cont.) in Section 5.2 for further discussion.

4 Construction and Basic Results for Uniform H

4.1 Preliminarily Binned Estimators

The construction is built upon a sequence $m = m(n) \longrightarrow \infty$ such that

$$\frac{m(n)}{n^{1-\epsilon}} \longrightarrow 0 \text{ for some } \epsilon > 0. \tag{19}$$

Additional assumptions about m are described following (23), below.

Let $I_k = [(k-1)/m, k/m]$, $k = 1, \ldots, m$. This creates a preliminary binning of $[0, 1)$ into bins of length $1/m$. For $k = 1, \ldots, m$ let

$$\begin{aligned} N_k &= \#\{X_i \in I_k\}, \ \ \overline{Y}_k = \frac{1}{N_k} \sum_{X_i \in I_k} Y_i \\ \overline{\Phi}_k &= m \int_{I_k} \Phi(x)\, dx \text{ where } \Phi = \{\varphi_j : j = 1, \ldots\}. \end{aligned} \tag{20}$$

($\overline{Y}_k$ are of course the empirical averages of those Y_i whose X_i are in the k^{th} bin. The j^{th} coefficient of $\overline{\Phi}_k$ is the average of φ_j over the k^{th} bin. The choice of m will guarantee that all $N_k > 0$ with a probability exponentially close to 1. In the rare case that some $N_k = 0$ the above definitions need minor modification. Any well-defined modification suffices from a theoretical perspective, but a pragmatically satisfying choice would be to let $\overline{Y}_k$ then be a weighted average of neighboring values $\overline{Y}_{k'}$ having $N_{k'} > 0$.) Let $\overline{\Phi}^{(m)} = \{\overline{\Phi}_k : k = 1, \ldots, m\}$, an $m \times 1$ vector, and define

$$\widetilde{\mathbf{Z}} = \widetilde{\mathbf{Z}}_n = \frac{1}{m(n)} \sum_{k=1}^{m(n)} \overline{Y}_k \overline{\Phi}_k. \tag{21}$$

$\widetilde{\mathbf{Z}}_n$ has coordinates $(\widetilde{\mathbf{Z}}_n)_j = \widetilde{Z}_{nj}$.

Define the stepwise averaged regression functions $\overline{f}_m$ by a formula analogous to (20):

$$\overline{f}_m(x) = m \int_{I_k} f(t)\, dt \text{ for } x \in I_k \quad k = 1, \ldots, m. \tag{22}$$

Where convenient we sometimes write $\overline{f}$ in place of $\overline{f}_m$.

The assumptions for subsequent results are related to this binning step. For most results we need to assume $m(n)$ satisfies (19) and

$$n \int_0^1 (f(t) - \overline{f}_{m(n)}(t))^2 dt \longrightarrow 0, \text{ uniformly for } f \in \mathscr{F}. \tag{23}$$

Section 7 discusses the situations where either $\{\varphi_j\}$ is a Fourier basis or a compactly supported wavelet basis. It is shown in Lemmas 7.1 and 7.2 that if $\mathscr{F} \subset S_\alpha(B)$ for some $\alpha > 1/2$ then m can be chosen to satisfy both (19) and (23), and will satisfy

$$\frac{m(n)}{n^{1/2\alpha}} \longrightarrow \infty. \tag{24}$$

In Theorem 5.1 we need to make a stronger assumption to deal with wavelet bases. We assume there that (23) is satisfied by $m = m(n)$ such that

$$\frac{m(n)}{n^{1/2-\epsilon}} \longrightarrow 0. \tag{25}$$

It also follows from Lemma 7.2 that this condition can be satisfied by wavelet bases whenever $\mathscr{F} \subset S_\alpha(B)$ for $\alpha > 1$. (Theorem 6.1 does not require this stronger assumption.)

For other orthogonal bases one may need to directly assume that (23) is satisfied as well as $\mathscr{F} \subset S_\alpha(B)$ for some $\alpha > 1/2$.

4.2 Truncation to m coordinates

The $\widetilde{\mathbf{Z}}_n$ of (21) asymptotically provide an almost entirely satisfactory substitute for the $\mathbf{Z}_n$ of the original infinite dimensional location model. (This contrasts with the counterexample in Section 3.1, where the $\hat{\mathbf{Z}}_n$ do not.) However, a preliminary truncation step is also needed. The necessity for such a step should be intuitively clear, since the coefficients of $\widetilde{\mathbf{Z}}_n$ are linear functions of the m values $\{\overline{Y}_k : k = 1, \ldots, m\}$. Thus, for $J > m$, the coordinates $\{\widetilde{\mathbf{Z}}_{nj} : j = m+1, \ldots, J\}$ will usually be linear functions of $\widetilde{\mathbf{Z}}_n^{(m)} = \{\widetilde{\mathbf{Z}}_{nj} : j = 1, \ldots, m\}$. This strongly contrasts with the situation for $\mathbf{Z}_n$ where $\{Z_{nj} : j = m+1, \ldots, J\}$ is independent of $Z_n^{(m)} = \{\mathbf{Z}_{nj} : j = 1, \ldots, m\}$. This difference is not a serious problem since as we now show $\{\mathbf{Z}_{nj} : j = m+1, \ldots, \}$ do not carry any asymptotically useful information about $f \in \mathscr{F}$.

For each n, let $\{T_j : j = 1, \ldots\}$ be an independent auxiliary sequences of independent Gaussian random variables with $T_j \sim N(0, 1/n)$. For notational convenience we have suppressed the dependence of T on n. Then define $\mathbf{Z}_n^o$ to have coordinates

$$\mathbf{Z}_{nj}^o = \begin{cases} Z_{nj} & j = 1, \ldots, m \\ T_j & j = m+1, \ldots \end{cases} \tag{26}$$

The following lemma shows that no asymptotically useful information is lost in passing from $\mathbf{Z}_n$ to $\mathbf{Z}_n^\circ$.

Lemma 4.1 *Assume H is uniform. Let $\alpha > 1/2$ and let $\mathscr{F} \subset S_\alpha(B)$. Assume $m = m(n)$ satisfies (24). Then $\{\mathbf{Z}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n\}$ according to $\{\mathbf{Z}_n^{(m)}\} :\longrightarrow \{\mathbf{Z}_n^\circ\} \approx \{\mathbf{Z}_n\}$.*

Proof: $\mathbf{Z}_n^\circ$ and $\mathbf{Z}_n$ each have independent Gaussian coordinates with variance $1/n$. By properties of S_α they also satisfy

$$n\,||E(\mathbf{Z}_n^\circ) - E(\mathbf{Z}_n)||^2 = \sum_{j=m+1}^{\infty} \theta_j^2 \le O\left(\frac{n}{m^{2\alpha}} \sum_{j=m+1}^{\infty} c_j \theta_j^2\right) \longrightarrow 0 \tag{27}$$

since $n/m^{2\alpha} \to 0$ by assumption (24). The assertion that $\mathbf{Z}_n^\circ \approx \mathbf{Z}_n$ then follows from (18). The remaining assertion concerning is trivial since $\mathbf{Z}_n^\circ$ is produced from $\mathbf{Z}_n^{(m)}$ by the randomizations described in (26). $\Box$.

4.3 Asymptotic equivalence for H uniform

Many current applications of our theory involve uniform H. This situation is also notationally simpler than the case of general H and Lemma 4.1 makes possible a slightly more convenient conclusion. Consequently we state a special version here of our theorem even though this is really a special case of the more general result in the next section. To save duplication we defer the major part of the proof of Theorem 4.1 to the next section, where it will follows as a special case of Theorem 5.1.

Theorem 4.1 *Assume H is uniform and (23) holds. Assume $\mathscr{F} \subset S_\alpha(B)$ for some fixed, known values of α, B. Let $m(n)$ be any sequence satisfying (24) and the following:*

If Assumption B also holds assume $\alpha > 1/2$ and (19) holds.

If Assumption B does not hold assume $\alpha > 1$ and (25) holds.

Then

$$\{X_i, Y_i : i = 1, \ldots, n\} :\longrightarrow \{\widetilde{\mathbf{Z}}_n\} :\longrightarrow \{\widetilde{\mathbf{Z}}_n^{(m)}\} \approx \{\mathbf{Z}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n^{\circ}\} \approx \{\mathbf{Z}_n\} \tag{28}$$

Proof: The first map in (28) is given by (20),(21). The second is a simple truncation. The last two parts of (28) are described in (26) and proved in Lemma 4.1. The assertion that $\widetilde{\mathbf{Z}}_n^{(m)} \rightsquigarrow \mathbf{Z}_n^{(m)}$ follows as a special case of Theorem 5.1. That theorem also describes a randomized map that explicitly yields this assertion of asymptotic sufficiency. $\square$

5 Basic results, general H

For the situation where H is not uniform the definition of the location problem involving $\mathbf{Z}$ needs to be generalized. The definitions of $\widetilde{\mathbf{Z}}_n$, $m = m(n)$ and $\widetilde{\mathbf{Z}}_n^{(m)}$ remain exactly as before. The main theorem of this section shows that $\widetilde{\mathbf{Z}}_n^{(m)}$ and $\mathbf{Z}_n^{(m)}$ are asymptotically equivalent under suitable assumptions like those in Theorem 4.1.

In Section 5.3 we show that $\mathbf{Z}^{(m)}$ is asymptotically equivalent to $\mathbf{Z}$. However, in many situations one can proceed to directly develop results and theory based only on $\mathbf{Z}^{(m)}$. This seems in any case a fairly natural course of action since $m(n)/n \to 0$ only slowly and so to use only $\mathbf{Z}^{(m)}$ usually does not prove a serious practical limitation.

5.1 The general location problem

Let $\Phi = \{\varphi_j : j = 1, \ldots,\}$ denote the infinite dimensional column vector of basis functions φ_j. We continue to assume that these basis functions are orthonormal in the usual $L_2[0, 1]$. (This assumption is convenient but could be somewhat relaxed.)

Define a covariance operator (matrix) via the infinite dimensional matrix formula

$$\mathbf{M} = \int_0^1 \Phi(t)\Phi'(t)\frac{1}{h(t)}\,dt \tag{29}$$

Let $\mathbf{M}^{(m)}$ denote the upper left $m \times m$ submatrix of H. (Recall that $\inf_t h(t) \geqslant \epsilon > 0$, so $\mathbf{M}$ is well defined.) Then define $\mathbf{Z}_n$ to be the infinite dimensional Gaussian vector with

$$E(\mathbf{Z}_n) = \boldsymbol{\theta}, \quad cov\,\mathbf{Z}_n = \mathbf{M}/n. \tag{30}$$

See Kuo (1975) and Mandelbaum (1984) for existence and properties of such infinite dimensional random variables.

As before we interpret the coordinates of $\boldsymbol{\theta}$ according to (12). To motivate the definition of $\mathbf{M}$ note that the conditional covariance matrix of $\widetilde{\mathbf{Z}}_n^{(m)}$ given N_k is

$$condcov(\widetilde{\mathbf{Z}}_n^{(m)}) = \frac{1}{m^2}E(\sum \frac{1}{N_k}\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})'|\{N_k\}) \tag{31}$$

where $\overline{\Phi}_k^{(m)}$ is the m-dimensional column vector with coordinates

$$(\overline{\Phi}_k^{(m)})_j = m\int_{I_k} \varphi_j(t)\,dt, \quad j = 1, \ldots, m.$$

Consequently

$$condcov(\widetilde{\mathbf{Z}}_n^{(m)})_{ij} = \frac{\mathbf{M}_{ij}}{n}(1 + o_p(1)) \tag{32}$$

as $n \to \infty$. A second related motivation may be found through the white-noise equivalence result in Brown and Low (1996).

5.2 Main Theorems, known H

When H is known a result like Theorem 4.1 and its corollary continues to hold to relate $\widetilde{\mathbf{Z}}^{(m)}$ and $\mathbf{Z}^{(m)}$. We assume below that h satisfies a condition like (23), that is

$$n\int (h(t) - \overline{h}_{m(n)}(t))^2 dt \leqslant B \;\; for\ all\ n. \tag{33}$$

Theorem 5.1 *Assume H is a known function and satisfies (33). Make the other assumptions of Theorem 4.1, and let $m(n)$ be as in that theorem. Then with $\mathbf{Z}$ as in (30)*

$$\{X_i, Y_i : i = 1, \ldots, n\} :\rightarrow \{\widetilde{\mathbf{Z}}_n\} :\rightarrow \{\widetilde{\mathbf{Z}}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n^{(m)}\}. \tag{34}$$

The final step in (34) is described more concretely by

$$\{\widetilde{\mathbf{Z}}_n^{(m)} + V_n^{(m)}\} \approx \{\mathbf{Z}_n^{(m)}\} \tag{35}$$

where

$$V_n \sim N_m(0, A_n^{(m)}/n) \textit{ independent of } \widetilde{\mathbf{Z}}_n^{(m)}$$

with

$$A_n^{(m)} = \mathbf{M}^{(m)} - \int \frac{1}{h(t)} \overline{\Phi}^{(m)}(t)(\overline{\Phi}^{(m)}(t))'\, dt \tag{36}$$

Remark 5.1: Formula (36) and the randomization step (35) can be described in alternative ways. For example, let $\overline{h}_k = \overline{h}(k/(m+1))$ and $\overline{\Phi}_k = \overline{\Phi}(k/(m+1))$. Then two other descriptions of A are

$$\begin{aligned} A_n^{(m)} &= \int \frac{1}{h(t)}(\Phi(t) - \overline{\Phi}^{(m)}(t))(\Phi(t) - \overline{\Phi}^{(m)}(t))'\, dt \\ &= \mathbf{M}^{(m)} - \frac{1}{m}\sum_{k=1}^{m} \frac{1}{\overline{h}_k} \overline{\Phi}_k^{(m)} (\overline{\Phi}_k^{(m)})'. \end{aligned} \tag{37}$$

Formulas (37) and (83) show that when H is uniform and $\{\varphi_j\}$ is a Fourier basis then A is a diagonal matrix.

Remark 5.2: The randomization in (35) can also be implemented as follows: Assume for convenience that $L = n/m$ is an integer. Generate auxiliary independent standard normal random variables $\{W_{kl} : l = 1, \dots L,\ k = 1, \dots m\}$. Let

$$Y_{kl}^* = \frac{1}{\sqrt{\overline{h}_k}}(W_{kl} - W_{k\cdot}) + \overline{Y}_k, \quad k = 1, \dots, m,\ l = 1, \dots, L. \tag{38}$$

Let X_{kl}^*, $l = 1, \dots, L$ be independent uniform variables on I_k, $k = 1, \dots, m$. Then define

$$\widetilde{\mathbf{Z}}_n^{(m)*} = \frac{1}{n}\sum_{k=1}^{m}\sum_{l=1}^{L} \Phi^{(m)}(X_{kl}^*) Y_{kl}^*. \tag{39}$$

It can be checked directly that $\{\widetilde{\mathbf{Z}}_n^{(m)*}\} \approx \{\widetilde{\mathbf{Z}}_n^{(m)} + V_n^{(m)}\}$.

Remark 5.3: It may be computationally more convenient to use a binned version of $\mathbf{M}$. Define

$$\overline{\mathbf{M}} = \frac{1}{m}\sum \frac{1}{\overline{h}_k} \overline{\Psi}_k \tag{40}$$

where $\overline{\Psi}_k = m \int_{I_k} \Phi(t)\Phi'(t)\, dt$. Then calculations like those in the proof of the theorem show that $\overline{\mathbf{M}}$ may be substituted for $\mathbf{M}$ in the statement of the theorem.

Remark 5.4: In the case where H is uniform Theorem 4.1 shows that $\{\mathbf{Z}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n^\circ\} \approx \{\mathbf{Z}_n\}$ in addition to (34). Theorem 5.2 below, shows that $\{\mathbf{Z}_n\} \rightsquigarrow \{\widetilde{\mathbf{Z}}_n^{(m)}\}$. Thus in the special case when H is uniform the randomization contained in the relation $\{\widetilde{\mathbf{Z}}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n^{(m)}\}$ does not lose any asymptotically useful information. This pleasant state of affairs holds for general H as well, also as a consequence of Theorem 5.2.

Remark 5.5: Suppose one has satisfactory procedures $\delta_n(\mathbf{Z}_n^{(m)})$ under some asymptotic criterion. Then the theorem says that $\delta_n(\widetilde{\mathbf{Z}}_n^{(m)} + V_n^{(m)})$ will have the same asymptotic distributional properties, uniformly over $f \in \mathscr{F}$. Note that the term $V_n^{(m)}$ represents a very small amount of additional noise added to $\widetilde{\mathbf{Z}}^{(m)}$. In many situations $\delta_n(\widetilde{\mathbf{Z}}_n^{(m)})$ can be shown to have the same asymptotic distribution as $\delta_n(\widetilde{\mathbf{Z}}_n^{(m)} + V_n^{(m)})$. In this case $\delta_n(\widetilde{Z}_n^{(m)})$ is somewhat simpler to use and is asymptotically entirely satisfactory. The following is a very simple illustration.

Example 2.1 (cont.) In most current literature orthogonal series estimators are defined for the independent coordinates situation of Section 2.2. The present correlated coordinates setting is a variant that will often require additional considerations. Such consideration are beyond the scope of our paper, and we include here only a few remarks about the simplest of such situations.

It appears that theory for linear estimators can easily be generalized. This includes kernel estimators for predetermined non-adaptive bandwidths and smoothing spline estimators with non-adaptive weight functions. This could be accomplished either by directly calculating the distributions of the proposed estimators or by linearly transforming the problems via the mapping $\mathbf{Z} \to \mathbf{M}^{-1/2}\mathbf{Z}$. One computational problem with the latter approach is that it may require knowledge of properties of $\mathbf{M}^{-1/2}\mathscr{F}$. Even if $\mathscr{F}$ has convenient properties such as orthosymmetry (see Donoho, Liu and MacGibbon (1990)) it need not be the case that $\mathbf{M}^{-1/2}\mathscr{F}$ also has these properties.

It seems that much of the theory for coordinatewise projection estimators can also be generalized in a relatively straightforward fashion. This is because the performance of these estimators depends primarily on only the diagonal terms Σ_{ii}. For example, it appears one should begin by modifying the formula (9) to redefine

$$\breve{\theta}_{k,m} = \begin{cases} Z_{km}, \text{ if } k \leqslant K_0 \\ sgn(Z_{km})(|Z_{km}| - \lambda_{n,km})_+ \text{ if } K > K_0 \text{ and } 2^k \leqslant n \\ 0 \text{ if } 2^k > n \end{cases}$$

where $\lambda_{n,km} = \sqrt{\Sigma_{km,km}}\sqrt{2(\log n)/n}$. This formula could also be truncated by setting $\breve{\theta}_{km} = 0$ for $2^k > m(n)$. It can easily be checked there is no asymptotic loss of precision if this is done.

It is also easy to check that this estimator has asymptotically identical performance when $\widetilde{\mathbf{Z}}_n^{(m)}$ is used in place of $\widetilde{\mathbf{Z}}_n^{(m)} + V_n^{(m)}$. The key fact is that $var[(\widetilde{\mathbf{Z}}_n^{(m)})_j]/var[(\widetilde{\mathbf{Z}}_n^{(m)})_j + (V_n^{(m)})_j] \to 1$, along with the coordinatewise structure of $\breve{\theta}$.

Proof of Theorem 5.1: The first step of the proof is to verify that the problem with $\overline{h}_m$, $\overline{f}_m$ in place of h, f is asymptotically equivalent to the original problem. Let q ($\overline{q}_m$, respectively) denote the density of X, Y under h, f ($\overline{h}, \overline{f}$). Thus

$$q(x,y) = h(x)\varphi(y - f(x)). \tag{41}$$

with $\overline{q}_m$ defined similarly but with $\overline{h}$, $\overline{f}$ in place of h, f. According to standard results involving the Hellinger metric the asymptotic equivalence will follow if

$$nH_n^2 = n\int (q^{1/2}(x,y) - \overline{q}_m^{1/2}(x,y))^2\, dx\, dy \to 0. \tag{42}$$

(See Brown and Low (1996).). Then

$$\begin{aligned} n\,H_n^2 &\leqslant 2n\left[\int (h^{1/2}(x) - \overline{h}_m^{1/2}(x))^2\, dx \right. && (43) \\ &\quad \left. + \int\int (\varphi^{1/2}(y - f(x)) - \varphi^{1/2}(y - \overline{f}(x)))^2\, dy\, \overline{h}(x)\, dx \right] && (44) \\ &\leqslant 2n\left[\left(\frac{1}{4\epsilon}\right)\int (h(x) - \overline{h}_m(x))^2 dx \;+\; \left(\frac{C}{4}\right)\int (f(x) - \overline{f}_m(x))^2\, dx\right] && (45) \end{aligned}$$

here, $C = \sup_{x\in\mathscr{X}} \overline{h}_m(x) \leqslant \sup_{x\in\mathscr{X}} h(x)$. It follows from (23) and (33) that $nH_n^2 \to 0$, as desired. Consequently when deriving results about the asymptotic distribution, uniformly over $f \in \mathscr{F}$, we may calculate these distribution as if the true densities are $\overline{h}$, $\overline{f}$ for $f \in \mathscr{F}$.

Similar reasoning shows that the family of distributions of $\mathbf{Z}$, under h, f is asymptotically equivalent to the same family over $\overline{h}_m$, $\overline{f}_m$, for $f \in \mathscr{F}$. See Brown and Low (1996) for details.

Under $\overline{h}, \overline{f}$ the statistics $\{N_k, \overline{Y}_k : k = 1, \ldots, m\}$ are sufficient for $\{X_i, Y_i : i = 1, \ldots, m\}$. Under $\overline{h}, \overline{f}$ and conditional on $\{N_k : k = 1, \ldots, m\}$ the distribution of $\widetilde{\mathbf{Z}}_n^{(m)}$ is multivariate normal. It has mean matrix

$$\overline{\Theta}^{(m)} = E_{\overline{h},\overline{f}}(\widetilde{\mathbf{Z}}_n^{(m)}) = \int \overline{\Phi}^{(m)}(t)\overline{f}(t)dt = E_{\overline{h},\overline{f}}(\mathbf{Z}_n^{(m)}).$$

It has covariance matrix

$$\Sigma^{(m)} = \frac{1}{m^2}\sum \frac{1}{N_k}\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})'. \tag{46}$$

Consider, the matrix D, say, where

$$\begin{aligned} D &= \frac{1}{m}\sum \frac{1}{\overline{h}_k}\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})' - \Sigma^{(m)} && (47) \\ &= \frac{1}{m}\sum(\frac{1}{\overline{h}_k} - \frac{1}{mN_k/n})\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})'. && (48) \end{aligned}$$

Note that the N_k are multinomial $(n, \overline{h})$. Hence, in particular, $E(mN_k/n) = \overline{h}$ and $var(mN_k/n) = m/n$.

Let ν be any vector in $\mathbf{R}^{(m)}$ with $||\nu|| = 1$. Under Assumption B (uniform boundedness of $|\varphi_j|$) we have that $\nu' D\nu$ is asymptotically normal with mean 0 and variance

$$\frac{1}{m^2}\sum \frac{1}{\overline{h}_k^2}(\nu_k'\overline{\Phi}_k^{(m)})^2\frac{m}{n} = O_p(\frac{1}{n}). \tag{49}$$

If ν is an eigenvector of D we thus have that the corresponding eigenvalue, λ, satisfies

$$\lambda^2 = (\nu' D\nu)^2 = O_p(1/n). \tag{50}$$

Bearing in mind that D is an $m \times m$ symmetric matrix and $m = O(n^{1-\epsilon})$ we have

$$tr\, D^2 = \sum_{k=1}^{m}(eig\, D)_k^2 \to 0. \tag{51}$$

It follows from (18) that $\widetilde{\mathbf{Z}}_n^{(m)} \approx \mathbf{Z}_n^*$, say, where $\overline{\Theta}^{(m)}$ is $m-$dimensional normal with mean $\overline{\Phi}^{(m)}$ and covariance matrix

$$\frac{1}{m}\sum \frac{1}{\overline{h}_k}\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})'.$$

The expression (37) then shows that $A_n^{(m)}$ is positive semidefinite and yields (35). This completes the proof when Assumption B holds.

In case Assumption B fails we take $m(n) = O(n^{1/2-\beta})$ for some sufficiently small $\beta > 0$. The equivalence reduction to $\overline{h}, \overline{f}$ is still valid, since, now, $\alpha > 1$. We now use the fact that $\varphi_j(x) = O(\sqrt{j})$ uniformly in j, x by (8). The reasoning at (49) then yields for an individual eigenvalue

$$\lambda^2 = O_p(\frac{m}{n}) \tag{52}$$

in place of (38). Thus,

$$tr\, D^2 \to 0 \tag{53}$$

since now D is $m \times m$ with $m = O(n^{1/2-\beta})$. The proof can now be completed as before. □

5.3 The converse; full equivalence

The main result in Theorem 5.1 shows how to begin with the nonparametric regression and arrive at the series problem $\mathbf{Z}_n^{(m)}$ with no asymptotic loss of effectiveness. Two further results are needed in order to guarantee suitability of this process. The asymptotic mapping

$$\{\mathbf{Z}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n\} \tag{54}$$

guarantees that knowledge of $\{\mathbf{Z}_n^{(m)}\}$ is asymptotically as effective as knowledge of $\{\mathbf{Z}_n\}$. And the mapping

$$\{\mathbf{Z}_n\} \rightsquigarrow \{X_i, Y_i : \, i = 1, \ldots, n\} \tag{55}$$

completes the circle by establishing that no asymptotic capability has been lost in the transition $\{X_i, Y_i\} \rightsquigarrow \{\mathbf{Z}_n\}$.

In particular, Theorem 5.1 and (54) and (55) means that if an asymptotic minimax result is desired in a problem with observations $\{X_i, Y_i\}$ then one may solve this problem using either $\{\mathbf{Z}_n^{(m)}\}$ or $\{\mathbf{Z}_n\}$. The asymptotic minimax solution derived in $\{\mathbf{Z}_n^{(m)}\}$ or in $\{\mathbf{Z}_n\}$ will then have the same asymptotic minimax property for $\{X_i, Y_i\}$ because of (54), (55).

The maps in (54) and (55) can be explicitly described, although they are somewhat harder to implement than those in (34). Further the main impact of (54) and (55) is to guarantee no asymptotic loss of effectiveness in the maps (34); we doubt (1) and (2) will need to often be explicitly implemented. Hence we only state the main converse theorem here but defer the description of these maps to Section 7, along with the proof.

Additional regularity conditions on $\{\varphi_j\}$ seem to be needed. In order to allow for wide generality we state an abstract technical form of these assumptions here. In Lemmas 7.3 and 7.4 we show these assumptions are satisfied by Fourier basis under a very mild condition on h. It is easy to check they are also satisfied by the Haar wavelet basis under a similar condition on h. We believe they are also satisfied for a wide range compactly supported wavelet bases under the conditions of Theorem 5.1 so long as h^{-1} is sufficiently smooth.

The proof in Section 7, as well as the technical assumption below requires introduction of a sequence $m^* = m^*(n)$ that satisfies the assumptions of Theorem 4.1, including (23), and also

$$\frac{m^*(n)}{m(n)} \to 0. \tag{56}$$

Let $M = \{M_{ij}\}$ be as in (28). We need to assume

$$\{\sum_{j=(m+m^*)/2}^{\infty} \sum_{i=1}^{m^*} + \sum_{j=m}^{\infty} \sum_{i=1}^{(m+m^*)/2} \} M_{ij}^2 \to 0 \tag{57}$$

We also need to assume

$$\{\bar{\Phi}_k : k = 1, \ldots, m \text{ is a linearly independent set of vectors } \} \tag{58}$$

for every sufficiently large n. As noted, Assumptions (57) and (58) are easily checked for Fourier bases and they seem also to be true for all the familiar wavelet and orthogonal polynomial bases.

Theorem 5.2 *Assume H is known. Let the assumptions of Theorem 5.1 be satisfied. Then (57) implies (54) and (58) implies (55). The randomized mapping corresponding to (54) is described in (100). The mapping corresponding to (55) is also explicitly described in the proof of the theorem in Section 7.*

5.4 Main Theorem, Unknown H

Only the final step (iii), of the equivalence construction involves knowledge of $h = H'$. This is needed in order to define the covariance matrix $\mathbf{M}$ that appears in the definition (46) of the auxiliary variable V_n. Under suitable assumptions it is possible to replace the true $\mathbf{M}$ by an estimated value, $\hat{\mathbf{M}}$, and still achieve the conclusion of Theorem 5.1.

For the results here we have found it necessary to assume h is known to lie in a Hölder ball of suitable smoothness. The smoothness assumption on h is thus somewhat more restrictive than that on f. This seems to be appropriate for most applications, and so this assumption seems to us not to be a significant constraint.

Definition: Let $\langle\gamma\rangle = \max\{k : k \text{ an integer}, k \geqslant 0, k < \gamma\}$. The Hölder ball of smoothness γ and radius B on $[0,1]$ is defined as the set of continuous functions, f, possessing continuous derivatives of order $\langle\gamma\rangle$ that satisfy $|f(y)| < B$ and

$$\left| f(y) - \sum_{j=0}^{\langle\gamma\rangle} f^{(j)}(x)\frac{(y-x)^j}{j!} \right| \leqslant B|y-x|^{\langle\gamma\rangle+1} \tag{59}$$

for all $x, y \in [0,1]$. Let $H_\gamma(B)$ denote this set. Note that $H_\gamma(B) \subset S_\gamma(B)$ when $\{\varphi_j\}$ is a Fourier basis.

When $\gamma > 1/2, \epsilon > 0$ and B are given and h is known to satisfy $h \in H_\gamma(B)$ and $h \geqslant \epsilon > 0$ (as in (1)) then it is possible to use $X_1, \ldots, X_n$ to construct an estimate $\hat{h}_n$ satisfying $\hat{h}_n \in H_\gamma(2B)$ and for any $\beta > 0$

$$E(\sup_{x\in[0,1]} (\hat{h}_n(x) - h(x))^2) = O(n^{-\frac{2\gamma}{(1+2\gamma)}+\beta}) \tag{60}$$

uniformly for $h \in H_\gamma(B)$. For example, one may use a standard kernel estimator with bandwidth the order of $n^{-1/(1+2\gamma)}$. See, for example, Fan and Gijbels (1996).

Define

$$\hat{\mathbf{M}} = \int_0^1 \Phi(t)\Phi'(t)\frac{1}{\hat{h}(t)}\,dt \tag{61}$$

and define $\hat{V}_n \sim N_,(0, \hat{A}_n^{(m)}/n)$ where

$$\hat{A}_n^{(m)} = \left[\hat{\mathbf{M}}^{(m)} - \int \frac{1}{h(t)}\overline{\Phi}^{(m)}(t)(\overline{\Phi}^{(m)}(t))'\,dt\right]_+ = \left[\hat{\mathbf{M}}^{(m)} - \frac{1}{m}\sum_{k=1}^{m}\frac{1}{\overline{h}_k}\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})'\,dt\right]_+. \tag{62}$$

(Here A_+ denotes the positive definite part of A. Thus if $A = Q'DQ$ with Q orthogonal and D diagonal then $A_+ = Q'D_+Q$ where $(D_+)_{ii} = (D_{ii})_+$.)

The analog of Theorem 5.1 is as follows. Note that that conditions below on h and on α are somewhat more restrictive than in the earlier theorem.

Theorem 5.3 *Let $\gamma > 1/2, \epsilon > 0, B$ be given. About H, assume it is known only that $h \in H_\gamma(B)$. Make the other assumptions of Theorem 5.1, including the assumptions about $m(n)$. Assume also* *that*

$$m(n) = O(n^{2\gamma/(2\gamma+1)-\beta}) \tag{63}$$

for some $\beta > 0$*. Then*

$$\{X_i, Y_i : i = 1, \ldots, n\} :\longrightarrow \{\hat{h}_n, \widetilde{\mathbf{Z}}_n\} :\longrightarrow \{\hat{h}_n, \widetilde{\mathbf{Z}}_n\} \approx \{\hat{h}_n, \mathbf{Z}_n^{(m)}\} \rightsquigarrow \{\mathbf{Z}_n^{(m)}\}. \tag{64}$$

The final step in (34) is given more precisely by

$$\{\widetilde{\mathbf{Z}}_n^{(m)} + \hat{V}_n^{(m)}\} \approx \{\mathbf{Z}_n^{(m)}\}. \tag{65}$$

with $\hat{V}_n^{(m)}$ *as in (62).*

Remark: If $\alpha = \gamma$ then $m(n)$ satisfying the conditions of the theorem exists if and only if

$$\alpha = \gamma > (1 + \sqrt{5})/4 \tag{66}$$

If $\alpha > 1$, as required when Assumption B is not satisfied then since $\gamma > 1/2$ condition (5.3) is automatically satisfied.

Remark: It is easy to check that one may also use $\bar{\hat{h}}$ in defining $\hat{\mathbf{M}}$. Thus, if it is more convenient one may use $\overline{\hat{\mathbf{M}}}$ in place of $\hat{\mathbf{M}}$ in the above formulas, where

$$\overline{\hat{\mathbf{M}}} = \frac{1}{m} \sum \frac{1}{\bar{\hat{h}}_k} \overline{\Psi}_k^{(m)}$$

with $\overline{\Psi}$ as in (40).

Proof: Theorem 5.1 shows that $\{\widetilde{\mathbf{Z}}_n^{(m)} + V_n^{(m)}\} \approx \{\mathbf{Z}_n^{(m)}\}$. So Theorem 5.3 will be proved when it is shown that $\{\widetilde{\mathbf{Z}}_n^{(m)} + \hat{V}_n^{(m)}\} \approx \{\widetilde{\mathbf{Z}}^{(m)} + V_n^{(m)}\}$. According to (18) it will suffice to show

$$tr\,(\mathbf{I}^{(m)} - (\mathbf{M}^{(m)})^{-1}\hat{\mathbf{M}}^{(m)})^2 \to 0 \text{ in probability} \tag{67}$$

uniformly over $h \in H_\gamma(B)$. The eigenvalues of $\mathbf{M}^{(m)}$ are uniformly bounded away from 0 and ∞ since $\int(\Phi(t)\Phi'(t))\,dt = \mathbf{I}$ and since $\frac{1}{B} \leqslant \frac{1}{h(t)} \leqslant \frac{1}{\epsilon}$ for some $\epsilon > 0$. Hence (67) will follow if (and only if)

$$tr\,(\mathbf{M}^{(m)} - \hat{\mathbf{M}}^{(m)})^2 \to 0 \text{ in probability} \tag{68}$$

uniformly over $h \in H_\gamma(B)$.

Let ν be a unit vector in $\mathbf{R}^{(m)}$. Then,

$$\begin{aligned}
\nu'(\mathbf{M}^{(m)} - \hat{\mathbf{M}}^{(m)})\nu &= \int \left(\frac{1}{h(t)} - \frac{1}{\hat{h}_n(t)} \right) \nu'\Phi(t)\Phi'(t)\nu \, dt && (69)\\
&\leqslant \sup_{x\in[0,1]} \left| \frac{1}{h(x)} - \frac{1}{\hat{h}_n(x)} \right| \int \nu'\Phi(t)\Phi'(t)\nu \, dt && (70)\\
&= O_p(n^{-\frac{\gamma}{2\gamma+1} + \beta/3}) && (71)
\end{aligned}$$

uniformly over $h \in H_\gamma(B)$.

It follows upon using (63) that

$$tr\,(\mathbf{M}^{(m)} - \hat{\mathbf{M}}^{(m)})^2 = O_p(m\, n^{-\frac{2\gamma}{2\gamma+1} + 2\beta/3}) = O_p(n^{-\frac{\beta}{3}}) \to 0,$$

as desired. □

6 Asymptotic Relative Sufficiency

When Assumption B is not satisfied we have needed to require $\alpha > 1$ in Theorems 4.1, 5.1 and 5.3. This requirement may be undesirably strong for some applications. The following theorem proves a result involving only the requirement $\alpha > 1/2$. This result is formally weaker than that in the earlier theorems but should serve nearly as well in practice.

The notion of Pitman efficiency provides the motivation for the formulation that follows, as well as suggesting the terminology we use. That notion compares two sequences $\{\varphi_{1,n}\}$ and $\{\varphi_{2,n}\}$ of level α statistical tests for a particular testing problem. $\{\varphi_{1,n}\}$ is asymptotically as efficient as $\{\varphi_{2,n}\}$ if $\varphi_{1,n}$ has asymptotically the same (or greater) power than $\varphi_{2,n*}$ at all local alternatives where $n^* = n^*(n)$ may have $n^* < n$ but satisfies $\lim n^*/n \to 1$.

Corresponding to this idea we consider two sequences $\{V_n\}$ and $\{W_n\}$ as in Section 3.4. We say $\{V_n\}$ is "asymptotically relatively sufficient" for $\{W_n\}$ if $\{V_n\} \rightsquigarrow \{W_{n^*}\}$ for some sequence $n^* = n^*(n)$ with

$$\limsup_{n\to\infty} \frac{n^*(n)}{n} \geqslant 1. \tag{72}$$

In practice one often has a sequence of procedure $\{\delta_n\}$ based on W_n whose risks $R(f, \delta_n)$ (based on uniformly bounded losses $\{L_n\}$) satisfy

$$\limsup_{n\to\infty} \sup_{f\in\mathscr{F}} \frac{R(f,\delta_n)}{R(f,\delta_{n*})} \leqslant 1. \tag{73}$$

If this is the case and if $\{V_n\}$ is relatively asymptotically sufficient for $\{W_n\}$ then we can construct a procedure δ_n^* from V_n according to $\delta_n^* = \delta_{n*}(\nu_n(V_n))$ where ν_n is the (randomized) sufficiency map from V_n to W_{n^*}, and

$$\limsup_{n\to\infty} \sup_{f\in\mathscr{F}} \frac{R(f,\delta_n)}{R(f,\delta_{n*})} \to 1, \tag{74}$$

no matter what are δ_n and L_n.

We state the following result only for the case where H is known. An analogous result holds when H is unknown as in Section 5.3.

Theorem 6.1 *Assume H is a known function satisfying (33). Assume (19) and (23) hold and $\mathscr{F} \subset S_\alpha(B)$ for some fixed known values of α, B. Then there is an $n^* = n^*(n) \leqslant n$ with*

$$\lim \frac{n^*(n)}{n} = 1 \tag{75}$$

such that

$$\{X_i, Y_i : i = 1, \ldots, n\} :\to \widetilde{\mathbf{Z}}_n^{(m)} \sim \{\mathbf{Z}_{n*}^{(m)}\}. \tag{76}$$

Hence $\{X_i, Y_i : i = 1, \ldots, n\}$ is asymptotically relatively sufficient for $\{\mathbf{Z}_n^{(m)}\}$. The randomized mapping yielding (76) is described in (79) below.

Proof: Through the assertion (46) the proof is the same as that of Theorem 5.1. Then choose $n^* < n$ but satisfying (75) so that

$$\lim_{n\to\infty} P\{\inf_{1\leqslant k\leqslant m} mN_k > n^*\,\overline{h}_k\} = 1. \tag{77}$$

(It suffices to choose $n^* = n - n^{1/2+\epsilon}$ for any $0 < \epsilon < 1/2$.) Next, define

$$A_n^* = \left[\mathbf{M}^{(m)} - \frac{1}{m}\sum \frac{1}{\overline{h}_k}\overline{\Phi}_k^{(m)}(\overline{\Phi}_k^{(m)})'\right]_+ . \tag{78}$$

It is easy to show using (77) that $P(A_n^* \text{ is positive definite }) \to 1$. (In other words, with probability tending to 1 the "positive part" symbol is unnecessary in (78).) Let

$$V_n^* \sim N_m(0, A_n^*/n^*). \tag{79}$$

Then $\widetilde{Z}_n^{(m)} + V_n^*$ has the same distribution as $\mathbf{Z}_{n^*}^{(m)}$. This proves the final assertion of (76). □

7 Proofs

The following two lemmas verify assertions made in Section 4.1.

Lemma 7.1 *Assume $\{\varphi_j\}$ is the cosine basis of Section 2.4 or the usual Fourier sine-cosine basis. Assume $\mathscr{F} \subset S_\alpha(B)$ for some $\alpha > 1/2$. Then (23) is satisfied whenever*

$$\frac{m(n)}{n^{\frac{1}{2\alpha}}} \longrightarrow \infty. \tag{80}$$

Remark: Since $\alpha > 1/2$ sequences exist satisfying both (19) and (80). With further preparation this lemma could be generalized to apply to a variety of other useful bases related to the Fourier basis.

Proof: To be specific we consider the cosine basis. (The proof for the sine-cosine basis is similar.) Direct calculation yields for $2 \leqslant j < l$ and $x \in I_k$

$$\begin{aligned} \overline{\varphi}_j(x)\overline{\varphi}_l(x) &= \frac{m^2}{8\pi^2 jl}\left[\sin(\frac{\pi k(j-1)}{m})(1-\cos\frac{\pi(j-1)}{m}) + \cos(\frac{\pi k(j-1)}{m})\sin(\frac{\pi(j-1)}{m})\right] \\ &\cdot \left[\sin(\frac{\pi k(l-1)}{m})(1-\cos\frac{\pi(l-1)}{m}) + \cos(\frac{\pi k(l-1)}{m})\sin(\frac{\pi(l-1)}{m})\right] \end{aligned} \tag{81}$$

It follows that for $2 \leqslant j < l$

$$\int \overline{\varphi}_j(t)\overline{\varphi}_l(t)dt = 0 \tag{82}$$

since

$$\sum_{k=1}^{m} \sin(\frac{\pi k(j-1)}{m})\cos(\frac{\pi k(l-1)}{m}) = 0$$

and similarly for the other relevant sums over k.

Hence, for $j \neq l$

$$\int (\varphi_j(t) - \overline{\varphi}_j(t))(\varphi_l(t) - \overline{\varphi}_l(t))dt = 0 \tag{83}$$

We can then write

$$\begin{aligned} n\int (f(t) - \overline{f}(t))^2 dt &= n\int \left[\sum_{j=1}^{\infty} \theta_j(\varphi_j(t) - \overline{\varphi}_j(t))\right]^2 dt \\ &= \sum_{j=1}^{\infty} \theta_j^2 \int (\varphi_j(t) - \overline{\varphi}_j(t))^2 dt. \end{aligned} \tag{84}$$

For $1 \leqslant j \leqslant m$ write

$$\begin{aligned}\int (\varphi_j(t) - \overline{\varphi}_j(t))^2 dt &< \frac{1}{m^2} \int |\varphi_j'(t)|^2 dt \\ &< \frac{j^2}{m^2},\end{aligned} \tag{85}$$

and for $j \geqslant m+1$

$$\int (\varphi_j(t) - \overline{\varphi}_j(t))^2 dt \leqslant \int \varphi_j^2(t) dt = 1.$$

Combining these simple facts with (84) and using $\mathscr{F} \subset S_\alpha(B), \alpha > 1/2$ and (80) yields

$$\begin{aligned} n \int (f(t) - \overline{f}(t))^2 dt &= \frac{n}{m^2} \sum_{j=1}^{m} j^2 \theta_j^2 + n \sum_{j=m+1}^{\infty} \theta_j^2 \\ &\leqslant \frac{n}{m^{2\alpha}} \sum_{j=1}^{m} j^{2\alpha} \theta_j^2 + \frac{n}{m^{2\alpha}} \sum_{j=m+1}^{\infty} j^{2\alpha} \theta_j^2 \\ &\leqslant \frac{n}{m^{2\alpha}} B \longrightarrow 0. \end{aligned} \tag{86}$$

This verifies (23). □

The next result applies to compactly supported wavelet bases. The Haar basis has special, convenient properties. Otherwise we need to assume that the basis functions are differentiable and satisfy

$$\int (\varphi_j'(t))^2 dt \leqslant c j^2 \quad j = 1, \ldots, \tag{87}$$

for some $c < \infty$. For wavelet bases presented in the double index subscript system of Section 2.3 this condition will be satisfied so long as

$$\int (\varphi'(t))^2 dt < \infty.$$

Lemma 7.2 *Let $\{\varphi_j(t)\}$ correspond to a compactly supported wavelet basis. Assume that either $\{\varphi_j(t)\}$ is the Haar basis or that (87) is satisfied. Suppose $\mathscr{F} \subset S_\alpha(B)$ for some $\alpha > 1/2$. Then (23) satisfied whenever*

$$\frac{m(n)}{n^{1/2\alpha} \log n} \longrightarrow \infty \tag{88}$$

and $m(n) = 2^{k(n)}$ for some $k = 1, \ldots$.

Proof: For the Haar basis and $m(n) = 2^{J_1}$, a power of 2, we write (as in Brown et.al. (1999))

$$n \int (f(t) - \overline{f}(t))^2 dt = n \sum_{j=m+1}^{\infty} \theta_j^2 \leqslant \frac{n}{m^{2\alpha}} \sum_{j=m+1}^{\infty} j^{2\alpha} \theta_j^2 \leqslant \frac{Bn}{m^{2\alpha}} \longrightarrow 0. \tag{89}$$

For other wavelet bases we need to write expressions in the double index subscript form.

$$\int (f(t)-\overline{f}(t))^2 dt = \left(\sum_{j=J_0} \theta_{J_0,k}(\varphi_{J_0,k} - \overline{\varphi}_{J_0,k}(t)) + \sum_{j=J_0+1}^{J_1} \sum_{k=0}^{2^{j-1}-1} \theta_{j,k}(\psi_{j,k}(t) - \overline{\psi}_{j,k}(t)) \right)^2 \quad (90)$$

$$\leqslant J_1 \left[(\sum_{j=J_0} \theta_{J_0,k}(\varphi_{J_0,k} - \overline{\varphi}_{J_0,k}(t)))^2 + \sum_{j=J_0+1}^{J_1} (\sum_{k=0}^{2^{j-1}-1} \theta_{j,k}(\psi_{j,k}(t) - \overline{\psi}_{j,k}(t)))^2 \right]$$

Since the wavelets are compactly supported there is a constant D such that $|k'-k| \geqslant D$ implies

$$(\psi_{j,k}(t) - \overline{\psi}_{j,k}(t))^2 (\psi_{j,k'}(t) - \overline{\psi}_{j,k'}(t))^2 = 0 \ \forall \ t. \quad (91)$$

Using (87) as in (85) we then have

$$\int (\sum_{k=0}^{2^{j-1}-1} \theta_{j,k}(\psi_{j,k}(t) - \overline{\psi}_{j,k}(t))^2 dt \leqslant 2D \sum_{k=0}^{2^{j-1}-1} \theta_{j,k}^2 (\psi_{j,k}(t) - \overline{\psi}_{j,k}(t))^2 dt$$

$$\leqslant 2DC \sum_{k=0}^{2^{j-1}-1} \frac{2^{2j}}{m^2} \theta_{j,k}^2. \quad (92)$$

Combining the above and using the assumption that $\mathscr{F} \subset S_\alpha(B)$ as at (86) yields

$$n \int (f(t)-\overline{f}(t))^2 dt \leqslant 4nDC(J_1 \sum_{j=J_0}^{J_1} 2^j \sum_{k=0}^{(2^{j-1}-1)\vee(2^{J_0-1}-1)} \theta_{j,k}^2 + \sum_{j=J_1+1}^{\infty} \sum_{k=0}^{2^{j-1}-1} \theta_{j,k}^2)$$

$$= O(\frac{J_1 n}{m^{1/2\alpha}}) \longrightarrow 0 \quad (93)$$

uniformly for $f \in \mathscr{F}$, since $J_1 = O(\log n)$. □

Remark: A variant of the above arguments can yield a (weaker) result for more general bases. Let $\{\varphi_j\}$ be any basis for which φ_j are differentiable and satisfy (87). Suppose $\mathscr{F} \subset S_\alpha(B)$ for some $\alpha > 1$. Then (23) is satisfied whenever

$$\frac{m(n)}{n^{1/\alpha}} \longrightarrow \infty. \quad (94)$$

The argument does not use the orthogonality in (83) or (91) but instead directly uses the Cauchy-Schwartz inequality to write

$$n \int (f(t)-\overline{f}(t))^2 dt \leqslant 2n\{m \sum_{j=1}^{m} \theta_j^2 \int (\varphi_j(t) - \overline{\varphi}_j(t))^2 dt + \sum_{j=m+1}^{\infty} \theta_j^2\}$$

$$= O(\frac{m(n)}{n^{1/\alpha}}) \longrightarrow 0. \quad (95)$$

We conclude with the proof of Theorem 5.2 and a discussion of its regularity conditions.

The following two lemmas prove assertions made in Section 5.3.

Lemma 7.3 *Assume $\{\varphi_j\}$ is as in Lemma 7.1. Assume $h^{-1} \in S_\alpha(B)$, $\alpha > 1/2$, for this basis. Let m, m^* each satisfy (23) for some $\alpha > 1/2$ and also (56). Then (57) is satisfied.*

Proof: Again, take $\{\varphi_j\}$ to be the Fourier cosine basis as defined above (7). Note that for $i, j \geqslant 2$

$$\int_0^1 \varphi_i(t)\varphi_j(t)\varphi_k(t)dt = \begin{cases} \sqrt{2}/2 & \text{if } k-1 = |j-i| \neq 0 \text{ or } k-1 = (i-1)+(j-1) \\ 0 & \text{otherwise} \end{cases} \tag{96}$$

Analogous expressions when $\min(i, j) = 1$ or $i = j$ are easily derived.

Since $h^{-1} \in S_\alpha(B)$ we can write

$$h^{-1}(x) = \sum_{k=1}^{\infty} \beta_k \varphi_k(x)$$

where

$$\sum k^{2\alpha} \beta_k^2 < \infty. \tag{97}$$

Then for $i < j$

$$\begin{aligned} M_ij &= \int (\sum_{k=1}^{\infty} \beta_k \varphi_k(x))\varphi_i(x)\varphi_j(x)dx \\ &= \begin{cases} (1/\sqrt{2})(\beta_{j-i} + \beta_{i+j-1}) & \text{if } i \geqslant 2 \\ \beta_k & \text{if } i = 1 \end{cases} \end{aligned} \tag{98}$$

This yields

$$\begin{aligned} \sum_{i=1}^{m^*} \sum_{j=(m+m^*)/2}^{\infty} M_{ij}^2 &\leqslant 2\sum_{i=1}^{m^*} \sum_{j=(m+m^*)/2}^{\infty} (\beta_{j-i}^2 + \beta_{i+j-1}^2) \\ &\leqslant 4\sum_{i=1}^{m^*} \sum_{j=(m+m^*)/2}^{\infty} \frac{1}{(\frac{m+m^*}{2} - i)^{2\alpha}} (j-i)^{2\alpha} \beta_{j-i}^2 \\ &= O(\sum_{i=1}^{m^*} \frac{1}{(m-m^*)^{2\alpha}}) \\ &= O(\frac{m^*}{(m-m^*)^{2\alpha}}) \to 0 \end{aligned} \tag{99}$$

since $\alpha > 1/2$. The other double sum in (56) is handled similarly. $\square$.

Lemma 7.4 *Assume $\{\varphi_j\}$ is as in Lemma 7.1. Then the $\{\overline{\Phi}_k : k = 1, \ldots, m\}$ are orthogonal and non-zero. Hence, in particular, they are linearly independent and so satisfy (57).*

Proof : The orthogonality follows from (79). A direct computation shows they are all non-zero, since $m/n \to 0$. $\square$

Proof of Theorem 5.2:

We first prove (54). Partition Z_n as $(Z_n) = (Z'_{n(1)}, Z'_{n(2)}, Z'_{n(3)})$. $\mathbf{Z}_{n(1)}$ is m^* dimensional and $\mathbf{Z}_{n(2)}$ is $(m - m^*)$ dimensional. Similarly write $\mathbf{Z}_n^{(m)} = (\mathbf{Z}_{n(1)}^{(m)\,\prime}, \mathbf{Z}_{n(2)}^{(m)\,\prime})$. In an analogous fashion partition M as

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \tag{100}$$

where M_{11} is $m^* \times m^*$ and M_{22} is $(m - m^*) \times (m - m^*)$, and the remaining blocks are determined from this. Given $Z_n^{(m)}$ define $Z_n^0 = (Z_{n(1)}^{(0)\,\prime}, Z_{n(2)}^{(0)\,\prime}, Z_{n(3)}^{(0)\,\prime})$ by

$$\mathbf{Z}_{n(1)}^0 = \mathbf{Z}_{n(1)}^{(m)},\ \mathbf{Z}_{n(2)}^0 = \mathbf{Z}_{n(2)}^{(m)} \tag{101}$$

and

$$\mathbf{Z}_{n(3)}^0 = V + M_{32} M_{22}^{-1} \mathbf{Z}_{n(2)}^{(m)}$$

with

$$V \sim N(0, M_{33} - M_{32} M_{22}^{-1} M_{23}).$$

As a consequence of (23) it suffices as in (26) to prove asymptotic equivalence when

$$E(\mathbf{Z}_{n(2)}) = E(\mathbf{Z}_{n(2)}^{(m)}) = 0 = E(Z_{n(3)}).$$

We henceforth assume (101) holds. Then

$$E(Z_n^0) = E(\mathbf{Z}_n). \tag{102}$$

Let m^0 denote the infinite dimensional covariance matrix of Z^0 and partition M^0 analogous to (100). Then

$$M_{(ij)}^0\ M_{(ij)} \text{ except for } i = 1, j = 3 or i = 3, j = 1.$$

$$M_{(13)}^0 = M_{(12)} M_{(22)}^{-1} M_{(23)} = M'_{(31)}.$$

It follows from this as in (62) that (uniformly in n)

$$\begin{aligned} \mathrm{tr}(I - M^{-1}M^0)^2 &= O(\mathrm{tr}(M - M^0)^2) \\ &= O(\mathrm{tr}(M^0_{(13)}M^0_{(31)} + M_{(31)}M_{(13)})). \end{aligned} \tag{103}$$

Now,

$$\mathrm{tr}(M_{(31)}M_{(13)}) = \sum_{i=1}^{m^*}\sum_{j=m}^{\infty} M_{ij}^2 \to 0 \tag{104}$$

by (57).

The eigenvalues of $M^{-1}_{(22)}$ are bounded away from 0, as has already been exploited in (104). Hence

$$\begin{aligned} \mathrm{tr}(M^0_{(13)}M^0_{(31)}) &= \mathrm{tr}(M^0_{(12)}M^{-1}_{(22)}M_{(23)}M_{(32)}M^{-1}_{(22)}M_{(21)}) \\ &= O(\mathrm{tr}(M^0_{(12)}M^0_{(23)}M^0_{(32)}M^0_{(21)})). \end{aligned} \tag{105}$$

Finally,

$$\mathrm{tr}(M^0_{(12)}M^0_{(23)}M^0_{(32)}M^0_{(21)}) = 2\sum_{i=1}^{m^*}\sum_{j=m^*+1}^{m}\sum_{k=m+1}^{\infty} M_{ij}^2 M_{jk}^2 \tag{106}$$

$$O\left\{\sum_{i=1}^{m^*}\sum_{j=\frac{m+m^*}{2}+1}^{m} M_{ij}^2 + \sum_{j=m^*+1}^{\frac{m+m^*}{2}}\sum_{k=m+1}^{\infty} M_{jk}^2\right\} \tag{107}$$

since

$$\sum_{j=1, j\neq i}^{\infty} M_{ij}^2 = O(1). \tag{108}$$

(108) is a general result about positive definite matrices with eigenvalues uniformly bounded away from 0 and ∞.

Let $A = \begin{pmatrix} a_{(11)} & A_{(12)} \\ A_{(21)} & A_{(22)} \end{pmatrix}$. Then

$$0 < (A^{-1})_{11} = (a_{(11)} - A_{(12)}A^{-1}_{(22)}A_{(21)})^{-1}.$$

Furthermore $\alpha =$ maxeig $A^{-1}_{22} \leqslant ($ mineig $A)^{-1}$. Hence

$$0 < a_{(11)} - \alpha A_{12}A_{21} = a_{(11)} - \alpha\sum_{j\neq 1} a_{1j}^2 \leqslant \alpha^{-1} - \alpha\sum_{j\neq 1} a_{1j}^2.$$

(104) and (106) combined with the assumption (57) prove

$$\mathrm{tr}(I - M^{-1}M^0)^2 \to 0.$$

This, together with (102) yields (54), in view of (17).

The proof of (55) proceeds in three steps, with matching descriptions of the asymptotic equivalence mappings. We first show

$$\{\mathbf{Z}_n\} \rightsquigarrow \{\mathbf{Z}_n^*\} \approx \{\widetilde{\mathbf{Z}}_n^{(m)}\} \tag{109}$$

with $\mathbf{Z}_n^*$ as defined below (50).

Let

$$Z(t) = \sum_{j=1}^{\infty} \int_0^t (\mathbf{Z}_n)_j \varphi_j(x) dx.$$

$Z(t)$ is the solution to the stochastic differential equation

$$dZ(t) = f(t)dt + \frac{1}{\sqrt{nh(t)}} dB(t) \tag{110}$$

where $B(t)$ is standard Brownian motion. See, for example, Steele (2001). Let

$$\overline{\mathbf{Z}}_n^{(m)} = \int \overline{\Phi}^{(m)}(t) dZ(t).$$

Then $\overline{\mathbf{Z}}_n^{(m)}$ is normal with mean

$$E(\overline{\mathbf{Z}}_n^{(m)}) = \int \overline{\Phi}^{(m)}(t) f(t) dt = \overline{\Theta}^{(m)}$$

and

$$Cov(\overline{\mathbf{Z}}_n^{(m)}) = \int \overline{\Phi}^{(m)}(t) (\overline{\Phi}^{(m)}(t))' \frac{1}{nh(t)} dt = (M^{(m)} - A_n^{(m)})/n$$

with $M^{(m)}$ and $A_n^{(m)}$ as in the proof of Theorem 5.1. Hence $\overline{\mathbf{Z}}_n^{(m)}$ produced from $\mathbf{Z}_n$ has the same distribution as $\mathbf{Z}_n^*$ defined below (51). It is shown in the proof of Theorem 5.1 that $\{\mathbf{Z}_n^*\} \approx \{\mathbf{Z}_n^{(m)}\}$. This proves the assertion contained in (108). $\square$

The next step of (58) is to establish

$$\{\widetilde{Z}_n^{(m)}\} \rightsquigarrow \{\bar{Y}_k\} \tag{111}$$

To do this note that the formula in (21) corresponds to a linear mapping of $\{\overline{Y}_k : k = 1, \ldots, m\}$ to $\widetilde{Z}_n^{(m)}$. If the set of vectors $\{\overline{\Phi}_k : k = 1, \ldots, m\}$ are linearly independent as in (58) then this mapping is invertible, and hence $\{\widetilde{Z}_n^{(m)}\} :\longrightarrow \{\overline{Y}_k\}$ in a direct way. This yields (111).

Finally we need to establish that

$$\{\bar{Y}_k\} \rightsquigarrow \{(X_i, Y_i)\} \tag{112}$$

This assertion is proved in Brown and Low (1996) under the condition (22). The necessary randomization is described in detail these and in related to that used in Remark 5.2.

References

[1] Antoniadis, A., Gregoire, G. and McKeague, I. W. (1994). Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.* **89**, 1340-1353.

[2] Brown, L.D. and Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-98.

[3] Brown, L.D., Cai, T., Low, M.G. and Zhang, C. (1999). On asymptotic equivalence of white noise model and nonparametric regression with random designs. Technical report.

[4] Brown, L.D. and Zhang, C.H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Ann. Statist.* **26**, 279-287.

[5] Cai, T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898-924.

[6] Cai, T., Low, M. and Zhao, L. (2000). Sharp adaptive estimation by a blockwise method. Technical report.

[7] Donoho, D.L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22**, 238-270.

[8] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 1200-1224.

[9] Donoho, D.L. and Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26**, 879-921.

[10] Donoho, D.L. and Johnstone, I.M. (1999). Asymptotic minimaxity of wavelet estimators with sampled data. *Statistica Sinica.* **9**, 1-32.

[11] Donoho, D. L., Liu, R. C., MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **3**, 1416–1437.

[12] Efromovich, S. (1999). *Nonparametric Curve Estimation, Methods, Theory, and Applications*. Springer-Verlag, New York.

[13] Efromovich, S. and Pinsker, M. S. (1984). A self-training algorithm for nonparametric filtering. (Russian) *Avtomat. i Telemekh.* **11**, 58–65.

[14] Eubank, R. L. and Speckman, P. (1991). Convergence rates for trigonometric and polynomial-trigonometric regression estimators. *Statist. Probab. Lett.* **11**, 119-124.

[15] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall.

[16] Härdle, W.K. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.

[17] Johnstone, I. (2000). *Function Estimation in Gaussian Noise: Sequence Models*. Stanford University.

[18] Kuo, H. (1975). *Gaussian measures in Banach spaces*. Lecture Notes in Mathematics. Vol. 463. Springer-Verlag, Berlin-New York.

[19] Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.* **1**, 277-329.

[20] Le Cam, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Stat.* **35**, 1419-1455.

[21] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

[22] Loader, C. (1999). *Local Regression and Likelihood*. Springer-Verlag, New York.

[23] Mandelbaum, A. (1984). All admissible linear estimators of the mean of a Gaussian distribution on a Hilbert space. *Ann. Statist.* **4**, 1448–1466.

[24] Nadaraya, E.A. (1964). On estimating regression. *Theory Prob. Appl.* **10**, 186-190.

[25] Parzen, E. (1962). On the estimation of a probability density and mode. *Ann. Math. Statist.* **33**, 1065-1076.

[26] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832-835.

[27] Steele, M. (2001). *Stochastic calculus and financial applications.* Springer-Verlag.

[28] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM

[29] Watson, G.S. (1964). Smooth regression analysis. *Sankhyā*. Series A, **26**, 359-372, 1964

[30] Zhang, C. (2000). General empirical Bayes wavelet methods and exactly adaptive minimax estimation. Technical Report, Department of Statistics, Rutgers University.