

MINIMAXITY OF THE METHOD OF REGULARIZATION ON STOCHASTIC PROCESSES¹

BY KER-CHAU LI

Purdue University

The idea of smoothing-spline interpolation is generalized to propose an estimator for the mean function of a stochastic process. A minimax property of the proposed estimator is then demonstrated under the usual squared loss function.

1. Introduction. Suppose we observe a stochastic process $\{Y(t) : 0 \leq t \leq 1\}$ satisfying $Y(t) = f(t) + \varepsilon_t$, where f is a smooth function and $\{\varepsilon_t\}$ is a stochastic process with mean 0 and covariance structure known up to a constant. The problem is to estimate the true function f . Most literature on this subject assumes that f belongs to a finite dimensional subspace of the reproducing kernel Hilbert space (RKHS) generated by the error process $\{\varepsilon_t\}$, and the least squares estimators are commonly used. See, for example, Parzen (1961) for a Gauss-Markov type theorem.

In this paper, we consider the case where the finite dimensional model is contaminated by some small smooth quantity. For example, if $\{\varepsilon_t\}$ is the first order autoregressive (continuous parameter) Gaussian process (Parzen, 1961) and if \mathbf{H} is the set of real functions on $[0, 1]$ with absolutely continuous first derivative and square integrable second derivative, then we might assume that $f \in \mathbf{H}$ and $\int_0^1 f''(x)^2 dx < \delta^2$ for some known constant δ . This becomes a "nearly-linear" model (Sacks and Ylvisaker, 1978) and a minimax linear estimator can be thought of as a robust regression estimator. We find that given any $t \in [0, 1]$ under the squared loss function, the unique minimax linear estimator for estimating $f(t)$ is the solution function of the following minimization problem, evaluated at the point t :

$$(1.1) \quad \min_{f \in \mathbf{H}} \frac{1}{\delta^2} \int_0^1 \{f''(x)\}^2 dx + (f, f)_K - 2(f, Y)_K,$$

where $(\cdot, \cdot)_K$ is the inner product of the RKHS generated by $\{\varepsilon_t\}$. Note that although most often Y does not belong to the RKHS, $(f, Y)_K$ is a well-defined random variable (Parzen, 1961).

The minimization problem (1.1) is similar to the following more famous minimization problem which yields a smoothing spline \hat{f} :

$$(1.2) \quad \min_{f \in \mathbf{H}} \int_0^1 \{f''(x)\}^2 dx + \frac{\delta^2}{\sigma^2} \sum_{i=1}^n \{Y_i - f(x_i)\}^2,$$

where the Y_i are the observations made at the $x_i \in [a, b]$, assumed to be independent with the same variance σ^2 . A minimax justification for the use of smoothing splines for (1.2) is given by Speckman (1979). An earlier similar result on periodic splines is given by P. J. Laurent; see Micchelli and Wahba (1979). As indicated by Speckman, a minimax linear estimator for any bounded linear functional can be derived by the method of regularization (MOR), i.e., by operating the linear functional on the smoothing spline \hat{f} . Our method for

Received January 1981; revised January 1982.

¹ This research was sponsored by the National Science Foundation under grant No. MCS-78-25301, MCS-75-10376, and MCS-80-04022, and was completed in partial fulfillment of requirements for the Doctor of Philosophy degree at the University of California, Berkeley, Department of Statistics.

AMS 1970 subject classification. Primary 60G35, 62M10; secondary 62G35, 62J05.

Key words and phrases. Autoregressive processes, minimaxity, method of regularization, robustness, smoothing splines, spectrum, stochastic processes.

getting the desired results follows a variation of Speckman’s approach. In Section 2, we shall reformulate our problem in a generalized form to cover also the problem of Speckman. While weakening a crucial assumption in Speckman (1979) and still retaining the results given there, we find an important difference between the case (1.2) where the observation is finite dimensional and the case (1.1) where the observation is infinite dimensional. Although the MOR always works for finite dimensional observations, it does not work in some infinite dimensional cases. The example given in the second paragraph of this section is a case where the MOR works. But, if instead of assuming that $\{\varepsilon_t\}$ is a first order autoregressive process, we assume that $\{\varepsilon_t\}$ is a second order autoregressive process, then the MOR does not work. More precisely, although the minimax linear estimator for $f(t)$ exists, (1.1) does not have a solution. Moreover, for some cases, there exist two different error processes generating the same reproducing kernel such that the MOR works for one error process but not for the other, although the usual theory about linear estimators involves only the first and the second moments of the random errors. However, a theorem is given to characterize the conditions under which the method of regularization works. It is somehow surprising that although much work has been done on solving (1.2), there is little statistical literature on (1.1). Specifically, it is not known how to apply the cross-validation technique (Wahba and Wold, 1975) for estimating δ^2/σ^2 in (1.2) to the problem (1.1).

2. Main result. Let I be an index set, let H_1 and H_2 be Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ respectively, and suppose

(2.1) T is a bounded linear mapping from H_1 onto H_2 with finite dimensional null space.

This assumption, together with (2.4) below, is weaker than the Assumptions A2 and A3 in Speckman (1979).

Assume that ε is a stochastic process on the index set I with mean 0 and the covariance structure generating a reproducing kernel Hilbert space H_3 with the kernel K . As usual we use $(\cdot, \cdot)_K$ to denote the inner product of H_3 . Assume our observation Y is a stochastic process on I such that

(2.2)
$$Y = Af + \sigma \cdot \varepsilon,$$

where A is a bounded linear mapping from H_1 to H_3 , σ is a known positive constant, and

(2.3)
$$f \in H_1 \quad \text{and} \quad \langle Tf, Tf \rangle_2 \leq \delta^2$$

for some known constant δ . Assume also that

(2.4)
$$\text{if } Af = 0 \quad \text{and} \quad Tf = 0 \quad \text{then } f = 0.$$

Suppose y is a realization of Y such that

(2.5) the linear functional on H_1 defined by mapping an element g in H_1 to the corresponding realization of the random variable $(Ag, Y)_K$ is bounded.

We may write $\tilde{y} \in H_1$ to denote the representation of the above linear functional in H_1 . Thus we have

(2.6)
$$\langle g, \tilde{y} \rangle_1 = (Ag, y)_K.$$

LEMMA 2.1. Assume (2.1) and (2.4). Then $(\sigma^2/\delta^2) T^*T + A^*A$ is one-to-one, onto, and has a self-adjoint square root L which is invertible with the self-adjoint inverse L^{-1} .

PROOF. Let N be the null space of T . By (2.1), N is finite dimensional and thus is a closed subspace of H_1 . Let N^\perp be the orthogonal complement of N in H_1 . Let T_1 be the restriction of T to N^\perp . By (2.1) again, T_1 is one-to-one and onto, and thus $T_1^*T_1$ is one-to-one and onto.

Let \mathbf{A}_1 and \mathbf{A}_2 be the restrictions of \mathbf{A} to \mathbf{N}^\perp and \mathbf{N} , respectively. By (2.4), \mathbf{A}_2 is 1-1 and hence $\mathbf{A}_2^* \mathbf{A}_2$ is 1-1 and onto (because \mathbf{N} is of finite dimension). Therefore $(\sigma^2/\delta^2) \mathbf{T}_1^* \mathbf{T}_1 + \mathbf{A}_2^* \mathbf{A}_2$ is a 1-to-1, onto, and self-adjoint linear mapping on \mathbf{H}_1 . Since

$$(\sigma^2/\delta^2) \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A} = (\sigma^2/\delta^2) \mathbf{T}_1^* \mathbf{T}_1 + \mathbf{A}_2^* \mathbf{A}_2 + \mathbf{A}_1^* \mathbf{A}_1,$$

the rest of the proof follows immediately.

THEOREM 2.1. *Under (2.1) through (2.5), the minimization problem*

$$(2.7) \quad \min_{f \in \mathbf{H}_1} \left\{ \frac{\sigma^2}{\delta^2} \langle \mathbf{T}f, \mathbf{T}f \rangle_2 + \langle \mathbf{A}f, \mathbf{A}f \rangle_K - 2 \langle \mathbf{A}f, y \rangle_K \right\}$$

has a unique solution \hat{f} given by $((\sigma^2/\delta^2) \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A})^{-1} \tilde{y}$. Moreover, if (2.5) is not satisfied then the solution does not exist (infimum = $-\infty$).

PROOF. It is clear that, by (2.6) and Lemma 2.1,

$$\begin{aligned} \frac{\sigma^2}{\delta^2} \langle \mathbf{T}f, \mathbf{T}f \rangle_2 + \langle \mathbf{A}f, \mathbf{A}f \rangle_K - 2 \langle \mathbf{A}f, y \rangle_K &= \frac{\sigma^2}{\delta^2} \langle \mathbf{T}^* \mathbf{T}f, f \rangle_1 + \langle \mathbf{A}^* \mathbf{A}f, f \rangle_1 - 2 \langle f, \tilde{y} \rangle_1 \\ &= \langle \mathbf{L}f, \mathbf{L}f \rangle_1 - 2 \langle \mathbf{L}f, \mathbf{L}^{-1} \tilde{y} \rangle_1 \\ &= \langle \mathbf{L}f - \mathbf{L}^{-1} \tilde{y}, \mathbf{L}f - \mathbf{L}^{-1} \tilde{y} \rangle_1 - \langle \mathbf{L}^{-1} \tilde{y}, \mathbf{L}^{-1} \tilde{y} \rangle_1 \\ &\geq - \langle \mathbf{L}^{-1} \tilde{y}, \mathbf{L}^{-1} \tilde{y} \rangle_1. \end{aligned}$$

The equality is achieved when $f = \mathbf{L}^{-1} \cdot \mathbf{L}^{-1} \tilde{y} = ((\sigma^2/\delta^2) \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A})^{-1} \tilde{y}$. Hence the first part of the proof is complete, while the second is obvious. \square

Given $\ell \in \mathbf{H}_1$, under (2.1) through (2.4), we want to estimate $\langle \ell, f \rangle_1$. By Parzen (1961), any linear estimator with finite variance can be represented as $(\mathbf{e}, Y)_K$ for some $\mathbf{e} \in H_3$. The mean squared error is $E \{ \langle \ell, f \rangle_1 - (\mathbf{e}, Y)_K \}^2$, which is equal to $\{ \langle \ell, f \rangle_1 - (\mathbf{e}, \mathbf{A}f)_K \}^2 + \sigma^2 (\mathbf{e}, \mathbf{e})_K$. The following theorem characterizes the minimax linear estimator and suggests the method of regularization for computing it.

THEOREM 2.2. *Under (2.1) through (2.4), the minimax problem*

$$(2.8) \quad \min_{\mathbf{e} \in \mathbf{H}_3} \max_{\langle \mathbf{T}f, \mathbf{T}f \rangle_2 \leq \delta^2} \{ \langle \ell, f \rangle_1 - (\mathbf{e}, \mathbf{A}f)_K \}^2 + \sigma^2 (\mathbf{e}, \mathbf{e})_K.$$

has a unique solution \mathbf{e}_0 given by $\mathbf{A}((\sigma^2/\delta^2) \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A})^{-1} \ell$. Moreover, if y is a realization of Y satisfying (2.5), then

$$(\mathbf{e}_0, y)_K = \langle \ell, \hat{f} \rangle_1$$

where \hat{f} is the solution provided by Theorem 2.1.

PROOF. The first result follows by an algebraic computation similar to that carried out in the proof of Theorem 1 of Speckman (1979). The second result follows from the following computation, which uses (2.6) and Theorem 2.1,

$$\begin{aligned} (\mathbf{e}_0, y)_K &= \left(\mathbf{A} \left(\frac{\sigma^2}{\delta^2} \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A} \right)^{-1} \ell, y \right)_K = \left\langle \left(\frac{\sigma^2}{\delta^2} \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A} \right)^{-1} \ell, \tilde{y} \right\rangle_1 \\ &= \left\langle \ell, \left(\frac{\sigma^2}{\delta^2} \mathbf{T}^* \mathbf{T} + \mathbf{A}^* \mathbf{A} \right)^{-1} \tilde{y} \right\rangle_1 = \langle \ell, \hat{f} \rangle_1. \end{aligned}$$

Note that in the model (2.2), if we take $I = \{1, \dots, n\}$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ with $E\varepsilon_i \varepsilon_j = 0$ and $E\varepsilon_i^2 = 1$ for all $i \neq j$, then our model is reduced to the model in Speckman (1979). Observe also that in this reduced case, (2.7) becomes

$$\min_{f \in \mathbf{H}_1} \left\{ \frac{\sigma^2}{\delta^2} \langle \mathbf{T}f, \mathbf{T}f \rangle_2 + (\mathbf{A}f)' \mathbf{A}f - 2(\mathbf{A}f)' y \right\},$$

which is equivalent to the solution given in Speckman (1979):

$$\min_{f \in \mathbf{H}_1} \left\{ \frac{\sigma^2}{\delta^2} \langle \mathbf{T}f, \mathbf{T}f \rangle_2 + (\mathbf{A}f - y)'(\mathbf{A}f - y) \right\}.$$

As was demonstrated by Speckman, (1.2) is a special case of the above minimization problem. It is therefore from this point of view that we consider (1.1) as being proposed by using a generalized idea of smoothing spline interpolation. Note also that in (2.7) we cannot replace $(\mathbf{A}f, \mathbf{A}f)_K - 2(\mathbf{A}f, y)_K$ by $(\mathbf{A}f - y, \mathbf{A}f - y)_K$ because in most cases $(y, y)_K$ is $+\infty$.

EXAMPLE 1. Ridge regression on stochastic process. Take $\mathbf{H}_1 = R^n$, $\mathbf{H}_2 = R^m$ ($n \geq m$), \mathbf{T} to be any linear mapping from R^n onto R^m and \mathbf{A} to be any one-to-one linear mapping from R^n to the reproducing kernel Hilbert space \mathbf{H}_3 . It is easy to verify (2.1) and (2.4), and that (2.5) holds with probability 1. An element in R^n is usually called a vector of unknown parameters and (2.3) restricts the parameters to lie in an ellipsoid with center at the origin. Thus Theorem 2.2 provides a ridge estimator for any linear functional on the parameter space and claims also that the method of regularization always works. For the case where $m = n$ and \mathbf{T} is the identity mapping with $\delta = +\infty$, this reduces to the usual least squares estimator for which a Gauss-Markov type theorem has been well known (Parzen, 1961).

EXAMPLE 2. Robust regression on a stochastic process. Consider the example given in the second paragraph of Section 1. Let $\mathbf{I} = [0, 1]$. Take \mathbf{H}_1 to be \mathbf{H} and equip it with the Sobolev norm obtained from the inner product $\langle f, g \rangle_1 = \sum_{k=0}^2 \int_0^1 f^{(k)} g^{(k)}(x) dx$. Take \mathbf{H}_2 to be $L^2[0, 1]$ and define \mathbf{T} by $\mathbf{T}(f) = f''$. The RKHS \mathbf{H}_3 contains all differentiable functions on $[0, 1]$ (c.f. Parzen, 1961). The inner product of \mathbf{H}_3 is given by

$$(f, g)_K = \int_0^1 (f' + \beta f)(g' + \beta g) dt + 2\beta f(0)g(0),$$

where β is a constant.

Take \mathbf{A} to be the identity map. Then (2.1) through (2.4) hold. To verify that (2.5) holds with probability 1, we may write

$$(f, \varepsilon)_K = \frac{1}{2\beta} \left\{ \beta^2 \int_0^1 f(t)\varepsilon(t) dt - \int_0^1 \varepsilon(t)f''(t) dt + f'(1)\varepsilon(1) - f(0)\varepsilon(0) \right\} + \frac{1}{2} \{ f(0)\varepsilon(0) + f(1)\varepsilon(1) \},$$

where $f \in \mathbf{H}$. By the Cauchy-Schwartz inequality, we have

$$\int_0^1 f(t)\varepsilon(t) dt \leq \left(\int_0^1 f^2(t) dt \right)^{1/2} \left(\int_0^1 \varepsilon^2(t) dt \right)^{1/2} \leq \langle f, f \rangle_1^{1/2} \cdot \left(\int_0^1 \varepsilon^2(t) dt \right)^{1/2},$$

which shows that the linear functional on \mathbf{H}_1 which maps f to $\int_0^1 f(t)\varepsilon(t) dt$ is bounded. A similar argument applies to the linear functional on \mathbf{H}_1 which maps f to $\int_0^1 f''(t)\varepsilon(t) dt$. Finally, it is a well-known fact that $f'(1)$, $f(0)$ and $f(1)$ are bounded functionals on \mathbf{H}_1 . Therefore (2.5) holds with probability 1. Thus Theorem 2.1 and 2.2 hold. The method of regularization works in this case. Moreover, we have the following global optimality

$$\begin{aligned} \min_{\mathbf{e}(\cdot) \in \mathbf{H}_1} \max_{\int_0^1 f''(x)^2 dx \leq \delta^2} \max_{t \in [0, 1]} E \{ f(t) - (\mathbf{e}(t), Y)_K \}^2 \\ = \max_{\int_0^1 f''(x)^2 dx \leq \delta^2} \max_{t \in [0, 1]} E \{ f(t) - \hat{f}(t) \}^2, \end{aligned}$$

where \hat{f} is the solution provided by Theorem 2.1. This will be seen by changing the order of the two maximizing operations on both sides of the equality and applying the minimax result for point estimation.

Similar results hold for the case where f belongs to an m th order Sobolev space and ϵ is a m' th order autoregressive process with $m > m'$.

EXAMPLE 3. Assume that ϵ is a second order autoregressive Gaussian process and that other conditions are the same as in Example 2. We see that the probability that (2.5) holds is zero because of the facts that the RKHS \mathbf{H}_3 contains all twice differentiable functions on $[0, 1]$, \mathbf{H}_1 is homeomorphic to \mathbf{H}_3 , and Y does not belong to \mathbf{H}_3 . Hence the method of regularization does not work in this case. To avoid computation, we look at another similar model. Let \mathbf{H}_1 and \mathbf{H}_2 be equal to \mathbf{H}_3 given above. Let \mathbf{T} and \mathbf{A} be the identity map. Assume (2.2) and (2.3) hold. It is easy to verify that (2.1) and (2.4) hold. But the probability that (2.5) holds is zero, because ϵ , which is Gaussian, does not belong to \mathbf{H}_3 , i.e. $(\epsilon, \epsilon)_K = +\infty$; see Parzen (1961) for details. Hence the solution of (2.7) does not exist. However, the minimax linear estimator for $f(t)$ provided by Theorem 2.2 is easily identified to be $\delta^2/(\delta^2 + \sigma^2) \cdot Y(t)$.

Assume hereafter that \mathbf{H}_1 and \mathbf{H}_3 are separable. Assume also that

$$(2.9) \quad \text{the spectrum of } \mathbf{A}^* \mathbf{A} \text{ is discrete.}$$

(See Rudin (1973) for all related statements hereafter.)

Now we characterize the conditions on \mathbf{A} such that (2.5) holds with probability 1. By (2.2) it is clear that we only have to consider the case where $f = 0$. Let P be the probability that (2.5) holds. If \mathbf{H}_1 or \mathbf{H}_3 is of finite dimension, then $P = 1$. Therefore, we assume that both \mathbf{H}_1 and \mathbf{H}_3 are of infinite dimension. By (2.9), there exist orthonormal systems $\{\psi_i\}_{i=1}^\infty$ and $\{\phi_i\}_{i=1}^\infty$ on \mathbf{H}_1 and \mathbf{H}_3 respectively, such that $\{\psi_i\}_{i=1}^\infty$ is complete and $\mathbf{A}\psi_i = \lambda_i \phi_i$ for a sequence of numbers $\{\lambda_i\}_{i=1}^\infty$. The orthonormal systems $\{\psi_i\}_{i=1}^\infty$ are the eigenvectors of $\mathbf{A}^* \mathbf{A}$ with the corresponding eigenvalues $\{\lambda_i^2\}_{i=1}^\infty$. Note that (2.9) holds if \mathbf{A} is a compact linear operator. In particular, the operators in Examples 1 and 2 satisfy (2.9).

THEOREM 2.3. Under (2.2), $P = 1$ iff

$$(2.10) \quad \Pr \left\{ \sum_{i=1}^\infty \lambda_i^2 (\phi_i, \epsilon)_K^2 < \infty \right\} = 1.$$

In particular, if

$$(2.11) \quad \sum_{i=1}^\infty \lambda_i^2 < \infty,$$

then $P = 1$. Moreover, if $P = 1$, and $\{(\phi_i, \epsilon)_K\}_{i=1}^\infty$ are independent and identically distributed, then $\sum_{i=1}^\infty \lambda_i^2 < \infty$.

PROOF. It is clear that

$$\begin{aligned} P &= \Pr \left\{ \sup_{(g, \epsilon)_{i=1}} (\mathbf{A}g, \epsilon)_K^2 < \infty \right\} = \Pr \left\{ \sup_{\sum_{i=1}^\infty a_i^2 \leq 1} \left(\sum_{i=1}^\infty a_i \lambda_i \phi_i, \epsilon \right)_K^2 < \infty \right\} \\ &= \Pr \left\{ \sup_{\sum_{i=1}^\infty a_i^2 \leq 1} \left(\sum_{i=1}^\infty a_i \lambda_i (\phi_i, \epsilon)_K \right)^2 < \infty \right\} = \Pr \left\{ \sum_{i=1}^\infty \lambda_i^2 (\phi_i, \epsilon)_K^2 < \infty \right\}. \end{aligned}$$

Thus the first result is proved. Since $E(\phi_i, \epsilon)_K^2 = (\phi_i, \phi_i)_K = 1$, a simple application of Fatou's lemma (c.f. Royden, 1968) leads to the conclusion that (2.11) implies (2.10). Thus the second result is proved.

Moreover, if $P = 1$ and $\{(\phi_i, \epsilon)_K\}_{i=1}^\infty$ are independent, then by Kolmogorov's "three series theorem" (c.f. Chung, 1968) we have, for any $a > 0$,

$$(2.12) \quad \sum_{i=1}^\infty \Pr \{ \lambda_i^2 (\phi_i, \epsilon)_K^2 > a \} < \infty$$

and

$$(2.13) \quad \sum_{i=1}^\infty E \{ \lambda_i^2 (\phi_i, \epsilon)_K^2 \cdot \mathbf{1}([0, a]) \} < \infty,$$

where $\mathbf{1}([0, a])$ is the indicator of $[0, a]$.

When $\{(\phi_i, \epsilon)_K\}_{i=1}^\infty$ are identically distributed, we have $\lim_{i \rightarrow \infty} \lambda_i = 0$; otherwise, upon

choosing $a = \min\{1, \frac{1}{2} \limsup_{i \rightarrow \infty} \lambda_i\}$, we see that (2.12) does not hold. Now rewrite (2.13) as

$$\sum_{i=1}^{\infty} \lambda_i^2 E \left\{ (\phi_i, \varepsilon)_K^2 \cdot \mathbf{1} \left(\left[0, \frac{a}{\lambda_i} \right] \right) \right\} < \infty.$$

This implies (2.11) because $E \left\{ (\phi_i, \varepsilon)_K^2 \cdot \mathbf{1} \left(\left[0, \frac{a}{\lambda_i} \right] \right) \right\} \rightarrow 1$.

REMARK. One can easily construct dependent random variables $\{(\phi_i, \varepsilon)_K\}_{i=1}^{\infty}$ such that (2.10) holds but (2.11) does not. For example, let x be an absolutely continuous symmetric random variable with infinite variance, and let $(\phi_i, \varepsilon)_K$ be (x/λ_i) when $i - 1 \leq |x| < i$ and 0 otherwise, where $\lambda_i = [E\{x^2 \cdot \mathbf{1}([i - 1, i])\}]^{1/2}$. There also exist examples where $\{(\phi_i, \varepsilon)_K\}_{i=1}^{\infty}$ are independent but non-identically distributed such that (2.10) holds but (2.11) does not. These counter-examples illustrate the probabilistic nature of the method of regularization, while the general theory of linear estimators usually does not depend on more than the first and the second moments of the random error. More precisely, there exist two different error processes ε_1 and ε_2 such that they yield the same reproducing kernel, but in order to estimate the same linear functionals, it is legitimate to use the method of regularization for ε_1 but not for ε_2 (e.g., ε_1 can be taken as the one generated by the above counter-example and ε_2 can be the Gaussian one with the same reproducing kernel).

Acknowledgments. I am indebted to Professor Jack Kiefer for his encouragement and many helpful discussions. I also thank the associate editor for useful comments and suggestions.

REFERENCES

- CHUNG, K. L. (1968). *A Course in Probability*. Academic, New York.
- MITCHELLI, C. A. and WAHBA, G. (1979). Design problems for optimal surface interpolation. Technical Report No. 565, University of Wisconsin, Madison.
- PARZEN, E. (1961). Regression analysis of continuous parameter time series. In *Proceedings of the Fourth Berkeley Symposium* 1 649-689.
- ROYDEN, H. L. (1968). *Real Analysis*. MacMillan, New York.
- RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York.
- SACKS, J. and YLVIKAKER, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* 6 1112-1137.
- SPECKMAN, P. (1979). Minimax estimates of linear functionals in a Hilbert space. Unpublished manuscript.
- WAHBA, G. and WOLD, S. (1975). A completely automatic French curve: fitting spline functions by cross-validation. *Commun. Statist.* 4 1-17.

STATISTICS DEPARTMENT
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907