MINIMAXITY OF THE METHOD OF REGULARIZATION

ON STOCHASTIC PROCESSES

by

Ker-Chau Li*

Technical Report #81-57

Purdue University

Department of Statistics

December 1981

*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 regiments for the Doctor of Philosophy degree in the University of California, Berkeley, Department of Statistics.

MINIMAXITY OF THE METHOD OF REGULARIZATION ON STOCHASTIC PROCESSES*

Short Title . METHOD OF REGULARIZATION

Ker-Chau Li

Statistics Department
Purdue University
West Lafayette, Indiana 47907

SUMMARY

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.

KEY WORDS AND PHRASES: Autoregressive processes, minimaxity, method of regularization, robustness, smoothing splines, spectrum, stochastic processes.

*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 regiments for the Doctor of Philosophy degree in the University of California, Berkeley, Department of Statistics.

AMS Subject Classification: Primary 60G35, 62M10; Secondary 62G35, 62J05.

MINIMAXITY OF THE METHOD OF REGULARIZATION ON STOCHASTIC PROCESSES

1 Introduction

Suppose we observe a stochastic process Y on the interval [0,1] satisfying model $Y(t) = f(t) + \varepsilon_t$, 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, e.g. Parzen (1961), who established a Gauss-Markov type theorem.

In this note, we shall 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 (c.f. Parzen (1961)) and we write $\underset{\sim}{H} = \{g |$ g is a real function on [0,1] with absolutely continuous first derivative and square integrable second derivative$\}$, then we might assume that $f \in \underset{\sim}{H}$ and $\int_0^1 f''(x)^2 dx < \delta^2$ for some known constant δ. This becomes a nearly linear model (see 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 usual 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:

$$(1.1) \qquad \underset{f \in \underset{\sim}{H}}{\text{Min}} \ \frac{1}{\delta^2} \int_0^1 f''(x)^2\, dx + (f,f)_k - 2(f,Y)_K$$

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 (see Parzen (1961)).

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

$$(1.2) \qquad \underset{f \in \underset{\sim}{H}}{\text{Min}} \int_0^1 f''(x)^2\, dx + \frac{\delta^2}{\sigma^2} \sum_{i=i}^{n} (Y_i - f(x_i))^2$$

where Y_i is the observation made at $x_i \in [a,b]$ and assumed to be independent and 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 periodical splines is given by P. J. Laurent (see Micchelli and Wahba (1979)).) As indicated there, a minimax linear estimator for any bounded linear functional can be derived by *the method of regularization* (MOR hereafter), i.e., by operating the linear functional on the smoothing spline $\hat{f}$. Our method for getting the desired results follows a variation of Speckman's appraoch. 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) (see the next section for details) 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 genrating 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 MOR works. It is somehow surprising that although much work has been done on how to solve (1.2), there is little statistical literature on (1.1). Specifically, how to apply the cross-validation technique used in estimating δ^2/σ^2 in (1.2) to the problem (.1.1) is still unknown.

2 Main Result

Let $\underset{\sim}{I}$ be an index set, let $\underset{\sim}{H}_1$ and $\underset{\sim}{H}_2$ 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 $\underset{\sim}{H}_1$ onto $\underset{\sim}{H}_2$ 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 $\underset{\sim}{I}$ with mean 0 and the covariance structure generating a reproducing kernel Hilbert space $\underset{\sim}{H}_3$ with the kernel K. As usual we use $(\cdot,\cdot)_K$ to denote the inner product of $\underset{\sim}{H}_3$. Assume our observation Y is a stochastic process on $\underset{\sim}{I}$ satisfying the following model:

$$(2.2) \qquad Y = \underset{\sim}{A}f + \sigma \cdot \varepsilon$$

where $\underset{\sim}{A}$ is a bounded linear mapping from $\underset{\sim}{H}_1$ to $\underset{\sim}{H}_3$, σ is a known positive constant, and

$$(2.3) \qquad f \in \underset{\sim}{H}_1 \quad \text{and} \quad \langle \underset{\sim}{T}f, \underset{\sim}{T}f \rangle_2 \leq \delta^2$$

for some known constant δ. Assume also that

$$(2.4) \qquad \text{"}\underset{\sim}{A}f = 0 \quad \text{and} \quad \underset{\sim}{T}f = 0\text{"} \quad \text{imply} \quad \text{"}f = 0\text{"}.$$

Suppose y is a realization of Y such that

(2.5) the linear functional on $\underset{\sim}{H}_1$ defined by mapping an element g in $\underset{\sim}{H}_1$ to the corresponding realization of the random variable $(\underset{\sim}{A}g, Y)_K$ 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

$$(2.6) \qquad \langle g, \tilde{y} \rangle_1 = (\underset{\sim}{A}g, y)_K$$

where we write $(\underset{\sim}{A}f, y)_K$ to denote the realization of $(\underset{\sim}{A}f, Y)_K$ when y is the realization of Y.

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

Proof. Let $\underset{\sim}{N}$ be the null space of $\underset{\sim}{T}$. By (2.1), $\underset{\sim}{N}$ is finite dimensional and thus is a closed subspace of $\underset{\sim}{H}_1$. Let $\underset{\sim}{N}^{\perp}$ be the orthogonal complement of $\underset{\sim}{N}$ in $\underset{\sim}{H}_1$. Let $\underset{\sim}{T}_1$ be the restriction of $\underset{\sim}{T}$ to $\underset{\sim}{N}^{\perp}$. By(2.1) again, $\underset{\sim}{T}_1$ is one-to-one and onto, and thus $\underset{\sim}{T}_1^*\underset{\sim}{T}_1$ is one-to-one and onto.

Let $\underset{\sim}{A}_1$ and $\underset{\sim}{A}_2$ be the restriction of $\underset{\sim}{A}$ to $\underset{\sim}{N}^{\perp}$ and $\underset{\sim}{N}$, respectively. By (2.4), $\underset{\sim}{A}_2$ is 1-1 and hence $\underset{\sim}{A}_2^*\underset{\sim}{A}_2$ is 1-1 and onto (because $\underset{\sim}{N}$ is of finite dimension). Therefore $\frac{\sigma^2}{\delta^2}\underset{\sim}{T}_1^*\underset{\sim}{T}_1 + \underset{\sim}{A}_2^*\underset{\sim}{A}_2$ is a 1-to-1, onto, and self-adjoint linear mapping on $\underset{\sim}{H}_1$. Since $\frac{\sigma^2}{\delta^2}\underset{\sim}{T}^*\underset{\sim}{T} + \underset{\sim}{A}^*\underset{\sim}{A} =$ $\left(\frac{\sigma^2}{\delta^2}\underset{\sim}{T}_1^*\underset{\sim}{T}_1 + \underset{\sim}{A}_2^*\underset{\sim}{A}_2\right) + (\underset{\sim}{A}_1^*\underset{\sim}{A}_1)$, the rest of the proof follows immediately.

Theorem 2.1. Under (2.1) through (2.5), the following minimization problem has a unique solution $\hat{f}$ given by $\left(\frac{\sigma^2}{\delta^2}\underset{\sim}{T}^*\underset{\sim}{T} + \underset{\sim}{A}^*\underset{\sim}{A}\right)^{-1}\tilde{y}$:

$$\text{(2.7)}\qquad \underset{f\in \underset{\sim}{H}_1}{\text{Min}}\left\{\frac{\sigma^2}{\delta^2}\langle \underset{\sim}{T}f,\underset{\sim}{T}f\rangle_2 + (\underset{\sim}{A}f,\underset{\sim}{A}f)_K - 2(\underset{\sim}{A}f,y)_K\right\}.$$

Moreover, if (2.5) is not satisfied then the solution does not exist (infimum = $-\infty$).

Proof. It is clear that

$$\begin{aligned}
&\frac{\sigma^2}{\delta^2}\langle \underset{\sim}{T}f,\underset{\sim}{T}f\rangle_2 + (\underset{\sim}{A}f,\underset{\sim}{A}f)_K - 2(\underset{\sim}{A}f,y)_K\\
&= \frac{\sigma^2}{\delta^2}\langle \underset{\sim}{T}^*\underset{\sim}{T}f,f\rangle_1 + \langle \underset{\sim}{A}^*\underset{\sim}{A}f,f\rangle_1 - 2\langle f,\tilde{y}\rangle_1 \qquad \text{(by (.2.6))}\\
&= \left\langle\left(\frac{\sigma^2}{\delta^2}\underset{\sim}{T}^*\underset{\sim}{T} + \underset{\sim}{A}^*\underset{\sim}{A}\right)f, f\right\rangle_1 - 2\langle f,\tilde{y}\rangle_1\\
&= \langle \underset{\sim}{L}f,\underset{\sim}{L}f\rangle - 2\langle \underset{\sim}{L}f,\underset{\sim}{L}^{-1}\tilde{y}\rangle_1 \qquad \text{(by Lemma 2.1)}\\
&= \langle \underset{\sim}{L}f-\underset{\sim}{L}^{-1}\tilde{y}, \underset{\sim}{L}f-\underset{\sim}{L}^{-1}\tilde{y}\rangle_1 - \langle \underset{\sim}{L}^{-1}\tilde{y}, \underset{\sim}{L}^{-1}\tilde{y}\rangle_1\\
&\geq -\langle \underset{\sim}{L}^{-1}\tilde{y}, \underset{\sim}{L}^{-1}\tilde{y}\rangle_1 .
\end{aligned}$$

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

Given $\ell \in \underset{\sim}{H}_1$, 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 $(\underset{\sim}{e}, Y)_K$ for some $\underset{\sim}{e} \in H_3$. The risk (for squared error loss) is $E(\langle \ell, f \rangle_1 - (\underset{\sim}{e}, Y)_K)^2$, which is equal to $(\langle \ell, f \rangle_1 - (\underset{\sim}{e}, \underset{\sim}{A} f)_K)^2 + \sigma^2 (\underset{\sim}{e}, \underset{\sim}{e})_K$. 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 following minimax problem has a unique solution $\underset{\sim}{e}_0$ given by $\underset{\sim}{A} \left(\frac{\sigma^2}{\delta^2} \underset{\sim}{T}^* \underset{\sim}{T} + \underset{\sim}{A}^* \underset{\sim}{A} \right)^{-1} \ell$:

$$(2.8) \qquad \underset{\underset{\sim}{e} \in \underset{\sim}{H}_3}{\text{Min}} \; \underset{\langle \underset{\sim}{T} f, \underset{\sim}{T} f \rangle_2 \leq \delta^2}{\text{Max}} (\langle \ell, f \rangle_1 - (\underset{\sim}{e}, \underset{\sim}{A} f)_K)^2 + \sigma^2 (\underset{\sim}{e}, \underset{\sim}{e})_K .$$

Moreover, if y is a realization of Y satisfying (2.5), then

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

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:

$$\begin{aligned} (\underset{\sim}{e}_0, y)_K &= \left(\underset{\sim}{A} \left(\frac{\sigma^2}{\delta^2} \underset{\sim}{T}^* \underset{\sim}{T} + \underset{\sim}{A}^* \underset{\sim}{A} \right)^{-1} \ell, y \right)_K \\ &= \left\langle \left(\frac{\sigma^2}{\delta^2} \underset{\sim}{T}^* \underset{\sim}{T} + \underset{\sim}{A}^* \underset{\sim}{A} \right)^{-1} \ell, \tilde{y} \right\rangle_1 \qquad \text{(by (2.6))} \end{aligned}$$

$$= \langle \ell, \left(\frac{\sigma^2}{\delta^2} \underset{\sim}{T}^* \underset{\sim}{T} + \underset{\sim}{A}^* \underset{\sim}{A}\right)^{-1} \tilde{y} \rangle_1$$

$$= \langle \ell, \hat{f} \rangle_1 \qquad \text{(by Theorem 2.1)} .$$

Note that in the model (2.2) if we take $I = \{1,\ldots,n\}$ and $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)'$ 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

$$\underset{f \in \underset{\sim}{H}_1}{\text{Min}} \left\{ \frac{\sigma^2}{\delta^2} \langle \underset{\sim}{T}f, \underset{\sim}{T}f \rangle + (\underset{\sim}{A}f)'\underset{\sim}{A}f - 2(\underset{\sim}{A}f)'y \right\} ,$$

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

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

As was demonstrated in Speckman (1979), (1.2) is a special case of the above minimization problem. It is therefore from this point of view we consider that (1.1) is proposed by using a generalized idea of smoothing spline interpolation. Note also that in (2.7) we cannot replace $(\underset{\sim}{A}f,\underset{\sim}{A}f)_K - 2(\underset{\sim}{A}f,y)_K$ by $(\underset{\sim}{A}f - y,\underset{\sim}{A}f - y)_K$ because in most cases $(y,y)_K$ is $+\infty$.

<u>Example 1</u>. Ridge regression on stochastic process.

Take $\underset{\sim}{H}_1 = R^n$, $\underset{\sim}{H}_2 = R^m$ (where $n \geq m$), $\underset{\sim}{T}$ to be any linear mapping from R^n onto R^m and $\underset{\sim}{A}$ to be any one-to-one linear mapping from R^n to the reproducing kernel Hilbert space $\underset{\sim}{H}_3$. 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 $\underset{\sim}{T}$ 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 the Introduction. Let $\underset{\sim}{I} = [0,1]$. Take $\underset{\sim}{H}_1$ to be $\underset{\sim}{H}$ and equip it with the Soblev 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 $\underset{\sim}{H}_2$ to be $L^2[0,1]$ and define $\underset{\sim}{T}$ by $\underset{\sim}{T}(f) = f''$. The RKHS $\underset{\sim}{H}_3$ contains all differentiable function on [0,1] (c.f. Parzen (1961)). The inner product of $\underset{\sim}{H}_3$ 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 $\underset{\sim}{A}$ to be the identity map. Then (2.1) through (2.4) hold. To verify that (2.5) holds with probability 1, we may write

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

where $f \in \underset{\sim}{H}$. By Cauchy-Schwartz inequality, we have $\int_0^1 f(t)\varepsilon(t)dt \le (\int_0^1 f^2(t)dt)^{\frac{1}{2}}(\int_0^1 \varepsilon^2(t)dt)^{\frac{1}{2}} \le \langle f,f\rangle_1^{\frac{1}{2}} \cdot (\int_0^1 \varepsilon^2(t)dt)^{\frac{1}{2}}$, which shows that the linear functional on $\underset{\sim}{H}_1$ which maps f to $\int_0^1 f(t)\varepsilon(t)dt$ is bounded. Similar argument applies to the linear functional on $\underset{\sim}{H}_1$ which maps f to $\int_0^1 f''(t)\varepsilon(t)dt$. Finally, it is a well-known fact that f'(1), f(0) are f(1) are bounded functionals on $\underset{\sim}{H}_1$. Therefore (2.5) holds with probability 1. Thus Theorem 2.1 and 2.2 hold. The MOR works in this case. Moreover, we have the following global optimality

$$\underset{\underset{\sim}{e}(\cdot)\in \underset{\sim}{H}_3}{\text{Min}} \quad \underset{\int_0^1 f''(x)^2 dx \leqslant \delta^2}{\text{Max}} \quad \underset{t\in[0,1]}{\text{Max}} \quad E\left(f(t) - (\underset{\sim}{e}(t),Y)_K\right)^2$$

$$= \underset{\int_0^1 f''(x)^2 dx \leqslant \delta^2}{\text{Max}} \quad \underset{t\in[0,1]}{\text{Max}} \quad E(f(t) - \hat{f}(t))^2$$

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

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

Example 3. Assume that ε is the second order autoregressive Gaussian process and 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 $\underset{\sim}{H}_3$ contains all twice differentiable functions on [0,1], $\underset{\sim}{H}_1$ is homeomorphic to $\underset{\sim}{H}_3$, and Y does not belong to $\underset{\sim}{H}_3$. Hence the MOR does not work in this case. To avoid computation, let's look at another similar model. Let $\underset{\sim}{H}_1$ and $\underset{\sim}{H}_2$ be equal to $\underset{\sim}{H}_3$ given above. Let $\underset{\sim}{T}$ and $\underset{\sim}{A}$ 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 $\underset{\sim}{H}_3$, i.e. $(\varepsilon,\varepsilon)_K = +\infty$, 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 $1/(1+\sigma^2/\delta^2)\cdot Y(t)$.

Assume that $\underset{\sim}{H}_1$ and $\underset{\sim}{H}_3$ are separable hereafter. Assume also that

(2.9) the spectrum of $\underset{\sim}{A}^*\underset{\sim}{A}$ is discrete.

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

Now we characterize the conditions on $\underset{\sim}{A}$ such that (2.5) holds with probabability 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 $\underset{\sim}{H}_1$ or $\underset{\sim}{H}_3$ is of finite dimension, then $P=1$. Therefore, we assume that both $\underset{\sim}{H}_1$ and $\underset{\sim}{H}_3$ are of infinite dimension. By (2.9), there exist orthonormal systems $\{\psi_i\}_{i=1}^{\infty}$ and $\{\phi_i\}_{i=1}^{\infty}$ on $\underset{\sim}{H}_1$ and $\underset{\sim}{H}_3$ respectively, such that $\{\psi_i\}_{i=1}^{\infty}$ is complete and $\underset{\sim}{A}\psi_i = \lambda_i\phi_i$ for a sequence of numbers $\{\lambda_i\}_{i=1}^{\infty}$. The orthonormal systems $\{\psi_i\}_{i=1}^{\infty}$ are the eigenvectors of $\underset{\sim}{A}^*\underset{\sim}{A}$ with the corresponding eigenvalues $\{\lambda_i^2\}_{i=1}^{\infty}$. Note that (2.9) holds if $\underset{\sim}{A}$ is a compact linear operator. In particular, the operators in Example .1 and 2 satisfy (2.9).

<u>Theorem .2.3</u>. Under (2.2), $P=1$ iff

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

In particular, if

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

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

Proof. It is clear that

$$P = \Pr\left\{ \sup_{\langle g,g\rangle_1 \leqslant 1} (\underset{\sim}{A}g,\varepsilon)_K^2 < \infty \right\}$$

$$= \Pr\left\{ \sup_{\sum_{i=1}^{\infty} a_i^2 \leqslant 1} \left(\sum_{i=1}^{\infty} a_i\lambda_i\phi_i,\varepsilon \right)_K^2 < \infty \right\}$$

$$= \Pr\left\{ \sup_{\sum_{i=1}^{\infty} a_i^2 \leqslant 1} \left(\sum_{i=1} a_i\lambda_i(\phi_i,\varepsilon)_K \right)^2 < \infty \right\}$$

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

Thus the first result is proved. Since $E(\phi_i,\varepsilon)_K^2 = (\phi_i,\phi_i)_K = 1$, 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 $\{(\phi_i,\varepsilon)_K^2\}_{i=1}^{\infty}$ are independent, then by Kolmogorov's "three series theorem" (c.f. Chung (1968)) we have for any $a>0$,

$$\sum_{i=1}^{\infty} \Pr\left\{ \lambda_i^2(\phi_i,\varepsilon)_K^2 > a \right\} < \infty \tag{2.12}$$

and

$$\sum_{i=1}^{\infty} E\left(\lambda_i^2(\phi_i,\varepsilon)_K^2 \cdot 1([0,a]) \right) < \infty \tag{2.13}$$

where $1([0,a])$ is the indicator of $[0,a]$.

When $\{(\phi_i,\varepsilon)_K\}_{i=1}^{\infty}$ are identically-distributed, we have $\lim_{i\to\infty} \lambda_i = 0$; otherwise, upon choosing $a = \text{Min}\{1, \frac{1}{2}\overline{\lim_{i\to\infty}}\lambda_i\}$, we see that (2.12) does not hold. Now rewrite (2.13) as

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

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

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 $\frac{x}{\lambda_i}$ when $i-1 \leq |x| < i$ and 0 otherwise, where $\lambda_i = \left(E\left(x^2 \cdot 1([i-1,i))\right)\right)^{\frac{1}{2}}$. There also exist examples where $\{(\phi_i,\varepsilon)_K^2\}_{i=1}^{\infty}$ 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).

Acknowledgements. I am indebted to Professor Jack Kiefer for his encouragements 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 Press, New York.

Mitchelli, C. A. and Wahba, G. (1979). Design problems for optimal surface interpolation. Technical Report No. 565, University of Wisconsin.

Parzen, E. (1961). Regression analysis of continuous parameter time series. In Proceedings of the Fourth Berkeley Symposium, Vol. 1, 649-489.

Royden, H. L. (1968). Real Analysis. MacMillan, New York.

Rudin, W. (1973). Functional Analysis, McGraw-Hill, New York.

Sacks, J. and Ylvisaker, D. (1978). Linear estimation for approximately linear models. Ann. Stat. 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.