The Annals of Statistics
1990, Vol. 18, No. 3, 1116–1138

BOUNDS ON THE EFFICIENCY OF LINEAR PREDICTIONS USING AN INCORRECT COVARIANCE FUNCTION[1]

By Michael L. Stein

The University of Chicago

Suppose $z(\cdot)$ is a random process defined on a bounded set $R \subset \mathbb{R}^1$ with finite second moments. Consider the behavior of linear predictions based on $z(t_1), \dots, z(t_n)$, where $t_1, t_2, \dots$ is a dense sequence of points in R. Stein showed that if the second-order structure used to generate the predictions is incorrect but compatible with the correct second-order structure, the obtained predictions are uniformly asymptotically optimal as $n \to \infty$. In the present paper, a general method is described for obtaining rates of convergence when the covariance function is misspecified but compatible with the correct covariance function. When $z(\cdot)$ is Gaussian, these bounds are related to the entropy distance (the symmetrized Kullback divergence) between the measures for the random field under the actual and presumed covariance functions. Explicit bounds are given when $R = [0, 1]$ and $z(\cdot)$ is stationary with spectral density of the form $f(\lambda) = (a^2 + \lambda^2)^{-p}$, where p is a known positive integer and a is the parameter that is misspecified. More precise results are given in the case $p = 1$. An application of this result implies that equally spaced observations are asymptotically optimal in the sense used by Sacks and Ylvisaker in terms of maximizing the Kullback divergence between the actual and presumed models when $z(\cdot)$ is Gaussian.

1. Introduction. This paper continues the investigation on the effect on optimal linear predictions of misspecifying the second-order structure of a random field begun by the author in previous works [Stein (1988, 1990)]. In particular, a new bound on the efficiency of predictions with an incorrect covariance function that is compatible (defined below) with the actual covariance function is given. Furthermore, explicit bounds are obtained for some simple one-dimensional processes. The general setup will be the same as in these previous works: We suppose $z(\cdot)$ is a continuous random field with finite second moments defined on a bounded region $R \subset \mathbb{R}^d$, and we consider the properties of linear predictions of $z(\cdot)$ based on an increasing number of observations in R. If the mean and covariance functions of the process are known, then optimal linear predictors based on a finite number of observations can be readily calculated; they are just the generalized least squares predictors. For example, if $Z_n = (z(t_1), \dots, z(t_n))'$, then the optimal linear predictor of $z(t_0)$ is

$$Ez(t_0) + \operatorname{cov}\left(z(t_0), Z_n'\right)\left[\operatorname{cov}\left(Z_n, Z_n'\right)\right]^{-1}(Z_n - EZ_n),$$

Received August 1988; revised April 1989.

[1]Author supported by a NSF Mathematical Sciences Postdoctoral Research Fellowship. Manuscript prepared using computer facilities supported in part by NSF Grants DMS-86-01732 and DMS-84-04941 to the Department of Statistics at The University of Chicago.

AMS 1980 *subject classifications*. Primary 62M20; secondary 41A25, 60G60.

Key words and phrases. Spatial statistics, approximation in Hilbert spaces, Kullback divergence, design for time series experiments.

assuming the matrix inverse exists. In practice, the mean or covariance function is at least partially unknown, so it is important to know how predictors generated using an incorrect mean or covariance function perform. Throughout this paper, we will assume the mean function is correctly specified and set it, without loss of generality, to be identically 0.

Suppose $K_0(s,t)$ and $K_1(s,t)$ are two continuous positive definite functions on $R \times R$. For $i = 0, 1$, corresponding to K_i there is a unique Gaussian measure P_i on the space of functions on R possessing mean 0 and covariance function K_i. We say that K_0 and K_1 are compatible on R if P_0 and P_1 are mutually absolutely continuous. See Stein (1988, 1990) for further discussion on the compatibility of covariance functions and Ibragimov and Rozanov (1978) and Yadrenko (1983) for results on the mutual absolute continuity of Gaussian measures. For example, suppose $K_0(s,t) = K_0(s-t)$ and $K_1(s,t) = K_1(s-t)$ are homogeneous covariance functions for processes in $\mathbb{R}^1$ with spectral densities f_0 and f_1, respectively, and R is any finite interval. Then by Theorem 17 in Chapter 3 of Ibragimov and Rozanov (1978), sufficient conditions for the compatibility of K_0 and K_1 on R are

$$0 < \liminf_{\lambda \to \infty} f_0(\lambda)\lambda^{\nu} \le \limsup_{\lambda \to \infty} f_0(\lambda)\lambda^{\nu} < \infty,$$

for some $\nu > 1$ and

$$\int_{|\lambda|>s} \left(\frac{f_0(\lambda) - f_1(\lambda)}{f_0(\lambda)} \right)^2 d\lambda < \infty, \tag{1.1}$$

for some $s < \infty$. Suppose $t_1, t_2, \ldots$ is a dense sequence of points in R, and we observe a zero-mean random process $z(\cdot)$ (not necessarily Gaussian) at $t_1, \ldots, t_n$. Furthermore, suppose K_1 is the covariance function used to generate predictions, K_0 is the actual covariance and K_0 and K_1 are compatible on R. For a covariance function $K(s,t)$ on $R \times R$, let $H(K)$ be the Hilbert space of random variables generated by $z(t)$, $t \in R$, with respect to the inner product defined by K. More specifically, $H(K)$ consists of all random variables of the form $a_1 z(s_1) + \cdots + a_m z(s_m)$, where $s_1, \ldots, s_m \in R$, plus all L^2 limits of these random variables assuming $\mathrm{cov}(z(s), z(t)) = K(s,t)$. We will also use K to denote the covariance operator on elements of $H(K)$, so that $h_1, h_2 \in H(K)$ implies $K(h_1, h_2) = \mathrm{cov}(h_1, h_2)$. In particular, $K(z(t_1), z(t_2)) = K(t_1, t_2)$, where K is interpreted as an operator on $H(K) \times H(K)$ on the left-hand side and as a function on $R \times R$ on the right-hand side. The compatibility of K_0 and K_1 implies $H(K_0) = H(K_1)$ [Ibragimov and Rozanov (1978), page 71]. Stein (1990) showed that the supremum, over all elements of $H(K_0)$, of the mean square prediction error using the incorrect K_1 divided by the mean square error of the optimal prediction based on K_0 is bounded by a quantity tending to 1 as $n \to \infty$. However, Stein (1990) only obtained rates of convergence when the mean function, and not the covariance function, is misspecified.

In the present paper, explicit bounds when the covariance function is misspecified are given for the first time. In Section 2, a different bound on the efficiency of predictions under an incorrect covariance function is given. When the process is weakly stationary, this bound can be expressed in terms of how well an element of an appropriate Hilbert space can be approximated by an element of a subspace generated by the observations, thus providing a general approach for calculating the bound. The bound can be related to the Kullback divergence and the variation distance between the zero-mean Gaussian measures with covariance functions K_0 and K_1. An explicit bound is obtained in the case $R = [0, 1]$ and the two spectral densities are $f_0(\lambda) = (a^2 + \lambda^2)^{-p}$ and $f_1(\lambda) = (b^2 + \lambda^2)^{-p}$, where p is a positive integer. It is shown that if $z(\cdot)$ and its $p - 1$ mean square derivatives are observed at $0 = t_0 < t_1 < \cdots < t_n = 1$, where the maximum gap between observations is $O(n^{-1})$, then the relative increase in prediction error variance caused by using the wrong spectral density is $O(n^{-\min(3, 2p)})$. This rather fast rate of convergence suggests that for purposes of predicting $z(\cdot)$ on $[0, 1]$, the penalty for misspecifying a is very small.

When $p = 1$, the inverse of the covariance matrix of the observations can be written down, allowing for much more precise calculations. In particular, an asymptotic expression of the Kullback divergence between the Gaussian measures with spectral densities f_0 and f_1 based on n observations on $[0, 1]$ is derived. An application of this result to the problem of choosing the observation points to maximize the Kullback divergence is given. This problem is similar to one considered by Sacks and Ylvisaker (1966, 1968, 1970) on designing time series experiments to obtain the best estimates of regression coefficients. In Section 5, it is shown that equally spaced observations are asymptotically optimal in terms of maximizing the Kullback divergence between the observations. This appears to be the first such result on asymptotically optimal designs of time series experiments for distinguishing between two stationary covariance functions.

2. General results. Suppose K_0 and K_1 are compatible covariance operators on a separable Hilbert space $H(K_0)$. Let $z_1, z_2, \ldots$ be a linearly independent basis for $H(K_0)$ and take $Z_n = (z_1, \ldots, z_n)'$. Using the same notation as Stein (1990), for $h \in H(K_0)$, define $e_i(h, n)$ to be the error in predicting h based on Z_n assuming K_i is the covariance operator, $e_i(h, 0) = h$, and let E_i denote expectation under K_i. Define

$$a_n(h) = \frac{E_0(e_1(h,n) - e_0(h,n))^2}{E_0 e_0(h,n)^2},$$

$$\tilde{a}_n(h) = \frac{E_1(e_0(h,n) - e_1(h,n))^2}{E_1 e_1(h,n)^2},$$

$$b_n(h) = \frac{E_1 e_0(h,n)^2 - E_0 e_0(h,n)^2}{E_0 e_0(h,n)^2},$$

and

$$\tilde{b}_n(h) = \frac{E_0 e_1(h,n)^2 - E_1 e_1(h,n)^2}{E_1 e_1(h,n)^2},$$

where $0/0$ is taken to be 0. Note that

$$1 + a_n(h) = E_0 e_1^2 / E_0 e_0^2, \tag{2.1}$$

which follows from the fact that $e_0(h,n)$ and $e_1(h,n) - e_0(h,n)$ are uncorrelated under K_0. Thus, $a_n(h)$ measures the relative increase in the mean square prediction error caused by using K_1 to define the linear predictor when K_0 is the correct covariance function. Further note that $e_1(h,n)$ is unchanged when K_1 is multiplied by a positive constant, so that all results in this paper on $a_n(h)$ that hold when K_0 and K_1 are compatible still hold when αK_1 is used to define $e_1(h,n)$, where α is a positive constant. Again taking K_1 to be the presumed covariance operator and K_0 the actual covariance operator, $\tilde{b}_n(h)$ measures the difference between what we think the mean square prediction error is, $E_1 e_1^2$, and what it actually is, $E_0 e_1^2$, relative to what we think it is. Stein (1990) showed

$$\lim_{n\to\infty} \sup_{h \in H(K_0)} a_n(h) \to 0$$

and

$$\lim_{n\to\infty} \sup_{h \in H(K_0)} |b_n(h)| \to 0$$

and the analogous results for $\tilde{a}_n(h)$ and $\tilde{b}_n(h)$. The purpose of this section is to derive new upper bounds on $a_n(h)$ and $b_n(h)$.

Let us define $b_n = b_n(z_{n+1})$ and $\tilde{b}_n = \tilde{b}_n(z_{n+1})$, and

$$M_n = \sum_{j=0}^{n-1} \left(b_j + \tilde{b}_j\right).$$

Now, K_0 and K_1 compatible implies there exists a constant $1 \le c < \infty$ such that

$$c^{-1} \le E_1 h^2 / E_0 h^2 \le c, \tag{2.2}$$

for all $h \in H(K_0)$.

THEOREM 1. *For K_0 and K_1 compatible and satisfying* (2.2),

$$\lim_{n\to\infty} M_n = M = \sum_{j=0}^{\infty} \left(b_j + \tilde{b}_j\right) \quad \textit{exists}$$

and for $n \ge 0$,

$$\sup_{h \in H(K_0)} a_n(h) \le \tfrac{1}{2}(c+1)(M - M_n) \tag{2.3}$$

and

$$\sup_{h \in H(K_0)} b_n(h)^2 \leq 2(M - M_n)\max\{1, 2(M - M_n)\}. \tag{2.4}$$

Analogous results hold for $\tilde{a}_n(h)$ and $\tilde{b}_n(h)$.

Define V_n to be the covariance matrix of Z_n under K_0 and W_n the covariance matrix of Z_n under K_1. The linear independence of $z_1, z_2, \dots$ imply V_n and W_n are invertible for all n. The following lemmas will be useful in proving Theorem 1.

LEMMA 1. *For K_0 and K_1 compatible and $n \geq 0$,*

$$\begin{aligned} b_n + \tilde{b}_n &= \operatorname{tr}(V_{n+1}^{-1} - W_{n+1}^{-1})(W_{n+1} - V_{n+1}) \\ &\quad - \operatorname{tr}(V_n^{-1} - W_n^{-1})(W_n - V_n), \end{aligned} \tag{2.5}$$

where $\operatorname{tr}(V_0^{-1} - W_0^{-1})(W_0 - V_0)$ *is defined to be* 0.

PROOF. The right-hand side of (2.5) equals

$$-2 + \operatorname{tr} V_{n+1}^{-1}W_{n+1} - \operatorname{tr} V_n^{-1}W_n + \operatorname{tr} W_{n+1}^{-1}V_{n+1} - \operatorname{tr} W_n^{-1}V_n,$$

so it suffices to show

$$b_n = -1 + \operatorname{tr} V_{n+1}^{-1}W_{n+1} - \operatorname{tr} V_n^{-1}W_n. \tag{2.6}$$

Let

$$V_{n+1} = \begin{pmatrix} V_n & \mathbf{v}_{n+1} \\ \mathbf{v}'_{n+1} & v_{n+1} \end{pmatrix},$$

and partition W_{n+1} similarly. Using the formula for the inverse of a partitioned matrix,

$$\begin{aligned} &\operatorname{tr} V_{n+1}^{-1}W_{n+1} - \operatorname{tr} V_n^{-1}W_n \\ &\quad = \left(v_{n+1} - \mathbf{v}'_{n+1}V_n^{-1}\mathbf{v}_{n+1}\right)^{-1}\left(\mathbf{v}'_{n+1}V_n^{-1}W_nV_n^{-1}\mathbf{v}_{n+1} - 2\mathbf{w}'_{n+1}V_n^{-1}\mathbf{v}_{n+1} + w_{n+1}\right. \end{aligned}$$

But

$$E_0 e_0(z_{n+1}, n)^2 = \mathbf{v}_{n+1} - \mathbf{v}'_{n+1}V_n^{-1}\mathbf{v}_{n+1}$$

and

$$\begin{aligned} E_1 e_0(z_{n+1}, n)^2 &= E_1\left(z_{n+1} - \mathbf{v}'_{n+1}V_n^{-1}Z_n\right)^2 \\ &= \mathbf{v}'_{n+1}V_n^{-1}W_nV_n^{-1}\mathbf{v}_{n+1} - 2\mathbf{w}'_{n+1}V_n^{-1}\mathbf{v}_{n+1} + w_{n+1}, \end{aligned}$$

so (2.6) and the lemma follow. □

LEMMA 2. *For K_0 and K_1 compatible and $n \geq 0$,*

$$\sup_{h \in H(K_0)} \left(b_n(h) + \tilde{b}_n(h)\right) \leq M - M_n.$$

PROOF. Using the compatibility of K_0 and K_1 along with Lemmas 4.4 and 4.5 and Theorem 4.3 in Kuo (1975), it follows that $\operatorname{tr}(V_n^{-1} - W_n^{-1})(W_n - V_n)$ increases monotonically in n to a finite limit. From Lemma 1, we see that this limit is given by M, thus establishing the first claim in Theorem 1. $H(K_0)$ separable implies that M is the same for any linearly independent basis for $H(K_0)$. For $h \in H(K_0)$ satisfying $E_0 e_0(h, n)^2 > 0$, we can choose $z'_{n+2}, z'_{n+3}, \dots$ such that $z_1, \dots, z_n, h, z'_{n+2}, z'_{n+3}, \dots$ forms a linearly independent basis for $H(K_0)$. The monotonicity of $\operatorname{tr}(V_n^{-1} - W_n^{-1})(W_n - V_n)$ implies that $(b_j + \tilde{b}_j) \geq 0$ for all j, and it follows that

$$0 \leq b_n(h) + \tilde{b}_n(h) \leq M - M_n.$$

Since this inequality trivially holds when $E_0 e_0(h, n)^2 = 0$, the lemma obtains. □

PROOF OF THEOREM 1. By definition of $b_n(h)$ and $\tilde{b}_n(h)$ and (2.1),

$$\begin{aligned}(1 + b_n(h))(1 + \tilde{b}_n(h)) &= \frac{E_1 e_0(h,n)^2}{E_0 e_0(h,n)^2} \frac{E_0 e_1(h,n)^2}{E_1 e_1(h,n)^2} \\ &= (1 + \tilde{a}_n(h))(1 + a_n(h)).\end{aligned}$$

Thus,

$$a_n(h) \leq b_n(h) + \tilde{b}_n(h) + b_n(h)\tilde{b}_n(h). \tag{2.7}$$

Using (2.2), it follows that $b_n(h)$ and $\tilde{b}_n(h)$ are bounded above by $c - 1$ and

$$b_n(h) + \tilde{b}_n(h) + b_n(h)\tilde{b}_n(h) \leq \tfrac{1}{2}(c + 1)\big(b_n(h) + \tilde{b}_n(h)\big). \tag{2.8}$$

Combining (2.7), (2.8) and Lemma 2, (2.3) obtains. Since $b_n + \tilde{b}_n \to 0$ as $n \to \infty$, for any $\varepsilon > 0$, $\frac{1}{2}(c + 1)$ can be replaced by $1 + \varepsilon$ for all n sufficiently large in (2.8) and (2.3).

From Lemma 2 and (2.7), we have

$$M - M_n \geq b_n(h) + \tilde{b}_n(h) \geq -b_n(h)\tilde{b}_n(h).$$

If $b_n(h)$ and $\tilde{b}_n(h)$ are both nonnegative,

$$M - M_n \geq b_n(h) \geq 0.$$

Now $b_n(h) + \tilde{b}_n(h)$ is nonnegative, so $b_n(h)$ and $\tilde{b}_n(h)$ cannot both be negative. Suppose $b_n(h) \geq 0$ and $\tilde{b}_n(h) < 0$, so that $|b_n(h)/\tilde{b}_n(h)| \geq 1$. If $|b_n(h)/\tilde{b}_n(h)| \geq 2$, then

$$M - M_n \geq b_n(h) + \tilde{b}_n(h) \geq \tfrac{1}{2}b_n(h) \geq 0,$$

and if $2 \geq |b_n(h)/\tilde{b}_n(h)| \geq 1$, then

$$M - M_n \geq -b_n(h)\tilde{b}_n(h) \geq \tfrac{1}{2}b_n(h)^2.$$

Finally, suppose $b_n(h) < 0$ and $\tilde{b}_n(h) \geq 0$, so that $|\tilde{b}_n(h)/b_n(h)| \geq 1$. If $|\tilde{b}_n(h)/b_n(h)| > 2$, then

$$M - M_n \geq b_n(h) + \tilde{b}_n(h) \geq -b_n(h) > 0,$$

and if $2 \geq |\tilde{b}_n(h)/b_n(h)| \geq 1$, then

$$M - M_n \geq -b_n(h)\tilde{b}_n(h) \geq b_n(h)^2.$$

Combining these inequalities, (2.4) obtains. □

3. Distances between equivalent Gaussian measures. If the process in question is in fact Gaussian, then $M - M_n$ can be interpreted as a measure of the information in $z_{n+1}, z_{n+2}, \dots$ not contained in Z_n for distinguishing between P_0 and P_1. The purpose of this section is to relate $M - M_n$ to more familiar measures of distances between probability measures. Thus, the bounds obtained in the next section on $M - M_n$ imply bounds on these other measures of distance. Following the notation in Blackwell and Dubins (1962), let $P_i^n(Z_n)$ be a conditional distribution for the "future" observations $Z_{-n} = (z_{n+1}, z_{n+2}, \dots)$ given the present observations Z_n under P_i. For P_0 and P_1 equivalent on a σ-algebra $\mathfrak{F}$, there exists a density q satisfying

$$P_1(A) = \int_A q\, dP_0,$$

for all $A \in \mathfrak{F}$. Define

$$q_n(Z_n) = \int q(Z_n, Z_{-n})\, dP_0^n(Z_{-n}|Z_n).$$

Based on Z_n, the Kullback divergence between the equivalent measures P_0 and P_1 is given by

$$I_n(P_0, P_1) = E_0 \log q_n^{-1}.$$

The Kullback divergence is used in Section 5 to define an optimality criterion for designing experiments to distinguish between two possible models for a Gaussian process. The entropy distance based on Z_n is a symmetrized Kullback divergence and is defined as [Ibragimov and Rozanov (1978), page 75]

$$r_n(P_0, P_1) = I_n(P_0, P_1) + I_n(P_1, P_0).$$

Then r_n is increasing and P_0 is equivalent to P_1 if and only if

$$r(P_0, P_1) = \lim_{n\to\infty} r_n(P_0, P_1)$$

is finite [Ibragimov and Rozanov (1978), page 77]. Using a general result on the convergence of Kullback divergences [see Kullback, Kegel and Kullback (1987), pages 33–35, for example],

$$\lim_{n\to\infty} I_n(P_0, P_1) = E_0 \log q^{-1},$$

which we will denote by $I(P_0, P_1)$. We thus have

$$r(P_0, P_1) = E_0 \log q^{-1} + E_1 \log q,$$

for P_0 and P_1 equivalent. For equivalent Gaussian measures P_0 and P_1 with zero means, we have

$$\log q_n^{-1} = -\tfrac{1}{2}\log|V_n| + \tfrac{1}{2}\log|W_n| - \tfrac{1}{2}Z_n'V_n^{-1}Z_n + \tfrac{1}{2}Z_n'W_n^{-1}Z_n,$$

so

$$\begin{aligned} r_n(P_0, P_1) &= E_0\left(-\tfrac{1}{2}Z_n'V_n^{-1}Z_n + \tfrac{1}{2}Z_n'W_n^{-1}Z_n\right) + E_1\left(\tfrac{1}{2}Z_n'V_n^{-1}Z_n - \tfrac{1}{2}Z_n'W_n^{-1}Z_n\right) \\ &= -n + \tfrac{1}{2}\mathrm{tr}W_n^{-1}V_n + \tfrac{1}{2}\mathrm{tr}V_n^{-1}W_n. \end{aligned}$$

Thus,

$$r_{n+1} - r_n = \tfrac{1}{2}\left(b_n + \tilde{b}_n\right),$$

so that for equivalent Gaussian measures $r_n = \frac{1}{2}M_n$ and $r = \frac{1}{2}M$.

Similarly, we can show

$$(3.1)\quad E_0 I(P_0^n(Z_n), P_1^n(Z_n)) + E_1 I(P_1^n(Z_n), P_0^n(Z_n)) = \tfrac{1}{2}(M - M_n).$$

Since the Kullback divergence is nonnegative, the right-hand side of (3.1) serves as a bound for both terms on the left-hand side. (3.1) can be used to bound the expected variation distance between $P_0^n(Z_n)$ and $P_1^n(Z_n)$. The variation distance between arbitrary measures P_0 and P_1 defined on the same σ-algebra $\mathfrak{F}$ is

$$\rho(P_0, P_1) = \sup_{A\in\mathfrak{F}} |P_0(A) - P_1(A)|.$$

If $I(P_0, P_1)$ is the entropy distance relative to this same σ-field, then it follows from Kullback (1967) that

$$I(P_0, P_1) \geq \tfrac{1}{2}\rho^2(P_0, P_1).$$

Replacing P_0 and P_1 by the conditional distributions $P_0^n(Z_n)$ and $P_1^n(Z_n)$, these relationships hold almost surely, so taking expectations under P_0, we obtain

$$E_0 I(P_0^n(Z_n), P_1^n(Z_n)) \geq \tfrac{1}{2}E_0\rho^2(P_0^n(Z_n), P_1^n(Z_n)),$$

so by (3.1) and the nonnegativity of Kullback divergences,

$$E_0\rho^2(P_0^n(Z_n), P_1^n(Z_n)) \leq M - M_n,$$

and by the Cauchy–Schwarz inequality,

$$(3.2)\qquad E_0\rho(P_0^n(Z_n), P_1^n(Z_n)) \leq (M - M_n)^{1/2}.$$

Blackwell and Dubins (1962) showed that if P_0 and P_1 are arbitrary equivalent measures,

$$\rho(P_0^n(Z_n), P_1^n(Z_n)) \to 0 \quad \text{almost surely}.$$

(3.2) gives a bound on the average rate of convergence when P_0 and P_1 are equivalent zero-mean Gaussian measures. Whereas Theorem 1 is a statement about the properties of linear predictors of linear functionals, (3.2) allows us to make statements about how much predictions of nonlinear functionals are changed when an incorrect Gaussian measure is used to produce the predictions. For example, in mining applications, one is often interested in the "tonnage" function: The probability that the mineral concentration at some site is greater than some cutoff grade z_0 [Matheron (1984)]. Suppose the mineral concentration $z(\cdot)$ which is necessarily nonnegative, is modeled so that its logarithm is a zero-mean Gaussian measure. If P_0 and P_1 are equivalent zero-mean Gaussian measures for $\log z(\cdot)$ on a region R, then by (3.2),

$$E_0|P_0(z(x) > z_0|Z_n) - P_1(z(x) > z_0|Z_n)| \le (M - M_n)^{1/2},$$

for all z_0 and all $x \in R$, where M and M_n are calculated using the covariance structure of $\log z(\cdot)$.

4. Explicit bounds. In this section, rates of convergence for $a_n(h)$ and $b_n(h)$ are obtained for some simple one-dimensional processes when $R = [0, 1]$. The main results are stated in Section 4.1; proofs are given in Section 4.3. A general approach to bounding $a_n(h)$ is discussed in Section 4.2.

4.1. *Main results.* Assume K_0 and K_1 are homogeneous covariance functions on $\mathbb{R}^1$ with corresponding spectral densities f_0 and f_1. Suppose $f_0(\lambda) = (a^2 + \lambda^2)^{-p}$ for a positive integer p and that K_0 and K_1 are compatible on $[0, 1]$. These conditions imply that $z(\cdot)$ has exactly $p - 1$ mean square derivatives under either K_0 or K_1. Define $T_n = (t_{0n}, \dots, t_{nn})$, where $0 = t_{0n} < t_{1n} < \cdots < t_{nn} = 1$. Assume

$$\max_{1 \le j \le n} (t_{jn} - t_{j-1,n}) = O(n^{-1}). \tag{4.1}$$

Suppose we observe $z^{(k)}(t_{jn})$ for $k = 0, \dots, p - 1$ and $j = 0, \dots, n$. That is, we not only observe $z(\cdot)$ at T_n, but also all of its mean square derivatives at these locations. We thus have $p(n + 1)$ "observations." Let us define quantities such as $a_{T_n, p-1}$, $b_{T_n, p-1}$ and $M_{T_n, p-1}$ as in Section 2 where the observations are taken to be $z(\cdot)$ and all $p - 1$ of its mean square derivatives on T_n.

PROPOSITION 1. *For* $f_0(\lambda) = (a^2 + \lambda^2)^{-p}$ *and* $f_1(\lambda) = (b^2 + \lambda^2)^{-p}$, *where* $a, b > 0$, *p is a positive integer and* $\{T_n\}$ *satisfies* (4.1),

$$\sup_{h \in H(K_0)} a_{T_n, p-1}(h) = O(n^{-\min(3, 2p)}). \tag{4.2}$$

When $p = 1$, this rate of convergence is shown to be the best possible in Section 5. I would conjecture that this rate of convergence is the best possible for all positive integers p. Using (2.3), Proposition 1 gives a rate of convergence for the supremum of $|b_n(h)|$ over $h \in H(K_0)$ of order n^{-1} when $p = 1$ and $n^{-3/2}$ when $p \ge 2$. By more direct means, we can obtain a sharper bound on $|b_n(h)|$.

PROPOSITION 2. *Suppose* $f_0(\lambda) = (a^2 + \lambda^2)^{-p}$ *and* T_n *satisfies* (4.1). *If* f_1 *satisfies*

$$\left| \frac{f_1(\lambda) - f_0(\lambda)}{f_0(\lambda)} \right| \leq \frac{C}{(1 + \lambda^2)^q}, \tag{4.3}$$

for some constant C *and positive integer* $q \leq p$, *then*

$$\sup_{h \in H(K_0)} |b_{T_n, p-1}| = O(n^{-\min(2q, p+1)}). \tag{4.4}$$

If

$$\frac{f_1(\lambda) - f_0(\lambda)}{f_0(\lambda)} \geq \frac{B}{(1 + \lambda^2)^q}, \tag{4.5}$$

for some constant $B > 0$ *and for positive integer* $q \leq p$, *then*

$$\sup_{h \in H(K_0)} b_{T_n, p-1}(h) \geq A n^{-2q}, \tag{4.6}$$

for some constant $A > 0$.

When $f_1(\lambda) = (b^2 + \lambda^2)^{-p}$, $p > 1$, and $b < a$, we see that under (4.1) we have

$$\sup_{h \in H(K_0)} n^2 b_{T_n, p-1}(h) \quad \text{and} \quad \sup_{h \in H(K_0)} \{-n^2 \tilde{b}_{T_n, p-1}(h)\}$$

are bounded away from 0 and ∞ as $n \to \infty$, but by (4.2)

$$\sup_{h \in H(K_0)} a_{T_n, p-1}(h) = O(n^{-3}).$$

We have that the maximum possible effect on the mean square errors of predictions from using the wrong covariance function is at least an order of magnitude less than the effect on the value of the mean square prediction error obtained by using the wrong covariance function. This result is in line with practical experience, which suggests that misspecifying the covariance structure tends to have a larger impact on what we think the mean square prediction error is than on the actual efficiency of the predictor [Starks and Sparks (1987)].

4.2. *General approach.* Suppose $H(K_0)$ is generated by $z(t)$ for $t \in R$, R a compact region in $\mathbb{R}^1$, and $K_0(s,t) = K_0(s-t)$ and $K_1(s,t) = K_1(s-t)$ are continuous covariance functions on $\mathbb{R}^1$ with spectral distributions $F_0(d\lambda)$ and $F_1(d\lambda)$, respectively, and are compatible on R. Define $L_{R \times R}(F_0 \times F_1)$ to be the L_2 closure with respect to the inner product

$$\langle \varphi, \psi \rangle_{F_0 \times F_1} = \int \varphi(\lambda, \mu) \overline{\psi(\lambda, \mu)}\, F_0(d\lambda) F_1(d\mu)$$

of functions of the form

$$\varphi(\lambda,\mu) = \sum_{j,k} c_{jk} e^{i(\lambda s_j - \mu t_k)},$$

where the range of summation is finite, $s_j, t_k \in R$, and the c_{jk}'s are real constants. By Theorem 8 in Chapter 3 of Ibragimov and Rozanov (1978), K_0 and K_1 are compatible on R if and only if there exists a function $\Psi(\lambda,\mu) \in L_{R\times R}(F_0 \times F_1)$ satisfying

$$(4.7)\quad K_0(s-t) - K_1(s-t) = \frac{1}{4\pi^2}\int e^{-i(\lambda s - \mu t)}\Psi(\lambda,\mu)F_0(d\lambda)F_1(d\mu),$$

for $s, t \in R$. Now the Hilbert space $H(K_0)$ is isomorphic to the Hilbert space of functions $L_R(F_0)$, where $L_R(F_0)$ is the closure with respect to the inner product

$$\langle \varphi, \psi\rangle_{F_0} = \int \varphi(\lambda)\overline{\psi(\lambda)}\, F_0(d\lambda)$$

of functions of the form

$$\sum_{j=1}^{n} c_j e^{i\lambda t_j},$$

$t_1,\dots,t_n \in R$. Let $\varphi_1, \varphi_2, \dots$ be a linearly independent basis for $L_R(F_0)$ and $z_1, z_2, \dots$ the corresponding elements in $H(K_0)$. For example, $\varphi_j(\lambda) = e^{i\lambda t_j}$, $j = 1, 2, \dots,$ where $t_1, t_2, \dots$ is dense in R, form a basis for $L_R(F_0)$; the element corresponding to φ_j in $H(K_0)$ is $z(t_j)$. Define V_n to be the $n \times n$ matrix with jkth element

$$\int \varphi_j(\lambda)\overline{\varphi_k(\lambda)}\, F_0(d\lambda) = E_0 z_j z_k$$

and W_n the $n \times n$ matrix with jkth element

$$\int \varphi_j(\lambda)\overline{\varphi_k(\lambda)}\, F_1(d\lambda) = E_1 z_j z_k.$$

Then V_n and W_n are nonsingular, so we can define c_{jk}^n for $j, k = 1, \dots, n$ to be the jkth element of $W_n^{-1} - V_n^{-1}$ and

$$\Psi_n(\lambda,\mu) = \sum_{j,k=1}^{n} c_{jk}^n \varphi_j(\lambda)\overline{\varphi_k(\mu)}\,.$$

By straightforward calculation

$$(4.8)\qquad \int |\Psi_n(\lambda,\mu)|^2 F_0(d\lambda)F_1(d\mu) = \operatorname{tr}\left(V_n^{-1} - W_n^{-1}\right)(W_n - V_n)..$$

Following Ibragimov and Rozanov (1978), pages 89–90, we have that

$$\Psi_n(\lambda,\mu) \to \Psi(\lambda,\mu)$$

in $L_{R\times R}(F_0 \times F_1)$,

$$\int |\Psi(\lambda,\mu) - \Psi_n(\lambda,\mu)|^2 F_0(d\lambda)F_1(d\mu) = M - M_n,$$

and $\Psi_n(\lambda,\mu)$ is the projection of $\Psi(\lambda,\mu)$ onto the subspace of $L_{R\times R}(F_0 \times F_1)$ of functions of the form

$$\sum_{j,k=1}^{n} a_{jk}\varphi_j(\lambda)\overline{\varphi_k(\mu)}\,,$$

with respect to the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_{F_0\times F_1}$, so that

$$M - M_n \le \int \left|\Psi(\lambda,\mu) - \sum_{j,k=1}^{n} a_{jk}\varphi_j(\lambda)\overline{\varphi_k(\mu)}\right|^2 F_0(d\lambda)F_1(d\mu), \tag{4.9}$$

with equality if $a_{jk} = c_{jk}^n$ for $j,k = 1,\dots,n$. It should be noted that Ibragimov and Rozanov (1978) say that (4.8) decreases in n when it in fact increases. We thus have an approach analogous to the one developed by Stein (1990) for a misspecified mean function for bounding the efficiency of linear predictions under an incorrect model compatible with the truth. The main problem in obtaining explicit bounds from (4.9) is that F_0 and F_1 implicitly define $\Psi(\lambda,\mu)$, and we need to know $\Psi(\lambda,\mu)$, or at least some of its properties, to bound $M - M_n$.

4.3. *Proofs.* In this section, we give proofs of Propositions 1 and 2. The proof of Proposition 1 makes use of an explicit formula for $\Psi(\lambda,\mu)$. Suppose $R = [0,1]$ and f_0 and f_1 satisfy

$$f_i(\lambda) \asymp (1+\lambda^2)^{-p}, \tag{4.10}$$

for some positive integer p, where $a(\lambda) \asymp b(\lambda)$ means there exist constants c_1 and c_2 such that

$$0 < c_1 \le a(\lambda)/b(\lambda) \le c_2 < \infty,$$

for all λ. Generalizing (1.3) on page 30 of Ibragimov and Rozanov (1978), it can then be shown that $L_{R\times R}(F_0 \times F_1)$ coincides with the class of functions

$$\begin{aligned}
&\sum_{j,k=0}^{p-1} \alpha_{jk}(i\lambda)^j(i\mu)^k + (1+i\lambda)^p \sum_{j=0}^{p-1}(i\mu)^j\int_0^1 \beta_j(u)e^{i\lambda u}\,du \\
&\quad + (1-i\mu)^p\sum_{j=0}^{p-1}(i\lambda)^j\int_0^1\gamma_j(v)e^{-i\mu v}\,dv \\
&\quad + (1+i\lambda)^p(1-i\mu)^p\int_0^1\int_0^1\delta(u,v)e^{i(\lambda u-\mu v)}\,du\,dv,
\end{aligned} \tag{4.11}$$

where the α_{jk}'s are real constants, the β_j's and γ_j's are real and square integrable on $[0,1]$, and $\delta(u,v)$ is real and square integrable on $[0,1]^2$. Then

$\Psi(\lambda, \mu)$ as defined in (4.7) must be of this form since it is an element of $L_{R\times R}(F_0 \times F_1)$.

Even under these additional conditions, it is still, in general, very difficult to obtain an explicit expression for $\Psi(\lambda, \mu)$. We now develop the form of such an expression when $f_0(\lambda) = (a^2 + \lambda^2)^{-p}$ and $f_1(\lambda) = (b^2 + \lambda^2)^{-p}$. Note that K_0 and K_1 are compatible on any bounded interval by (1.1). It will be convenient to rewrite (4.11) in the slightly different form

$$\begin{aligned}&\sum_{j,k=0}^{p-1} \alpha_{jk}(a+i\lambda)^j(b-i\mu)^k + (a+i\lambda)^p \sum_{j=0}^{p-1}(b-i\mu)^j \int_0^1 e^{i\lambda u}\beta_j(u)\,du\\ &+(b-i\mu)^p \sum_{j=0}^{p-1}(a+i\lambda)^j \int_0^1 e^{-i\mu v}\gamma_j(v)\,dv \qquad (4.12)\\ &+(a+i\lambda)^p(b-i\mu)^p \int_0^1\int_0^1 \delta(u,v)e^{i(\lambda u-\mu v)}\,du\,dv,\end{aligned}$$

which we can also show coincides with $L_{R\times R}(F_0 \times F_1)$ when $\alpha_{jk}, \beta_j, \gamma_j$ and δ satisfy the same conditions as before. To determine $\delta(u, v)$, substitute (4.12) for $\Psi(\lambda, \mu)$ in (4.7) and take $(a + \partial/\partial s)^p(b + \partial/\partial t)^p$ of both sides of (4.7). The left-hand side gives

$$\begin{aligned}&\left(a+\frac{\partial}{\partial s}\right)^p\left(b+\frac{\partial}{\partial t}\right)^p \int\left\{\frac{1}{(a^2+\lambda^2)^p}-\frac{1}{(b^2+\lambda^2)^p}\right\}e^{i\lambda(s-t)}\,d\lambda\\ &\qquad = \int \frac{(b^2+\lambda^2)^p-(a^2+\lambda^2)^p}{(a-i\lambda)^p(b+i\lambda)^p}e^{i\lambda(s-t)}\,d\lambda,\end{aligned} \qquad (4.13)$$

by differentiating inside the integral, which is justified by

$$\int \lambda^{2p}\left|\frac{1}{(a^2+\lambda^2)^p}-\frac{1}{(b^2+\lambda^2)^p}\right|d\lambda < \infty.$$

For $s > t$, (4.13) equals

$$\begin{aligned}&\frac{2\pi i^{-p+1}}{(p-1)!}\frac{d^{p-1}}{d\lambda^{p-1}}\left[e^{i\lambda(s-t)}\frac{(b^2+\lambda^2)^p-(a^2+\lambda^2)^p}{(a-i\lambda)^p}\right]\Bigg|_{\lambda=ib}\\ &\qquad = -2\pi\sum_{j=0}^{p-1}\binom{p}{j+1}\frac{1}{j!}(s-t)^j(a-b)^{j+1}e^{-b(s-t)}.\end{aligned}$$

A similar result holds for $s < t$, and we obtain that (4.13) equals

$$\begin{aligned}&2\pi(b-a)\sum_{j=0}^{p-1}\binom{p}{j+1}\frac{1}{j!}\{(a-b)(s-t)\}^j\\ &\qquad\times\left\{e^{-b(s-t)}I_{\{s>t\}}+e^{-a(t-s)}I_{\{s<t\}}\right\}.\end{aligned} \qquad (4.14)$$

Applying $(a + \partial/\partial s)^p(b + \partial/\partial t)^p$ to the right-hand side of (4.7) and interchanging integration and differentiation, we just get $\delta(s, t)$. This interchange can be justified along the lines of the argument on pages 97 and 98 of

Ibragimov and Rozanov (1978). Thus, $\delta(s,t)$ is given by (4.14). Applying $(a + \partial/\partial s)^p$ to both sides of (4.7), through lengthy calculations, it is possible to show that $\beta_j(u)$ has the form

$$\sum_{m=0}^{p-1} \beta_{jm} u^m e^{-bu},$$

where β_{jm}'s are real constants. Applying $(b + \partial/\partial t)^p$ to both sides of (4.7), it can further be shown that $\gamma_j(v)$ has the form

$$\sum_{m=0}^{p-1} \gamma_{jm} v^m e^{-av},$$

where γ_{jm}'s are real constants.

We now turn to the proof of Proposition 1. The element in $L_R(F_0)$ corresponding to $z^{(k)}(t_{jn})$ is $(i\lambda)^k e^{i\lambda t_{jn}}$. Let $L_{T_n, p-1}$ be the subspace of $L_{R\times R}(F_0 \times F_1)$ generated by functions of the form

$$(i\lambda)^j e^{i\lambda t_{ln}} (-i\mu)^k e^{-i\mu t_{mn}}.$$

By (4.9), we can bound $M - M_{T_n, p-1}$ by approximating $\Psi(\lambda, \mu)$ with an element of $L_{T_n, p-1}$. Following Stein (1990),

$$(a + i\lambda)^p (b - i\mu)^p \sum_{j,k=1}^{n} \sum_{l,m=0}^{p-1} c_{jklm} \int_0^{t_{jn}} \int_0^{t_{kn}} s^l t^m e^{(a+i\lambda)s + (b-i\mu)t}\, ds\, dt \in L_{T_n, p-1},$$

for c_{jklm}'s real constants. Thus, letting $g(s,t) = \delta(s,t) e^{-(as+bt)}$,

$$\begin{aligned}
&\int \left| (a + i\lambda)^p (b - i\mu)^p \int_0^1 \int_0^1 \delta(s,t) e^{i(\lambda s - \mu t)}\, ds\, dt \right.\\
&\qquad \left. - (a + i\lambda)^p (b - i\mu)^p \sum_{j,k=1}^{n} \sum_{l,m=0}^{p-1} c_{jklm} \int_0^{t_{jn}} \int_0^{t_{kn}} s^l t^m e^{(a+i\lambda)s+(b-i\mu)t}\, ds\, dt \right|^2 \\
&\qquad \times (a^2 + \lambda^2)^{-p} (b^2 + \mu^2)^{-p}\, d\lambda\, d\mu \\
&= \int \left| \int_0^1 \int_0^1 e^{i(\lambda s - \mu t)} \right. \\
&\qquad \left. \times \left[\delta(s,t) - \sum_{j,k=1}^{n} \sum_{l,m=0}^{p-1} c_{jklm} I_{\{s \le t_{jn}, t \le t_{kn}\}} s^l t^m e^{as+bt} \right] ds\, dt \right|^2 d\lambda\, d\mu \\
&= 4\pi^2 \int_0^1 \int_0^1 e^{2(as+bt)} \left[g(s,t) - \sum_{j,k=1}^{n} \sum_{l,m=0}^{p-1} c_{jklm} I_{\{s \le t_{jn}, t \le t_{kn}\}} s^l t^m \right]^2 ds\, dt \\
&= 4\pi^2 \sum_{j,k=1}^{n} \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{k-1,n}}^{t_{kn}} e^{2(as+bt)} \left[g(s,t) - \sum_{l,m=0}^{p-1} \sum_{j'=j}^{n} \sum_{k'=k}^{n} c_{j'k'lm} s^l t^m \right]^2 ds\, dt,
\end{aligned} \tag{4.15}$$

where the second equality is by Parseval's relation. For real constants b_{jklm}, we can choose the c_{jklm}'s such that

$$\sum_{l,m=0}^{p-1} \sum_{j'=j}^{n} \sum_{k'=k}^{n} c_{j'k'lm} s^l t^m = \sum_{l,m=0}^{p-1} b_{jklm}(s - t_{jn})^l (t - t_{kn})^m,$$

for $j, k = 1, \dots, n$. Let us first consider those terms in (4.15) for which $j = k$. Setting $b_{jjlm} = 0$ if $l > 0$ or $m > 0$, letting $b_{jj00} = g(t_{jn}, t_{jn})$, and using (4.1),

$$\sum_{j=1}^{n} \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{j-1,n}}^{t_{jn}} e^{2(as+bt)}\big(g(s,t) - g(t_{jn}, t_{jn})\big)^2 \, ds\, dt = O(n^{-3}).$$

Because $\delta(s,t) = \delta(s - t)$ has a discontinuity in its first derivative at $s - t = 0$, it does not appear possible to improve on this rate. For $j \neq k$, let $b_{jklm} = 0$ for $l + m \geq p$, and for $l + m \leq p - 1$,

$$b_{jklm} = \frac{1}{l!m!} \frac{\partial^l}{\partial s^l} \frac{\partial^m}{\partial t^m} g(s,t) \bigg|_{(t_{jn}, t_{kn})}.$$

It follows that

$$\sum_{j \neq k} \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{k-1,n}}^{t_{kn}} e^{2(as+bt)} \left[g(s,t) - \sum_{l,m=0}^{p-1} b_{jklm}(s - t_{jn})^l (t - t_{jn})^m \right]^2 ds\, dt$$
$$= O(n^{-2p}).$$

Thus, (4.15) is $O(n^{-\min(3,2p)})$.

Now consider approximating a term like $(a + i\lambda)^p (b - i\mu)^j \int_0^1 e^{i\lambda s} \beta_j(s)\, ds$ for $j = 0, \dots, p - 1$. Because we have observed $z(0), \dots, z^{(p-1)}(0)$, we can obtain $(b - i\mu)^j$ exactly, so we want to choose constants c_{jk} such that

$$\int \bigg| (b - i\mu)^j (a + i\lambda)^p \int_0^1 e^{i\lambda s} \beta_j(s)\, ds - (b - i\mu)^j (a + i\lambda)^p$$
$$\times \sum_{j=1}^{n} \sum_{k=0}^{p-1} c_{jk} \int_0^{t_{jn}} s^l e^{(a+i\lambda)s}\, ds \bigg|^2 (a^2 + \lambda^2)^{-p} (b^2 + \mu^2)^{-p}\, d\lambda\, d\mu$$
$$= O\left(\int \left| \int_0^1 e^{i\lambda s} \left[\beta_j(s) - \sum_{j=1}^{n} \sum_{k=0}^{p-1} c_{jk} I_{\{s \leq t_{jn}\}} s^l e^{as} \right] ds \right|^2 d\lambda \right)$$

is small. Using Theorem 4.1 from Stein (1990) and the fact that $\beta_j(s)e^{-as}$ has a square integrable pth derivative on [0, 1], this last expression can be made $O(n^{-2p})$. Since the terms $(a + i\lambda)^j (b - i\mu)^k$ for $j, k = 0, \dots, p - 1$ are in $L_{T_n, p-1}$,

$$M - M_{T_n, p-1} = O(n^{-\min(3,2p)}), \tag{4.16}$$

and Proposition 1 follows from (4.9) and (2.3).

In obtaining (4.16), the two factors that control the rate of convergence are p and the smoothness of $\delta(s,t)$. For f_0 and f_1 satisfying (4.10) for some positive integer p, the smoothness of $\delta(s,t)$ should be closely related to the relative difference between $f_0(\lambda)$ and $f_1(\lambda)$ for large λ. Suppose for a constant $\nu > \frac{1}{2}$,

$$0 < \liminf_{\lambda \to \infty} \frac{f_0(\lambda) - f_1(\lambda)}{f_0(\lambda)} \lambda^{\nu} \le \limsup_{\lambda \to \infty} \frac{f_0(\lambda) - f_1(\lambda)}{f_0(\lambda)} \lambda^{\nu} < \infty.$$

Then if f_0 and f_1 satisfy (4.10) and $\{T_n\}$ satisfies (4.1), I conjecture the following generalization of (4.16):

$$M - M_{T_n, p-1} = O\Big(\max\big(n^{-2p}, n^{-2\nu+1}(\log n)^{2g(\nu)}\big)\Big),$$

where $g(\nu) = 1$ if ν is an odd integer and is 0 otherwise. Note that if $\nu \le \frac{1}{2}$, K_0 and K_1 are not compatible.

We next prove Proposition 2. By (1.1), (4.3) guarantees that the corresponding covariance functions are compatible [Ibragimov and Rozanov (1978), page 105]. To obtain an upper bound on $|b_{T_n, p-1}(h)|$, it is equivalent to find an upper bound on $|E_1 h^2 - E_0 h^2|$ among those $h \in H(K_0)$ that are uncorrelated with the $p(n+1)$ observations and have variance 1 under f_0. In the isomorphic space $L_R(F_0)$, all elements are of the form [Ibragimov and Rozanov, (1978), page 30]

$$\sum_{j=0}^{p-1} c_j (i\lambda)^j + (a + i\lambda)^p \int_0^1 c(u) e^{i\lambda u}\, du,$$

where $c(u)$ is a square integrable on [0, 1]. Since $(i\lambda)^j \in L_{T_n, p-1}$, we can take $c_j = 0$ for all j. Furthermore, for $m = 0, \dots, p-1$,

$$\left\langle (a + i\lambda)^p \int_0^1 c(u) e^{i\lambda u}\, du, (i\lambda)^m e^{i\lambda t_{jn}} \right\rangle_{F_0}$$

$$= 2\pi \sum_{k=0}^{m} \binom{m}{k} \frac{(-a)^{m-k}}{(p-k-1)!} \int_0^{t_{jn}} c(u) e^{a(u - t_{jn})} (t_{jn} - u)^{p-k-1}\, du$$

and

$$\left\| (a + i\lambda)^p \int_0^1 e^{i\lambda u} c(u)\, du \right\|_{F_0}^2 = 2\pi \int_0^1 c(u)^2\, du,$$

where $\|\varphi(\lambda)\|_{F_0}^2 = \langle \varphi(\lambda), \varphi(\lambda) \rangle_{F_0}$. Thus, the elements of $H(K_0)$ that are uncorrelated with the observations and have variance 1 under f_0 correspond to those functions $c(u)$ satisfying

$$\int_{t_{j-1,n}}^{t_{jn}} c(u) e^{au} u^k\, du = 0 \quad \text{for } j = 1, \dots, n;\ k = 0, \dots, p-1, \tag{4.17}$$

and $2\pi \int_0^1 c(u)^2\, du = 1$. Define C_n to be the set of functions $c(\cdot)$ satisfying these conditions. We have

$$\left\|(a+i\lambda)^p\int_0^1 c(u)e^{i\lambda u}\,du\right\|_{F_1}^2-\left\|(a+i\lambda)^p\int_0^1 c(u)e^{i\lambda u}\,du\right\|_{F_0}^2$$

$$=\int_{-\infty}^{\infty}\frac{f_1(\lambda)-f_0(\lambda)}{f_0(\lambda)}\int_0^1 c(u)e^{i\lambda u}\,du\int_0^1 c(u)e^{i\lambda u}\,du\,d\lambda$$

$$=\int_0^1\int_0^1 c(u)c(v)\int_{-\infty}^{\infty}\frac{f_1(\lambda)-f_0(\lambda)}{f_0(\lambda)}e^{i\lambda(u-v)}\,d\lambda\,du\,dv,$$

where the order of integration can be changed because of the absolute integrability of the integrand. Now, since the Fourier transform of a positive integrable function is a positive definite function, we have by (4.3)

$$0\le\int_0^1\int_0^1 c(u)c(v)\left\{\int_{f_1(\lambda)>f_0(\lambda)}\frac{f_1(\lambda)-f_0(\lambda)}{f_0(\lambda)}e^{i\lambda(u-v)}\,d\lambda\right\}du\,dv$$

$$\le C\int_0^1\int_0^1 c(u)c(v)\left\{\int_{f_1(\lambda)>f_0(\lambda)}(1+\lambda^2)^{-q}e^{i\lambda(u-v)}\,d\lambda\right\}du\,dv.$$

Making a similar calculation on $f_1(\lambda)<f_0(\lambda)$, we can obtain for $c\in C_n$,

$$\left|\left\|(a+i\lambda)^p\int_0^1 c(u)e^{i\lambda u}\,du\right\|_{F_1}^2-\left\|(a+i\lambda)^p\int_0^1 c(u)e^{i\lambda u}\,du\right\|_{F_0}^2\right| \tag{4.18}$$

$$\le C\int_0^1\int_0^1 r(u-v)c(u)c(v)\,du\,dv,$$

where

$$r(t)=\int(1+\lambda^2)^{-q}e^{i\lambda t}\,d\lambda$$

$$=\frac{\pi e^{-|t|}}{2^{2(q-1)}}\sum_{k=0}^{q-1}\binom{2q-k-2}{q-1}\frac{(2|t|)^k}{k!}.$$

For $j<k$,

$$\int_{t_{j-1,n}}^{t_{jn}}\int_{t_{k-1,n}}^{t_{kn}} r(u-v)c(u)c(v)\,du\,dv$$

$$=\sum_{l+m=0}^{q-1}\alpha_{lm}\int_{t_{j-1,n}}^{t_{jn}}e^{-u}u^l c(u)\,du\int_{t_{k-1,n}}^{t_{kn}}e^{v}v^m c(v)\,dv,$$

where the α_{lm}'s are constants independent of n. We can choose constants

$\beta_{j0}, \dots, \beta_{j,q-1}$ such that

$$e^{-u}u^l - \sum_{m=0}^{q-1} \beta_{jm} e^{au} u^m = O(\Delta_{jn}^q)$$

on $(t_{j-1,n}, t_{jn})$, where $\Delta_{jn} = t_{jn} - t_{j-1,n}$, so that by (4.17)

$$\begin{aligned}
&\left| \int_{t_{j-1,n}}^{t_{jn}} e^{-u}u^l c(u)\,du \right| \\
&= \left| \int_{t_{j-1,n}}^{t_{jn}} \left(e^{-u}u^l - e^{au} \sum_{m=0}^{q-1} \beta_{jm} u^m \right) c(u)\,du \right| \\
&\le \left\{ \int_{t_{j-1,n}}^{t_{jn}} \left(e^{-u}u^l - e^{au} \sum_{m=0}^{q-1} \beta_{jm} u^m \right)^2 du \int_{t_{j-1,n}}^{t_{jn}} c(u)^2\,du \right\}^{1/2} \\
&= \left(\int_{t_{j-1,n}}^{t_{jn}} c(u)^2\,du \right)^{1/2} O\left(\Delta_{jn}^{(2q+1)/2}\right)
\end{aligned}$$

for $q \le p$. Thus, under (4.1) and for $c \in C_n$,

$$\begin{aligned}
&\left| \sum_{j \ne k} \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{k-1,n}}^{t_{kn}} r(u-v)c(u)c(v)\,du\,dv \right| \\
&= \sum_{j \ne k} \left\{ \int_{t_{j-1,n}}^{t_{jn}} c(u)^2\,du \int_{t_{k-1,n}}^{t_{kn}} c(v)^2\,dv \right\}^{1/2} O(n^{-(2q+1)}) \\
&= \left[\sum_{j=1}^{n} \left\{ \int_{t_{j-1,n}}^{t_{jn}} c(u)^2\,du \right\}^{1/2} \right]^2 O(n^{-(2q+1)}) \\
&= O(n^{-2q}),
\end{aligned} \tag{4.19}$$

since subject to $2\pi\int_0^1 c(u)^2\,du = 1$,

$$\sum_{j=1}^{n} \left\{ \int_{t_{j-1,n}}^{t_{jn}} c(u)^2\,du \right\}^{1/2}$$

is maximized by making all terms in the sum equal. In a neighborhood of the origin,

$$r(t) = \sum_{k=0}^{q-1} \gamma_k t^{2k} + \gamma_q |t|^{2q-1} + O(t^2 q),$$

for some constants $\gamma_0, \dots, \gamma_q$. It follows that for given j we can choose constants $a_0, \dots, a_{p-1}, b_0, \dots, b_{p-1}$ and integrable functions $g_1(\cdot)$ and $g_2(\cdot)$ such that

$$r(u-v) = e^{au} \sum_{l=0}^{p-1} a_l u^l g_1(v) + e^{av} \sum_{l=0}^{p-1} b_l v^l g_2(u) + O(|u-v|^{\min(p, 2q-1)}),$$

for $u, v \in [t_{j-1,n}, t_{jn}]$. Thus,

$$\begin{aligned}
&\left| \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{j-1,n}}^{t_{jn}} r(u-v) c(u) c(v)\, du\, dv \right| \\
&\quad = \left| \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{j-1,n}}^{t_{jn}} \left(r(u-v) - e^{au} \sum_{l=0}^{p-1} a_l u^l g_1(v) \right.\right. \\
&\qquad\qquad \left.\left. - e^{av} \sum_{l=0}^{p-1} b_l v^l g_2(u) \right) c(u) c(v)\, du\, dv \right| \\
&\quad \le \left\{ \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{j-1,n}}^{t_{jn}} \left(r(u-v) - e^{au} \sum_{l=0}^{p-1} a_l u^l g_1(v) \right.\right. \\
&\qquad\qquad \left.\left. - e^{av} \sum_{l=0}^{p-1} b_l v^l g_2(u) \right)^2 du\, dv \right\}^{1/2} \int_{t_{j-1,n}}^{t_{jn}} c(u)^2\, du \\
&\quad = \int_{t_{j-1,n}}^{t_{jn}} c(u)^2\, du \times O\left(\Delta_{jn}^{\min(2q, p+1)}\right),
\end{aligned}$$

so subject to (4.1), for $c \in C_n$,

$$\sum_{j=1}^{n} \left| \int_{t_{j-1,n}}^{t_{jn}} \int_{t_{j-1,n}}^{t_{jn}} r(u-v) c(u) c(v)\, du\, dv \right| = O(n^{-\min(2q, p+1)}). \tag{4.20}$$

(4.4) then follows from (4.18)–(4.20). It can be generalized to let $f_0(\lambda) = \prod_{j=1}^{p} (a_j^2 + \lambda^2)^{-1}$, where the a_j's are positive constants, which corresponds to the spectral density of an arbitrary continuous-time AR(p) process.

To prove (4.6), let us make the simplifying assumption that

$$t_{1n} \asymp n^{-1}, \tag{4.21}$$

although the following argument can be easily extended to include cases where (4.21) does not hold. Define

$$c_n(u) = \alpha_n e^{-au} P_p^*(u/t_{1n}) I_{\{u \le t_{1n}\}},$$

where $P_p^*(\cdot)$ is the shifted Legendre polynominal of order p [Abramowitz and Stegun (1965), page 774] and α_n is chosen so that $2\pi\int_0^1 c_n(u)^2\,du = 1$. Since $P_p^*(\cdot)$ is orthogonal to all polynominals of degree at most $p-1$ on $[0,1]$, $c_n(u) \in C_n$. Using (4.21), we can show $\alpha_n \asymp n^{1/2}$. Then

$$\left\|(a+i\lambda)^p\int_0^1 c_n(u)e^{i\lambda u}\,du\right\|_{F_1}^2 - \left\|(a+i\lambda)^p\int_0^1 c_n(u)e^{i\lambda u}\,du\right\|_{F_0}^2$$

$$\geq B\int_0^{t_{1n}}\int_0^{t_{1n}} r(u-v)c_n(u)c_n(v)\,du\,dv$$

$$(4.22)\qquad = B\int_0^{t_{1n}}\int_0^{t_{1n}}\left[\sum_{k=0}^{q-1}\gamma_k(u-v)^{2k} + \gamma_q|u-v|^{2q-1} + O(|u-v|^{2q})\right] \times c_n(u)c_n(v)\,du\,dv$$

$$= B\gamma_q\int_0^{t_{1n}}\int_0^{t_{1n}}|u-v|^{2q-1}c_n(u)c_n(v)\,du\,dv + O(n^{-2q-1}),$$

using (4.5), (4.17) and (4.21). Now,

$$B\gamma_q\int_0^{t_{1n}}\int_0^{t_{1n}}|u-v|^{2q-1}c_n(u)c_n(v)\,du\,dv$$

$$= \alpha_n^2 t_{1n}^{2q+1}\gamma_q\int_0^1\int_0^1|u-v|^{2q-1}P_p^*(u)P_p^*(v)\,du\,dv + o\left(\alpha_n^2 t_{1n}^{2q+1}\right).$$

Using the fact that $(-1)^q|t|^{2q-1}$ is a generalized covariance function of order $q-1$ [Matheron (1973)], it follows that

$$(4.23)\qquad (-1)^q\int_0^1\int_0^1|u-v|^{2q-1}P_p^*(u)P_p^*(v)\,du\,dv \geq 0,$$

since P_p^* is in the space Λ_{q-1} as defined by Matheron (1973). In fact, using the spectral representation of the generalized covariance function $(-1)^q|t|^{2q-1}$, the left-hand side of (4.23) can be written as

$$4\int_0^\infty (4\pi^2\rho^2)^{-q}\left|\int_0^1 P_p^*(u)e^{2\pi i\rho u}\,du\right|^2 d\rho,$$

which is clearly positive. Finally, $(-1)^q\gamma_q > 0$, so (4.6) follows from (4.22) and $\alpha_n^2 t_{1n}^{2q+1} \asymp n^{-2q}$.

5. AR(1) process. The spectral density $(a^2+\lambda^2)^{-p}$ corresponds to a continuous-time AR(p) process. When $p=1$, we can write down the inverse of the covariance matrix of the observations, which enables us to sharpen the calculations in the previous section. That is, let V_{T_n} be the $(n+1)\times(n+1)$ covariance matrix of the observation set T_n under the covariance function $K_0(s,t) = a^{-1}\exp\{-a|s-t|\}$, which corresponds to $f_0(\lambda) = 1/\{\pi(a^2+\lambda^2)\}$.

Setting $\rho_i = \exp(-a\Delta_{in})$, $V_{T_n}^{-1} = (v^{ij})$ is the tridiagonal matrix defined by

$$a^{-1}v^{ii} = \begin{cases} \left(1-\rho_1^2\right)^{-1}, & i = 1, \\ \left(1-\rho_n^2\right)^{-1}, & i = n+1, \\ \dfrac{1-\rho_{i-1}^2\rho_i^2}{\left(1-\rho_{i-1}^2\right)\left(1-\rho_i^2\right)}, & i = 2,\dots,n, \end{cases}$$

$$v^{i+1,i} = v^{i,i+1} = -a\rho_i\left(1-\rho_i^2\right)^{-1},$$

and all other elements 0, which can be verified by direct calculation. Furthermore,

$$\log\left|V_{T_n}\right| = -(n+1)\log a + \sum_{i=1}^{n} \log\left(1-\rho_i^2\right),$$

which can be obtained by row reducing $V_{T_n}^{-1}$ to make it upper triangular and then taking the product of the resulting diagonal elements. Define $\rho_i' = \exp(-b\Delta_{in})$ and $I_{T_n}(P_0, P_1)$ to be the Kullback divergence between the two Gaussian distributions of the random vector $(z(t_{0n}), \dots, z(t_{nn}))$ with zero mean and covariance functions $K_0(t) = a^{-1}e^{-a|t|}$ and $K_1(t) = b^{-1}e^{-b|t|}$, respectively. Using the above expressions for $V_{T_n}^{-1}$ and $|V_{T_n}|$, it follows that

$$I_{T_n}(P_0, P_1) = \frac{n+1}{2}\left(\log\frac{a}{b} - 1 + \frac{b}{a}\right) - \frac{1}{2}\sum_{i=1}^{n}\log\frac{1-\rho_i^2}{1-\rho_i'^2} + \frac{b}{a}\sum_{j=1}^{n}\frac{\rho_j'(\rho_j'-\rho_j)}{1-\rho_j'^2}.$$

Taking Taylor series, we obtain after lengthy calculations,

$$I_{T_n}(P_0, P_1) = \frac{1}{2}\log\frac{a}{b} + \frac{(a-b)^2}{4a} + \frac{b-a}{2a} - \frac{\left(a^2-b^2\right)^2}{48a}\sum_{i=1}^{n}\Delta_{in}^3\left(1 + O(\Delta_{in})\right). \tag{5.1}$$

It follows that

$$I(P_0, P_1) = \frac{1}{2}\log\frac{a}{b} + \frac{(a-b)^2}{4a} + \frac{b-a}{2a},$$

which is given by Kullback, Keegel and Kullback (1987), page 74.

Let $M_{T_n} = M_{T_n,0}$ and $a_{T_n} = a_{T_n,0}$. From (5.1), we have

$$M_{T_n} = \frac{(b-a)^2(2+a+b)}{2ab} - \frac{(b-a)^2(b+a)^3}{24ab}\sum_{j=1}^{n}\Delta_{jn}^3\left(1 + O(\Delta_{jn})\right),$$

so that

$$M - M_{T_n} = \frac{(b-a)^2(b+a)^3}{24ab} \sum_{j=1}^{n} \Delta_{jn}^3\big(1 + O(\Delta_{jn})\big),$$

which is $O(n^{-2})$ if (4.1) is satisfied, so that (4.16) gives the optimal rate of convergence for $M - M_{T_n, p-1}$ when $p = 1$. Furthermore, the bound in Proposition 1 on $a_{T_n}(h)$ is of the optimal rate when $K_0(t) = a^{-1}e^{-a|t|}$ and $K_1(t) = b^{-1}e^{-b|t|}$. Let

$$h_n = \sum_{j=1}^{n} z(\tau_{jn}),$$

where $\tau_{jn} = (t_{j-1,n} + t_{jn})/2$, in which case,

$$e_0(h_n, n) = \sum_{j=1}^{n} \left[z(\tau_{jn}) - \tfrac{1}{2}\operatorname{sech}\big(\tfrac{1}{2}a\Delta_j\big)\big(z(t_{j-1,n}) + z(t_{jn})\big)\right].$$

Using Taylor series, it can then be shown that under (4.1),

$$\frac{E_0\big(e_1(h_n, n) - e_0(h_n, n)\big)^2}{E_0 e_0(h_n, n)^2} \asymp n^{-2},$$

so that the bound in (4.2) is sharp when $p = 1$.

Finally, we consider an application of (5.1) to a problem in the design of time series experiments. Sacks and Ylvisaker (1966, 1968, 1970) consider the design problem of finding asymptotically optimal choices for T_n in terms of minimizing the variances of estimates of regression coefficients. For example, suppose $Ez(x) = \beta f(x)$, where $f(\cdot)$ is specified and β is unknown, and $\operatorname{cov}(z(x), z(x'))$ is assumed known. Let $\mathscr{D}_n$ be the collection of all sets $T_n = (t_{0n}, \dots, t_{nn})$ satisfying $0 = t_{0n} < \cdots < t_{nn} = 1$. Let $\sigma_{T_n}^2$ be the variance of the generalized least squares estimate β based on observing $z(\cdot)$ on T_n. If $f(\cdot)$ is sufficiently smooth, then the variance of the best linear unbiased estimate of β based on observing $z(\cdot)$ everywhere on $[0,1]$ is positive, and we will denote this variance by σ^2. Then Sacks and Ylvisaker (1966, 1968, 1970) consider choosing T_n so that

$$\lim_{n\to\infty} \frac{\sigma^{-2} - \sigma_{T_n}^{-2}}{\sigma^{-2} - \sup_{T\in\mathscr{D}_n} \sigma_T^{-2}} = 1.$$

For discriminating between two possible covariance functions K_0 and K_1 for a Gaussian process $z(\cdot)$, we might analogously consider choosing T_n so that

$$\lim_{n\to\infty} \frac{I(P_0, P_1) - I_{T_n}(P_0, P_1)}{I(P_0, P_1) - \sup_{T\in\mathscr{D}_n} I_T(P_0, P_1)} = 1. \tag{5.2}$$

A sequence of designs $T_1, T_2, \dots$ satisfying (5.2) will be called asymptotically optimal. It is easy to show that letting $t_{jn} = j/n$ yields an asymptotically optimal design when $K_0(t) = a^{-1}e^{-a|t|}$ and $K_1(t) = b^{-1}e^{-b|t|}$. This can be seen by first noting that $\max_j \Delta_{jn} \to 0$ is a necessary and sufficient condition for $I_{T_n}(P_0, P_1) \to I(P_0, P_1)$, and that if this condition is satisfied,

$$\sum_{j=1}^{n} \Delta_{jn}^3\big(1 + O(\Delta_{jn})\big) = (1 + o(1)) \sum_{j=1}^{n} \Delta_{jn}^3.$$

The result then follows by observing that $\Delta_{1n}^3 + \cdots + \Delta_{nn}^3$ is minimized by making the Δ_{jn}'s equal.

Acknowledgments. I would like to thank the referees and associate editor for many helpful suggestions in improving the organization of this paper.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions*, 9th ed. Dover, New York.

BLACKWELL, D. and DUBINS, L. E. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33** 882–886.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes* (translated by A. B. Aries). Springer, New York.

KULLBACK, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory* **IT-13** 765–766.

KULLBACK, S., KEEGEL, J. C. and KULLBACK, J. H. (1987). *Topics in Statistical Information Theory. Lecture Notes in Statist.* **42**. Springer, New York.

KUO, H. (1975). *Gaussian Measures in Banach Spaces. Lecture Notes in Math.* **463**. Springer, New York.

MATHERON, G. (1973). The intrinsic random functions and their applications. *Adv. in Appl. Probab.* **5** 439–468.

MATHERON, G. (1984). The selectivity of the distributions, and the second principle of geostatistics. In *Proceedings, 2nd NATO A. S. I., Geostatistics for Natural Resources Characterization* (G. Verly, ed.) 421–434. Reidel, Dordrecht.

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors *Ann. Math. Statist.* **37** 66–89.

SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors; Many parameters. *Ann. Math. Statist.* **39** 49–69.

SACKS, J. and YLVISAKER, D. (1970). Designs for regression problems with correlated errors III. *Ann. Math. Statist.* **41** 2057–2074.

STARKS, T. H. and SPARKS, A. R. (1987). Rejoinder to "Comment on 'Estimation of the generalized covariance function II. A response surface approach' by Thomas H. Starks and Allen R. Sparks." *Math. Geol.* **19** 789–792.

STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63.

STEIN, M. L. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 839–861.

YADRENKO, M. I. (1983). *Spectral Theory of Random Fields*. Optimization Software, New York.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637