ON LINEAR ESTIMATION FOR REGRESSION PROBLEMS ON TIME SERIES[1]

By N. Donald Ylvisaker

University of Washington

1. Introduction. The purpose of this paper is to discuss, in a general context, certain mean value problems of a single parameter nature. More specifically, suppose one observes the family $\{Y(t),\ t \varepsilon T\}$ of random variables satisfying $Y(t) = m(t, \beta) + X(t)$, $t \varepsilon T$, where β is an unknown parameter lying in some set Λ. Suppose further that

$$E_\beta X(t) \equiv 0, \qquad t \varepsilon T, \qquad \beta \varepsilon \Lambda,$$

$$E_\beta X(s)X(t) = K(s, t), \qquad s, t \varepsilon T, \qquad \beta \varepsilon \Lambda.$$

(Assumptions will be imposed on the set T and the kernel K where necessary.) The problem of interest is that of estimating, on the basis of the observation, a real valued function g defined on Λ under the criterion of squared error loss. This model emanates from regression analysis on time series and our attention is focussed in this direction. The estimators concerned will be linear only and we shall be interested in problems for which the mean value function is not essentially linear in β.

The results obtained proceed from the application of results from the area of reproducing kernel spaces. The statistical results include a specification of problems in which linear estimation makes sense, a precise lower bound for the risk function of a linear estimator, and a characterization of those problems which admit consistent (zero risk) linear estimators.

As previously mentioned, the tools arise in the theory of reproducing kernel spaces and consequently we begin, in Section 2, by listing the appropriate results of this discipline. Section 3 is devoted to the properties of linear estimators, and Section 4 is given over to examples and remarks.

As few assumptions are imposed as is feasible and specific problems are introduced only for purposes of illustration.

This work leans heavily, for its emphasis, on the work of Parzen, [4] and [5], in this same vein, as will frequently be noted in the sequel.

2. Reproducing kernel spaces. The theory of reproducing kernel spaces has received an extensive exposition in the paper of Aronszajn [1]. The purpose of this section will be to state succinctly the apparatus necessary to accomplish the next section. For a much broader discussion of the role of this theory in the area of time series analysis, the interested reader can consult [4] and [5].

Received February 9, 1962.

[1] This research was sponsored by the Office of Naval Research under contract Nonr 266 (59), Project Number NR 042-205 while the author was at Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Let T denote an abstract set and let $\mathcal{K}(T)$ denote the class of real, positive definite kernels defined on $T \times T$, equivalently, $\mathcal{K}(T)$ is the class of real covariance kernels on $T \times T$. It is easily verified that $\mathcal{K}(T)$ is closed with respect to pointwise addition, multiplication, and passage to limits. A partial ordering of $\mathcal{K}(T)$ is accomplished by the relation $\ll$ defined by

$$K_1 \ll K_2 \quad \text{if and only if} \quad K_2 - K_1 \in \mathcal{K}(T).$$

To each $K \in \mathcal{K}(T)$ there corresponds a real Hilbert space of functions defined on T called the reproducing kernel space with reproducing kernel K. This kernel space will be denoted by $H(K,T)$ and is determined by the conditions

(i) $K(\cdot, t) \in H(K,T)$ for each $t \in T$,

(ii) For each $f \in H(K,T)$ and each $t \in T$, $(f, K(\cdot, t))_K = f(t)$.

(We shall persist in labelling the inner product according to the kernel K associated with the kernel space in question.) Note that the family of functions $\{K(\cdot, t), t \in T\}$ is complete in $H(K,T)$. We shall assume throughout that $H(K,T)$ is a separable Hilbert space, this covering all normal situations. We proceed directly to the general theorems which will be of use in this study.

The first theorem forms the basis for applying the kernel space results in the next section.

THEOREM 2.1. *If $\{f_t, t \in T\}$ is a family of elements in a Hilbert space H, then the span of $\{f_t, t \in T\}$ in H, $V[f_t, t \in T]$, is isomorphic to the reproducing kernel space $H(K,T)$, where $K(s,t) = (f_s, f_t)_H$, $s, t \in T$, under the mapping Ψ for which, $\{\Psi f\}(t) = (f, f_t)_H$.*

PROOF. By assumption, $(f_s, f_t)_H = (K(\cdot, s), K(\cdot, t))_K$ and consequently, by the basic congruence theorem of [5], there is an isomorphism between $V[f_t, t \in T]$ in H and $H(K,T)$ for which $\{\Psi f_t\}(\cdot) = K(\cdot, t)$. Then, for $f \in V[f_t, t \in T]$,

$$(f, f_t) = (\Psi f, \Psi f_t) = (\Psi f, K(\cdot, t))_K = \{\Psi f\}(t),$$

which is the required result.

We shall require the effect on the kernel spaces obtained under addition of kernels and the relationship between two kernel spaces where one kernel dominates the other in the sense of the partial ordering $\ll$. Proofs of these two theorems may be found in [1].

THEOREM 2.2 *If K_1 and $K_2 \in \mathcal{K}(T)$, then $K = K_1 + K_2 \in \mathcal{K}(t)$ (pointwise addition) and $H(K,T)$ consists of all functions $f = f_1 + f_2$, $f_i \in H(K_i, T)$, with*

$$\| f \|_K^2 = \min_{f_1+f_2=f}[\| f_1 \|_{K_1}^2 + \| f_2 \|_{K_2}^2].$$

THEOREM 2.3. *If K_1 and $K_2 \in \mathcal{K}(T)$, $H(K_1, T) \subset H(K_2, T)$, if and only if $\exists\, B > 0 \ni K_1 \ll BK_2$.* Note that the inclusion relation in the preceding theorem is between classes of functions.

A slightly altered version of Theorem 2.3 is given next. A proof is supplied as the theorem is not stated explicitly in [1].

THEOREM 2.4. *If K_1 and $K_2 \in \mathcal{K}(T)$, then $H(K_1, T) \subset H(K_2, T)$, if and only if $\{\sum_{j=1}^{N(n)} c_{jn} K_2(\cdot, t_{jn})\}$ is a Cauchy sequence in $H(K_2, T)$ implies that*

$$\left\{\sum_{j=1}^{N(n)} c_{jn}K_1(\cdot, t_{jn})\right\}$$

is a Cauchy sequence in $H(K_1, T)$.

PROOF. The necessity follows directly from Theorem 2.2. Suppose then that $\{f_{2n}\} = \{\sum_{j=1}^{N(n)} c_{jn}K_2(\cdot, t_{jn})\}$ is a Cauchy sequence in $H(K_2, T)$. By hypothesis $\{f_{1n}\} = \{\sum_{j=1}^{N(n)} c_{jn}K_1(\cdot, t_{jn})\}$ is a Cauchy sequence in $H(K_1, T)$. Let g be an element of $H(K_1, T)$ which is not in $H(K_2, T)$. By the reproducing property (ii) and weak convergence

$$(g, f_{1n})_{K_1} = \sum_{j=1}^{N(n)} c_{jn}g(t_{jn}) \to \text{limit as } n \to \infty.$$

Now $\{f_{2n}\}$ may be chosen with limit the zero function and with lim $\sum_{j=1}^{N(n)} c_{jn}g(t_{jn}) = a > 0$, for otherwise

$$\phi(f) = \phi(\text{l.i.m.}\, f_{2n}) = \phi\left(\text{l.i.m.} \sum_{j=1}^{N(n)} c_{jn}\, K_2(\cdot, t_{jn})\right)$$
$$= \lim \sum_{j=1}^{N(n)} c_{jn}\, g(t_{jn})$$

is a bounded linear functional on $H(K_2, T)$ with $\phi(K_2(\cdot, t)) = g(t)$, a contradiction. For this choice of $\{f_{2n}\}$, let

$$h_{2n} = \begin{cases} f_{2n} & n \text{ odd} \\ 2f_{2n} & n \text{ even.} \end{cases}$$

$\{h_{2n}\}$ is a Cauchy sequence in $H(K_2,T)$ tending to the zero function and for the corresponding Cauchy sequence $\{h_{1n}\}$ in $H(K_1, T)$,

$$\liminf (g,h_{1n})_{K_1} = a, \qquad \limsup\, (g, h_{1n})_{K_1} = 2a,$$

a contradiction.

3. Linear estimates. It is assumed that an observation is taken on the family $\{Y(t),\ t \varepsilon T\}$ of real random variables of the form $Y(t) = m(t,\beta) + X(t)$, $t \varepsilon T$, $\beta \varepsilon \Lambda$, with

$$E_\beta X(t) \equiv 0, \qquad t \varepsilon T, \qquad \beta \varepsilon \Lambda$$
$$E_\beta X(s)X(t) = K(s,t), \qquad s,t \varepsilon T, \qquad \beta \varepsilon \Lambda.$$

Given a real valued function g defined on Λ, we consider here linear estimation of g subject to squared error loss. The first aim will be to establish conditions under which the problem of linear estimation has a particular meaning to be described below. Secondly, a precise lower bound is given for the risk function of a linear estimate, and finally the class of problems admitting consistent linear estimates is characterized.

By a linear estimate of g is meant a random variable of the form

$$\sum_{j=1}^{n} c_j Y(t_j),\ c_1, \cdots, c_n \text{ real},\ t_1, \cdots, t_n \varepsilon T,$$

or a limit in the mean of such random variables. The loss function being squared error, the risk function for an estimator Z is given by $E_\beta[Z - g(\beta)]^2$.

Since, in general, limits in the mean must be considered, a structural requirement will be imposed on the problem to insure that such estimators have meaning. Specifically, we make

DEFINITION 3.1. The span of $\{Y(t), t \varepsilon T\}$ in $L_2(dP_\beta)$, $V_\beta[Y(t), t \varepsilon T]$, is said to be operationally independent of β, if and only if a sequence of random variables of the form $\{\sum_{j=1}^{N(n)} c_{jn} Y(t_{jn})\}$ is a Cauchy sequence in $L_2(dP_\beta)$ for all β or no β.

The requirement of operational independence imposes a condition on the functions $m(\cdot, \beta)$ and the kernel K and some intermediate results are necessary in order to obtain the condition explicitly.

First of all, $E_\beta Y(s)Y(t) = m(s,\beta)m(t,\beta) + K(s,t) = K_\beta(s,t)$, say. According to Theorem 2.1, $V_\beta[Y(t), t \varepsilon T]$ is isomorphic to the reproducing kernel space $H(K_\beta, T)$ under a map $\Psi_\beta \ni$

$$\{\Psi_\beta Z\}(t) = E_\beta Z Y(t).$$

Thus $V_\beta[Y(t), t \varepsilon T]$ is operationally independent of β, if and only if a sequence of functions of the form $\{\sum_{j=1}^{N(n)} c_{jn} K_\beta(\cdot, t_{jn})\}$ is a Cauchy sequence in $H(K_\beta, T)$ for all β or no β. By Theorem 2.4, this last condition is equivalent to $H(K_\beta, T)$ independent of β (that is, the class of functions $H(K_\beta, T)$ does not depend on β).

Let now $M_\beta(s,t) = m(s,\beta)m(t,\beta)$, $s,t \varepsilon T$, $\beta \varepsilon \Lambda$. M_β as defined is an element of $\mathcal{K}(T)$ and if $m(\cdot, \beta) \not\equiv 0$,

LEMMA 3.1. *$H(M_\beta, T)$ consists of multiples of the function $m(\cdot, \beta)$ with $\| m(\cdot, \beta)\|_{M_\beta}^2 = 1$.*

With the aid of Lemma 3.1, the condition that $H(K_\beta, T)$ be independent of β, and hence that $V_\beta[Y(t), t \varepsilon T]$ is operationally independent of β, may be equated to a condition on the functions $m(\cdot, \beta)$ and the kernel K.

THEOREM 3.1. *$H(K_\beta, T)$ is independent of β, if and only if, for every ordered pair β, λ of Λ $\exists\, d_{\beta\lambda} \ni m(\cdot, \beta) - d_{\beta\lambda} m(\cdot, \lambda) \varepsilon H(K,T)$.*

PROOF. Suppose $H(K_\beta, T)$ is independent of β. By Theorem 2.2 ($K_\beta = M_\beta + K$), the elements of $H(K_\beta, T)$ are of the form

$$f(\cdot) = dm(\cdot, \beta) + h(\cdot),\ h \varepsilon H(K,T).$$

By assumption $m(\cdot, \lambda) \varepsilon H(K_\beta, T)$ for every $\lambda \varepsilon \Lambda$ and hence $m(\cdot, \lambda) = dm(\cdot, \beta) + h(\cdot)$ and the conclusion follows. Suppose then, to every ordered pair β,λ of Λ there corresponds $d_{\beta\lambda} \ni m(\cdot, \beta) - d_{\beta\lambda} m(\cdot, \lambda) \varepsilon H(K,T)$. If $f \varepsilon H(K_\beta, T)$, then

$$f(\cdot) = dm(\cdot, \beta) + h(\cdot),\ h \varepsilon H(K,T).$$

If $d = 0, f \varepsilon H(K,T) \subset H(K_\lambda, T)$ for any $\lambda \varepsilon \Lambda$. If $d \neq 0$, then with $m(\cdot, \beta) - d_{\beta\lambda} m(\cdot, \lambda) = h_{\beta\lambda}(\cdot) \varepsilon H(K,T)$,

$$f(\cdot) = dd_{\beta\lambda} m(\cdot, \lambda) + [dh_{\beta\lambda}(\cdot) + h(\cdot)]$$

and $f \varepsilon H(K_\lambda, T)$ for any $\lambda \varepsilon \Lambda$.

It may be seen from Theorem 3.1 that operational independence arises in two distinct ways. The case of particular interest is the case $m(\cdot, \beta) \varepsilon H(K,T)$ for all β, here any choice of constants $d_{\beta\lambda}$ will suffice and in fact $H(K_\beta, T) = H(K,T)$ for all $\beta \varepsilon \Lambda$. The second case necessarily involves the condition $m(\cdot, \beta) \not\varepsilon H(K,T)$ for all $\beta \varepsilon \Lambda$. An example of the latter case is the one parameter regression model;

$$m(\cdot, \beta) = \beta\varphi(\cdot), \qquad \beta \varepsilon \Lambda = R^1 - \{0\}, \qquad \varphi \not\varepsilon H(K,T),$$

here $d_{\beta\lambda} = \beta/\lambda$. In the second case of operational independence, the constants $d_{\beta\lambda}$ are unique.

The condition arrived at above for operational independence, is of the same nature as that of Theorem 9A of Parzen [4]. There it is shown that under the model of the present paper with $\{Y(t), t \varepsilon T\}$ a Gaussian process, the measures $\{P_\beta, \beta \varepsilon \Lambda\}$ induced by the mean functions $m(\cdot, \beta)$ are equivalent if and only if $m(\cdot, \beta) - m(\cdot, \lambda) \varepsilon H(K,T)$ for all β, (under some minor conditions on T and K). Thus equivalence of the measures would imply operational independence but not conversely, the example mentioned above being an appropriate counterexample.

Returning to consideration of linear estimators, we give now a precise (attainable) lower bound on the risk function of a linear estimate. The bound is given for fixed $\beta \varepsilon \Lambda$, in the case of operational independence it may be regarded as a functional lower bound.

THEOREM 3.2. *If $Z \varepsilon V_\beta[T(t), t \varepsilon T]$, then*

$$E_\beta[Z - g(\beta)]^2 \geqq E_\beta[Z_\beta - g(\beta)]^2 = \begin{cases} \dfrac{g^2(\beta)}{1 + \|m(\cdot, \beta)\|_K^2} & \text{if } m(\cdot, \beta) \varepsilon H(K, T) \\ 0 & \text{if } m(\cdot, \beta) \not\varepsilon H(K, T). \end{cases}$$

PROOF. Regarding $g(\beta)$ as a random variable in $L_2(dP_\beta)$,

$$E_\beta[Z - g(\beta)]^2 \geqq E_\beta[Z_\beta - g(\beta)]^2,$$

where Z_β denotes the projection of $g(\beta)$ onto $V_\beta[Y(t), t \epsilon T]$. Now Z_β satisfies

$$E_\beta Z_\beta Y(t) = E_\beta g(\beta) Y(t) = g(\beta) m(t,\beta) \quad \text{for} \quad t \varepsilon T,$$

and thus by Theorem 2.1,

$$E_\beta Z_\beta^2 = \| g(\beta) m(\cdot, \beta) \|_{K_\beta}^2 = g^2(\beta) \| m(\cdot, \beta) \|_{K_\beta}^2.$$

Theorem 2.2 and Lemma 3.1 enable us to compute $\| m(\cdot, \beta) \|_{K_\beta}^2$ in terms of $\| m(\cdot, \beta) \|_K^2$ when $m(\cdot, \beta) \varepsilon H(K,T)$. Indeed

$$\begin{aligned}\| m(\cdot, \beta)\|_{K_\beta}^2 &= \min_{f_1+f_2=m(\cdot,\beta)}[\| f_1 \|_{M_\beta}^2 + \| f_2 \|_K^2] \\ &= \min_u[u^2 \| m(\cdot, \beta)\|_{M_\beta}^2 + (1-u)^2 \| m(\cdot, \beta)\|_K^2] \\ &= \min_u [u^2 + (1-u)^2 \| m(\cdot, \beta) \|_K^2] = \frac{\| m(\cdot,\beta) \|_K^2}{(1 + \|m(\cdot,\beta)\|_K^2)}.\end{aligned}$$

Thus $E_\beta Z_\beta^2 = g^2(\beta) \dfrac{\| m(\cdot,\beta) \|_K^2}{(1+\|m(\cdot,\beta)\|_K^2)}$ and $E_\beta[Z_\beta - g(\beta)]^2 = \dfrac{g^2(\beta)}{(1+\|m(\cdot,\beta)\|_K^2)}$. As is easily checked, if $m(\cdot,\beta) \not\in H(K, T)$, $E_\beta Z_\beta^2 = g^2(\beta)$ and $E_\beta[Z_\beta - g^2(\beta)]^2 = 0$.

Theorem 3.2 gives then a lower bound for the risk function of a linear estimate, the bound being attained for the appropriate projection random variable.

We assume now the case of operational independence and characterize those problems which allow linear estimates with zero risk for all $\beta \varepsilon \Lambda$. To avoid annoying though trivial difficulties, it is assumed that the function g in question has no zeros.

THEOREM 3.3. *Suppose $V_\beta[Y(t), t \varepsilon T]$ is operationally independent of β. There exists a linear estimator with zero risk for $g(\cdot)$, if and only if $m(\cdot, \beta) \varepsilon H(K,T)$, for no $\beta \varepsilon \Lambda$, and $m(\cdot, \beta) - [g(\beta)/g(\lambda)]\, m(\cdot, \lambda) \varepsilon H(K,T)$, for all $\beta, \lambda \varepsilon \Lambda$.*

PROOF. Suppose $Z = \text{l.i.m.} \sum_{j=1}^{N(n)} c_{jn} Y(t_{jn})$ is a linear estimator with zero risk, i.e.,

$$\begin{aligned}E_\lambda[Z - g(\lambda)]^2 &= \lim_{n\to\infty} \Big[\sum_{j,k=1}^{N(n)} c_{jn}\, c_{kn}\, K_\beta(t_{jn}, t_{kn}) \\ &\qquad - 2g(\lambda) \sum_{j=1}^{N(n)} c_{jn}\, m(t_{jn}, \lambda) + g^2(\lambda) \Big] \\ &= \lim_{n\to\infty} \Big[\sum_{j,k=1}^{N(n)} c_{jn}\, c_{kn}\, K(t_{jn}, t_{kn}) + \Big(\sum_{j=1}^{N(n)} c_{jn}\, m(t_{jn}, \lambda) - g(\lambda) \Big)^2 \Big] \equiv 0.\end{aligned}$$

Consequently $\{\sum_{j=1}^{N(n)} c_{jn} K(\cdot, t_{jn})\}$, which is a Cauchy sequence in $H(K,T)$ by Theorem 2.4, tends to the zero function and $\lim_{n\to\infty} \sum_{j=1}^{N(n)} c_{jn} m(t_{jn}, \lambda) = g(\lambda)$ for all $\lambda \varepsilon \Lambda$. By operational independence $\exists$ constants $d_{\beta\lambda} \ni m(\cdot, \beta) - d_{\beta\lambda}\, m(\cdot,\lambda) = h_{\beta\lambda}(\cdot) \varepsilon H(K,T)$ and therefore

$$\begin{aligned}g(\beta) - d_{\beta\lambda}\, g(\lambda) &= \lim_{n\to\infty} \Big[\sum_{j=1}^{N(n)} c_{jn}\, m(t_{jn}, \beta) - d_{\beta\lambda} \sum_{j=1}^{N(n)} c_{jn}\, m(t_{jn}, \lambda) \Big] \\ &= \lim_{n\to\infty} \sum_{j=1}^{N(n)} c_{jn}\, h_{\beta\lambda}(t_{jn}) = \lim_{n\to\infty} (\textstyle\sum c_{jn}\, K(\cdot, t_{jn}), h_{\beta\lambda})_K = 0,\end{aligned}$$

or $d_{\beta\lambda} = g(\beta)/g(\lambda)$. The reverse implication follows immediately from Theorem 3.2 and the computation of the risk function above.

Theorem 3.3 contains the regression model $m(\cdot, \beta) = \beta\varphi(\cdot)$ with $\beta \varepsilon R^1 - \{0\}$ and $\varphi \not\in H(K,T)$ for the function $g(\beta) = \beta$. It is directly verified that adding the parameter point $\beta = 0$ will destroy the operational independence but will still allow a linear estimator with zero risk over R^1. Consequently Theorem 3.3 is a generalization of the corresponding results in [2], [3], and [4], see also [7].

4. Examples and remarks. Numerous examples of specific reproducing kernel spaces are known, that is, the norm structure is known explicitly (cf. [5]). Thus the notion of operational independence and the computation of the lower bound of Theorem 3.2 are not intractable. We turn to some examples.

Suppose T is a finite set, say $T = \{t_1, \cdots, t_n\}$ and let K denote here the matrix

$$\begin{pmatrix} K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \vdots & & \vdots \\ K(t_1, t_n) & \cdots & K(t_n, t_n) \end{pmatrix}$$

When T is finite, the concept of operational independence is not appropriate as any linear estimate is obtainable as a finite linear combination of the observations. The reproducing kernel space here consists of all functions on T in the row space of the matrix K, consisting of all functions when K is nonsingular and the dimension of $H(K,T)$ corresponding always to the rank of K. When K is nonsingular the bound of Theorem 3.2 is $g^2(\beta)/[1 + m(\cdot, \beta)'K^{-1}m(\cdot, \beta)]$ with $m(\cdot, \beta)' = (m(t_1, \beta), \cdots, m(t_n, \beta))$. And when K is singular the bound is (when non zero) $g^2(\beta)/[1 + c(\beta)'Kc(\beta)]$ where $c(\beta)$ is the vector solution of $Kc(\beta) = m(\cdot, \beta)$ (cf. [5]).

We next suppose T to be an infinite set.

EXAMPLE 1. T and K arbitrary, $m(\cdot, \beta) = \beta\varphi(\cdot), \beta \varepsilon R^1$. The bound of Theorem 3.2 is trivial if $\varphi \not\varepsilon H(K,T)$, this corresponding to the situation in Theorem 3.3. When $\varphi \varepsilon H(K,T)$ and $g(\beta) = \beta$, the bound is $\beta^2/(1 + \beta^2 \| \varphi \|_K^2)$.

EXAMPLE 2. $T = \Lambda = R^1$, $K(s,t) = k(s - t)$, and $m(t,\beta) = \varphi(t - \beta) = \varphi^\beta(t)$. Here $\varphi^\beta \varepsilon H(K,T)$ for all β or no β and in the former case $\| \varphi^\beta \|_K^2$ is independent of β (cf. [6]). In this instance the bound for $g(\beta) = \beta$ is $\beta^2/(1 + \| \varphi \|_K^2)$.

EXAMPLE 3. $T = [0,A]$, $\Lambda = [0, \infty)$, $K(s,t) = e^{-|s-t|}$ and $m(t, \beta) = \cos \beta t$. For estimating $g(\beta) = \beta$, we find [5] that $m(\cdot, \beta) \varepsilon H(K,T)$ for all β and

$$\|m(\cdot, \beta)\|_K^2 = \frac{1}{2}\int_0^A (\cos \beta t - \beta \sin \beta t)^2 \, dt + 1$$

$$= \frac{1}{4}(3 + A + \cos 2\beta A) - \frac{1}{8}\left(\frac{1 + \beta^2}{\beta}\right) \sin 2\beta A + \frac{\beta^2 A}{4}.$$

And the bound $\beta^2/[1 + \| m(\cdot, \beta)\|_K^2]$ is bounded in β.

We turn then to remarks on the feasibility of linear estimates. The effect of operational independence is indeed to insure proper structure for linear estimation when T is infinite. The bound obtained has some application although more largely in a negative direction. Thus in Example 2, even when $T = R^1$, any linear estimator has an unbounded risk function and the situation is worse for a smaller set of observations.

When linear unbiased estimators exist with finite risk, the bound of Theorem 3.2 will reflect this. Parzen has shown that these estimators exist if and only if $g \varepsilon H(M,\Lambda)$ where $M(\alpha,\beta) = (m(\cdot,\alpha), m(\cdot,\beta))_K$ (assuming $m(\cdot, \beta) \varepsilon H(K,T)$

for all β). If g has this property, then by the Cauchy-Schwartz inequality,

$$g^2(\beta) = (g, M(\cdot, \beta))_M^2 \leqq \| g \|_M^2 M(\beta,\beta) = \| g \|_M^2 \| m(\cdot, \beta)\|_K^2,$$

and for the bound

$$g^2(\beta)/[1 + \| m(\cdot, \beta)_K^2] \leqq \{\| g \|^2 \| m(\cdot, \beta)\|^2/[1 + \| m(\cdot, \beta)\|_K^2]\} \leqq \| g \|_M^2,$$

which last quantity is precisely the variance of the best linear unbiased estimator. In any event $\sup_\beta\{g^2(\beta)/[1 + \| m(\cdot, \beta)\|_K^2]\}$ is a lower bound for the variance of a linear unbiased estimator, if one exists.

In general, when $m(\cdot, \beta) \varepsilon H(K,T)$ for all β, the risk functions for admissible linear estimators (among the class of linear estimators) will be given by

$$((f, m(\cdot, \beta))_K - g(\beta))^2 + \| f \|_K^2,$$

where $f \varepsilon V[m(\cdot, \beta), \beta \varepsilon \Lambda]$ in $H(K,T)$. This may be seen in the calculations of Theorem 3.3.

REFERENCES

[1] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

[2] BALAKRISHNAN, A. V. (1959). On a characterization of covariances. *Ann. Math. Statist.* **30** 670–675.

[3] GRENANDER, U. and SZEGO, G. (1958). *Toeplitz Forms and Their Application.* Univ. of California Press, Berkeley.

[4] PARZEN, E. (1959). Statistical inference on time series by Hilbert space methods, I. Tech. Rpt. No. 23 (NR-042-993). Appl. Math. and Statist. Lab., Stanford Univ.

[5] PARZEN, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

[6] YLVISAKER, N. D. (1959). On time series analysis and reproducing kernel spaces. Tech. Rpt. No. 1 (DA-04-200-ORD-996). Appl. Math. and Statist. Lab., Stanford Univ.

[7] YLVISAKER, N. D. (1961). A generalization of a theorem of Balakrishnan. *Ann. Math. Statist.* **32** 1337–1339.