# REGRESSION ANALYSIS OF CONTINUOUS PARAMETER TIME SERIES

EMANUEL PARZEN
STANFORD UNIVERSITY

## 1. Summary

This paper attempts to show how the problem of regression analysis of time series can be treated, using Hilbert space techniques, in a manner which applies simultaneously to discrete and continuous parameter time series, and also to multiple time series. The idea of a Hilbert space representation of a time series and in particular the reproducing kernel Hilbert space representation, is discussed in section 3. Examples of such representations are given in section 4. A formula for the probability density functional of a normal time series is obtained in section 5. The problem of maximum likelihood estimation of the mean value function of a normal time series is treated in section 6. Minimum variance unbiased linear estimation of the mean value function is treated in section 7. Tests of hypotheses and simultaneous confidence bands for mean value functions are given in section 8. A method of iteratively evaluating reproducing kernel inner products is described in section 9.

## 2. Introduction

The problem of regression analysis of time series may be formulated as follows. A model often adopted for the analysis of an observed time series $X(t)$, $t \in T$, is to regard $X(t)$ as the sum of two functions

$$X(t) = m(t) + Y(t), \qquad t \in T. \tag{2.1}$$

We call $m(t)$ the mean value function and $Y(t)$ the fluctuation function.

The stochastic process $Y(t)$ is assumed to possess finite second moments, and to have zero means and covariance kernel

$$K(s, t) = E[Y(s)Y(t)]. \tag{2.2}$$

The mean value function is assumed to belong to a known class $M$ of functions. Very often $M$ is taken to be the set of all linear combinations of $q$ known functions $w_1(t), \cdots, w_q(t)$. Then, for $t$ in $T$,

**This paper was prepared under Contract Nonr 225(21) for the Office of Naval Research.**

Reprinted by permission from the *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, University of California Press, Berkeley, Calif., 1961, pp. 469–489

$$m(t) = \beta_1 w_1(t) + \cdots + \beta_q w_q(t), \tag{2.3}$$

for some coefficients $\beta_1, \cdots, \beta_q$ to be estimated.

In regard to the index set $T$, there are cases of particular importance. One may be observing (i) a discrete parameter time series $X(t)$, in which case one assumes $T$ is a finite set of points written $T = \{1, 2, \cdots, N\}$; (ii) a continuous parameter time series, in which case $T$ is a finite interval written $T = \{t: 0 \leq t \leq L\}$; or (iii) a multiple (discrete or continuous parameter) time series $\{[X_1(t), \cdots, X_k(t)], t \in T'\}$ which may be written as a time series $\{X(t), t \in T\}$ whose index set $T = \{(j, t): j = 1, \cdots, k \text{ and } t \in T'\}$.

Various methods of forming estimates of $m(t)$ are available. The most important methods are classical least squares estimation and minimum variance linear unbiased estimation. In the case of normally distributed observations, one has in addition the methods of maximum likelihood estimation and minimum variance unbiased estimation. In this paper we show how Hilbert space techniques may be used to form explicit expressions for these estimates in terms of certain so-called reproducing kernel inner products.

It is generally accepted that maximum likelihood estimates under the assumption of normality are equivalent to minimum variance unbiased linear estimates. Many researchers in time series analysis (notably Whittle [16], [17]) have very effectively utilized this principle to develop a "least squares" theory of time series analysis (for a lucid statement of this philosophy, see [17], p. 132). The methods discussed here permit a development of the least squares theory from first principles.

## 3. Hilbert space representations of time series

In the 1940's probabilists began to use Hilbert space methods to study time series. In this section we define the notion of a Hilbert space representation of a time series, and show the fundamental role played in the representation theory of time series by *reproducing kernel Hilbert spaces.*

*Definition* 3A. Given a time series $\{X(t), t \in T\}$ with finite second moments we call

$$K(s, t) = E[X(s)X(t)] \tag{3.1}$$

its covariance kernel, and

$$K(s, t) = \operatorname{Cov}[X(s), X(t)] \tag{3.2}$$

its proper covariance kernel.

*Definition* 3B. By an abstract Hilbert space is meant a set $H$ whose members $u, v, \cdots$ are usually called vectors or points which possesses the following properties

(I) $H$ is a linear space. That is, for any vectors $u$ and $v$ in $H$, and real number $a$, there exist vectors, denoted by $u + v$ and $au$ respectively, which satisfy the usual algebraic properties of addition and multiplication; also there exists a zero vector 0 with the usual properties under addition.

(II) $H$ is an inner product space. That is, to every pair of points $u$ and $v$ in $H$ there corresponds a real number, written $(u, v)$ and called the inner product of $u$ and $v$, possessing the following properties: for all points $u$, $v$, and $w$ in $H$, and every real number $a$,

(i) $(au, v) = a(u, v)$
(ii) $(u + v, w) = (u, w) + (v, w)$
(iii) $(v, u) = (u, v)$
(iv) $(u, u) > 0$ if and only if $u \neq 0$.

(III) $H$ is a complete metric space under the norm $||u|| = (u, u)^{1/2}$. That is, if $\{u_n\}$ is a sequence of points such that $||u_m - u_n|| \to 0$ as $m, n \to \infty$ then there is a vector $u$ in $H$ such that $||u_n - u||^2 \to 0$ as $n \to \infty$.

*Definition* 3C. We call $(Q, \mathbf{B}, \mu)$ a measure space if $Q$ is a set, $\mathbf{B}$ is a $\sigma$-field of subsets of $Q$, and $\mu$ is a measure on the measurable space $(Q, \mathbf{B})$. We denote by $L_2(Q, \mathbf{B}, \mu)$ the Hilbert space of all $\mathbf{B}$-measurable real-valued functions defined on $Q$ satisfying

$$(f, f)_{\hat{\mu}} = \int_Q f^2 \, d\mu < \infty. \tag{3.3}$$

*Definition* 3D. Let $T$ be an index set, and let $\{u(t), t \in T\}$ be a family of members of a Hilbert space $H$. The linear manifold spanned by the family $\{u(t), t \in T\}$, denoted $L[u(t), t \in T]$, is defined to be the set consisting of all vectors $u$ in $H$ which may be represented in the form $u = \sum_{i=1}^{n} c_i u(t_i)$ for some integer $n$, some constants $c_1, \cdots, c_n$, and some points $t_1, \cdots, t_n$ in $T$. The Hilbert space spanned by the family $\{u(t), t \in T\}$, denoted $V[u(t), t \in T]$ or $L_2[u(t), t \in T]$ if $H$ is the space of square integrable functions on some measure space, is defined to be the set of vectors which either belong to the linear manifold $L[u(t), t \in T]$ or may be represented as a limit of vectors in $L[u(t), t \in T]$. If $V[u(t), t \in T]$ coincides with $H$, we say that $\{u(t), t \in T\}$ *spans* $H$.

LEMMA 3a. *The family $\{u(t), t \in T\}$ spans $H$ if and only if the vector $g = 0$ is the only vector in $H$ satisfying $[g, u(t)] = 0$ for every $t$ in $T$.*

*Definition* 3E. The Hilbert space spanned by a time series $\{X(t), t \in T\}$ is denoted by $L_2[X(t), t \in T]$ and is defined to consist of all random variables $U$ which are either finite linear combinations of the random variables $\{X(t), t \in T\}$ or are limits of such finite linear combinations in the norm corresponding to the inner product defined on the space of square integrable random variables by $(U, V) = E(UV)$. In words, $L_2[X(t), t \in T]$ consists of all linear functionals in the time series.

*Definition* 3F. A Hilbert space $H$ is said to be a reproducing kernel Hilbert space, with reproducing kernel $K$, if the members of $H$ are functions on some set $T$, and if there is a kernel $K$ on $T \otimes T$ having the two properties: for every $t$ in $T$, where $K(\cdot, t)$ is the function defined on $T$, with value at $s$ in $T$ equal to $K(s, t)$,

$$K(\cdot, t) \in H \tag{3.4}$$

$$[g, K(\cdot, t)] = g(t) \tag{3.5}$$

for every $g$ in $H$.

*Definition* 3G. Let $T$ be an index set, and let $K$ be a real-valued function of two variables defined on $T \otimes T$. The function (or kernel) $K$ is called a *nonnegative kernel* if for any integer $n$, and $n$ points $\{t_1, \cdots, t_n\}$ in $T$, and any set of $n$ real numbers $\{a_1, \cdots, a_n\}$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(t_i, t_j) \geqq 0. \tag{3.6}$$

The kernel $K$ is said to be symmetric if, for all $s$ and $t$ in $T$, $K(s, t) = K(t, s)$. It is to be noted that if (3.6) were required to hold for all complex numbers, then (3.6) would imply symmetry.

THEOREM 3A (see Loève [11], p. 466). *$K$ is the covariance kernel of a time series if, and only if, $K$ is a symmetric nonnegative kernel.*

THEOREM 3B (Moore-Aronszajn [1]). *A symmetric nonnegative kernel $K$ generates a unique Hilbert space, which we denote by $H(K)$, of which $K$ is the reproducing kernel.*

LEMMA 3b. *If $K$ is a reproducing kernel for the Hilbert space $H$, then the family of functions $\{K(\cdot, t), t \in T\}$ spans $H$.*

THEOREM 3C. *Let $K$ be a covariance kernel. If there exist a measure space $(Q, \mathbf{B}, \mu)$, and a family of functions $\{f(t), t \in T\}$ in $L_2(Q, \mathbf{B}, \mu)$ such that for all $s$, $t$ in $T$*

$$K(s, t) = \int_Q f(s) f(t)\, d\mu, \tag{3.7}$$

*then the reproducing kernel Hilbert space $H(K)$ corresponding to the covariance kernel $K$ may be described as follows: $H(K)$ consists of all functions $g$, defined on $T$, which may be represented in the form*

$$g(t) = \int_Q g^* f(t)\, d\mu \tag{3.8}$$

*for some (necessarily unique) function $g^*$ in the Hilbert subspace $L_2[f(t), t \in T]$ of $L_2(Q, \mathbf{B}, \mu)$ spanned by the family of functions $\{f(t), t \in T\}$, with norm given by*

$$||g||^2 = \int_Q |g^*|^2\, d\mu. \tag{3.9}$$

PROOF. Verify that the set $H$ of functions of the form of (3.8), with norm given by (3.9), is a Hilbert space satisfying (3.4) and (3.5).

The definition we give of a representation of a time series is based on the following theorem.

BASIC CONGRUENCE THEOREM. *Let $H_1$ and $H_2$ be two abstract Hilbert spaces. Denote the inner product between two vectors $u_1$ and $u_2$ in $H_1$ by $(u_1, u_2)_1$. Similarly, denote the inner product between two vectors $v_1$ and $v_2$ in $H_2$ by $(v_1, v_2)_2$. Let $T$ be an* *index set. Let* $\{u(t), t \in T\}$ *be a family of vectors which span* $H_1$. *Similarly, let* $\{v(t), t \in T\}$ *be a family of vectors which span* $H_2$. *Suppose that, for every s and t in T,*

$$[u(s), u(t)]_1 = [v(s), v(t)]_2. \tag{3.10}$$

*Then there exists a congruence (a one-one inner product preserving linear mapping)* $\psi$ *from* $H_1$ *onto* $H_2$ *which has the property that*

$$\psi[u(t)] = v(t)_1, \; t \text{ in } T. \tag{3.11}$$

PROOF. Define the function $\psi$ from $H_1$ to $H_2$ as follows. For each vector in the family $\{u(t), t \in T\}$, define $\psi[u(t)] = v(t)$. For each vector $u$ in the linear manifold $L[u(t), t \in T]$, define

$$\psi(u) = \sum c_i v(t_i) \quad \text{if} \quad u = \sum c_i u(t_i). \tag{3.12}$$

We need to prove that the mapping $\psi$ is well defined, that is, it needs to be shown that two different representations of a vector, $u = \sum c_i u(t_i) = \sum c_i' u(t_i')$, lead to the same value $\psi(u) = \sum c_i v(t_i) = \sum c_i' v(t_i')$. To prove this it suffices to prove that

$$\sum c_i u(t_i) = 0 \quad \text{if and only if} \quad \sum c_i v(t_i) = 0, \tag{3.13}$$

which follows from the fact that

$$\begin{aligned} 0 = \|\sum c_i u(t_i)\|_1^2 &= \sum c_i c_j [u(t_i), u(t_j)]_1 \\ &= \sum c_i c_j [v(t_i), v(t_j)]_2 = \|\sum c_i v(t_i)\|_2^2. \end{aligned} \tag{3.14}$$

From the last equation we see that $\psi$ is a congruence from $L[u(t), t \in T]$ onto $L[v(t), t \in T]$. Consequently, it follows for any sequence $\{u_n\}$ in $L[u(t), t \in T]$ that $\{u_n\}$ is a Cauchy sequence in $H_1$ if and only if $\{\psi(u_n)\}$ is a Cauchy sequence in $H_2$, and $\lim_n u_n = 0$ if and only if $\lim_n \psi(u_n) = 0$. Therefore, for $u = \lim_n u_n$, define $\psi(u) = \lim_n \psi(u)$. In this way $\psi$ is defined for every $u$ in $H_1$. One may verify that $\psi$ is a congruence from $H_1$ onto $H_2$.

*Definition* 3H. A family of vectors $\{f(t), t \in T\}$ in a Hilbert space $H$ is said to be a representation of a time series $\{X(t), t \in T\}$ if, for every $s$ and $t$ in $T$,

$$[f(s), f(t)]_H = K(s, t) = E[X(s)X(t)]. \tag{3.15}$$

Then there is a congruence $\psi$ from $V[f(t), t \in T]$ onto $L_2[X(t), t \in T]$ satisfying

$$\psi[f(t)] = X(t) \tag{3.16}$$

and every random variable $U$ in $L_2[X(t), t \in T]$ may be written $U = \psi(g)$ for some unique vector $g$ in $V[f(t), t \in T]$.

Since $K(s, t) = [K(\cdot, s), K(\cdot, t)]_{H(K)} = E[X(s)X(t)]$ we immediately obtain the following important theorem.

THEOREM 3D. *Let* $\{X(t), t \in T\}$ *be a time series with covariance kernel K. Then the family* $\{K(\cdot, t), t \in T\}$ *of functions in* $H(K)$ *is a representation for* $\{X(t), t \in T\}$. *Given a function g in* $H(K)$, *we denote by* $(X, g)_K$ *or* $(g, X)_K$ *the random variable U in* $L_2[X(t), t \in T]$ *which corresponds to g under the congruence* *which maps $K(\cdot, t)$ into $X(t)$. We then have the following formal relations: for every $t$ in $T$, and $g, h$ in $H(K)$,*

$$[X, K(\cdot, t)]_K = X(t) \tag{3.17}$$

$$E[(X, h)_K(X, g)_K] = (h, g)_K, \tag{3.18}$$

*where we write $(h, g)_K$ for $(h, g)_{H(K)}$.*

*Definition* 3I. Let $(Q, \mathbf{B}, \mu)$ be a measure space and, for every $B$ in $\mathbf{B}$, let $Z(B)$ be a random variable. The family of random variables $\{Z(B), B \in \mathbf{B}\}$ is called an *orthogonal random set function with covariance kernel* $\mu$ if, for any two sets $B_1$ and $B_2$ in $\mathbf{B}$,

$$E[Z(B_1)Z(B_2)] = \mu(B_1B_2), \tag{3.19}$$

where, as usual, $B_1B_2$ denotes the intersection of $B_1$ and $B_2$.

The Hilbert space $L_2[Z(B), B \in \mathbf{B}]$ of random variables spanned by an orthogonal random set function may be defined, as was the Hilbert space spanned by a time series, to be the smallest Hilbert subspace of the Hilbert space of all square integrable random variables containing all random variables $U$ of the form $U = \sum_{i=1}^{n} c_i Z(B_i)$ for some integer $n$, subfamily $\{B_1, \cdots, B_n\} \subset \mathbf{B}$, and real constants $c_1, \cdots, c_n$. On the other hand, $L_2(Q, \mathbf{B}, \mu)$ may be described as the Hilbert space spanned under the norm (3.3) by the family of indicator functions $\{I_B, B \in \mathbf{B}\}$, where the indicator function $I_B$ of $B$ is defined by $I_B(q) = 1$ or $0$ according as $q \in B$ or $q \notin B$. Now for any $B_1, B_2$ in $\mathbf{B}$

$$(I_{B_1}, I_{B_2})_\mu = \mu(B_1B_2) = E[Z(B_1)Z(B_2)]. \tag{3.20}$$

Therefore, by the Basic Congruence Theorem, there is a congruence $\psi$ from $L_2(Q, \mathbf{B}, \mu)$ onto $L_2[Z(B), B \in \mathbf{B}]$ such that for any $B \in \mathbf{B}$,

$$\psi(I_B) = Z(B). \tag{3.21}$$

This fact justifies the following definition of the stochastic integral.

*Definition* 3J. Let $(Q, \mathbf{B}, \mu)$ be a measure space and let $\{Z(B), B \in \mathbf{B}\}$ be an orthogonal random set function with covariance kernel $\mu$. For any function $f$ in $L_2(Q, \mathbf{B}, \mu)$ one defines the stochastic integral of $f$ with respect to $\{Z(B), B \in \mathbf{B}\}$, denoted $\int_Q f\, dZ$, by

$$\int_Q f\, dZ = \psi(f), \tag{3.22}$$

where $\psi$ is the congruence from $L_2(Q, \mathbf{B}, \mu)$ onto $L_2(Z(B), B \in \mathbf{B})$ determined by (3.21).

THEOREM 3E. *Let $\{X(t), t \in T\}$ be a time series with covariance kernel $K$. Let $\{f(t), t \in T\}$ be a family of functions in a space $L_2(Q, \mathbf{B}, \mu)$, such that* (3.7) *holds. Then $\{f(t), t \in T\}$ is a representation for $\{X(t), t \in T\}$.*

*If further, $\{f(t), t \in T\}$ spans $L_2(Q, \mathbf{B}, \mu)$, then there is an orthogonal random set function $\{Z(B), B \in \mathbf{B}\}$ with covariance kernel $\mu$ such that*

$$X(t) = \int_Q f(t)\, dZ, \qquad t \in T, \tag{3.23}$$

*and every random variable U in* $L_2[X(t), t \in T]$ *may be represented*

$$U = \int_Q g\, dZ \tag{3.24}$$

*for some unique function g in* $L_2(Q, \mathbf{B}, \mu)$.

PROOF. Let $\psi$ be the congruence from $L_2[f(t), t \in T]$ *onto* $L_2[X(t), t \in T]$ satisfying (3.16). If $\{f(t), t \in T\}$ spans $L_2(Q, \mathbf{B}, \mu)$, define, for $B \in \mathbf{B}$, $Z(B) = \psi(I_B)$. It is immediate that $\{Z(B), B \in \mathbf{B}\}$ is an orthogonal random set function with covariance kernel $\mu$. By the definition of the stochastic integral, (3.23) is merely another way of writing the fact that $X(t) = \psi[f(t)]$.

In our opinion from the foregoing theorems one may draw the following moral. The notion of the representation of a time series as an integral with respect to an orthogonal random set function is a special case of the notion of representation given by definition 3H. One may choose representations of a time series in a multitude of ways. Indeed, if $(Q, \mathbf{B}, \mu)$ is a measure space such that $L_2[X(t), t \in T]$ and $L_2(Q, \mathbf{B}, \mu)$ have the same dimension, there are many families $\{f(t), t \in T\}$ of functions in $L_2(Q, \mathbf{B}, \mu)$ which are a representation for $\{X(t), t \in T\}$. What one desires is a family $\{f(t), t \in T\}$ of familiar functions, such as the family of complex exponentials $\{e^{i\lambda t}, -\infty < t < \infty\}$, which are a representation in a suitable space $L_2(Q, \mathbf{B}, \mu)$ for a stationary time series.

The representation of a time series with covariance kernel $K$ by the functions $\{K(\cdot, t), t \in T\}$ in the reproducing kernel Hilbert space $H(K)$ is in terms of a well-behaved family of functions. Further, it is in a sense a natural representation. In the sequel, we shall show that it is also the natural representation in terms of which to solve problems of statistical inference on time series.

## 4. Examples of reproducing kernel Hilbert space representations

To illustrate the notion of the representation of a time series by a reproducing kernel Hilbert space, let us here consider the case of a time series $\{X(t), a \leqq t \leqq b\}$ defined on a finite interval $T = \{a \leqq t \leqq b\}$ with continuous covariance kernel $K(s, t) = E[X(s)X(t)]$.

A representation for $K$ of the form of (3.7) is provided by Mercer's theorem, which may be stated as follows. *If one defines* $\{\varphi_n(t), n = 1, 2, \cdots\}$ *to be the sequence of normalized eigenfunctions and* $\{\lambda_n, n = 1, 2, \cdots\}$ *to be the sequence of corresponding nonnegative eigenvalues satisfying the relation*

$$\int_a^b K(s, t)\varphi_n(s)\, ds = \lambda_n \varphi_n(t), \qquad a \leqq t \leqq b, \tag{4.1}$$

$$\int_a^b \varphi_m(t)\varphi_n(t)\, dt = \delta(m, n), \tag{4.2}$$

*where* $\delta(m, n)$ *is the Kronecker delta function equal to* 1 *or* 0 *depending on whether* $m = n$ *or* $m \neq n$, *then the kernel* $K(s, t)$ *may be written*

$$K(s, t) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(s) \varphi_n(t) \tag{4.3}$$

*where the series converges absolutely and uniformly for* $a \leqq s, t \leqq b$.

For ease of exposition we assume that the eigenfunctions span the space of square integrable functions on the interval $a \leqq t \leqq b$.

It may be shown that the reproducing kernel Hilbert space $H(K)$ corresponding to the covariance kernel $K$ consists of all square integrable functions $h(t)$ on the interval $a \leqq t \leqq b$ such that

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} \left| \int_a^b h(t) \varphi_n(t)\, dt \right|^2 < \infty. \tag{4.4}$$

The reproducing kernel inner product between two such functions is given by

$$(h, g)_K = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \int_a^b h(t) \varphi_n(t)\, dt \int_a^b g(t) \varphi_n(t)\, dt. \tag{4.5}$$

The random variable $(h, X)_K$ in $L_2[X(t), a \leqq t \leqq b]$ corresponding to $h(\cdot)$ in $H(K)$ under the mapping described in theorem 3D is given by (4.5) with $g$ replaced by $X$.

We next consider the reproducing kernel Hilbert space corresponding to the covariance kernel of an autoregressive scheme $X(t)$ observed over a finite interval $a \leqq t \leqq b$.

A continuous parameter stationary time series $X(t)$ is said to be an autoregressive scheme of order $m$ if its covariance function $R(u) = E[X(t)X(t + u)]$ may be written (see Doob [3], p. 542)

$$R(s - t) = \int_{-\infty}^{\infty} \frac{e^{i(s-t)\omega}}{2\pi \left| \sum_{k=0}^{m} a_k (i\omega)^{m-k} \right|^2}\, d\omega \tag{4.6}$$

where the polynomial $\sum_{k=0}^{m} a_k z^{m-k}$ has no zeros in the right half of the complex $z$-plane. It may be shown that given observations of such a time series over a finite interval $a \leqq t \leqq b$, the corresponding reproducing kernel Hilbert space contains all functions $h(t)$ on $a \leqq t \leqq b$ which are continuously differentiable of order $m$. The reproducing kernel inner product is given by

$$(h, g)_K = \int_a^b (L_t h)(L_t g)\, dt + \sum_{j,k=0}^{m} d_{j,k} h^{(j-1)}(a) g^{(k-1)}(a) \tag{4.7}$$

where

$$L_t h = \sum_{k=0}^{m} a_k h^{(m-k)}(t) \tag{4.8}$$

$$\{d_{j,k}\}^{-1} = \left\{ \frac{\partial^{j+k-2}}{\partial t^{j-1}\, \partial u^{k-1}} R(t - u) \Big|_{t=a, u=a} \right\}. \tag{4.9}$$

The first and second autoregressive schemes are of particular importance.

A stationary time series $X(t)$ is said to satisfy a first order autoregressive scheme if it is the solution of a first order linear differential equation whose input is white noise $\eta'(t)$, the symbolic derivative of a process $\eta(t)$ with independent stationary increments

$$\frac{dX}{dt} + \beta X = \eta'(t), \qquad \beta > 0. \tag{4.10}$$

It should be remarked that from a mathematical point of view (4.10) should be written

$$dX(t) + \beta X(t)\, dt = d\eta(t). \tag{4.11}$$

Even then, by saying that $X(t)$ satisfies (4.10) or (4.11) we mean that

$$X(t) = \int_{-\infty}^{t} H(t - s)\, d\eta(s) \tag{4.12}$$

where $H(t - s) = \exp[-\beta(t - s)]$ is the one-sided Green's function of the differential operator $L_t f = f'(t) + \beta f(t)$.

The covariance function of the stationary time series $X(t)$ is

$$R(t - u) = \frac{1}{2\beta} e^{-\beta|u-t|}. \tag{4.13}$$

The corresponding reproducing kernel Hilbert space $H(K)$ contains all differentiable functions. The inner product is given by

$$(h, g) = \int_a^b (f' + \beta f)(g' + \beta g)\, dt + 2\beta f(a)g(a). \tag{4.14}$$

More generally, corresponding to the covariance function

$$K(s, t) = Ce^{-\beta|s-t|} \tag{4.15}$$

the reproducing kernel inner product is

$$\begin{aligned} (h, g)_K &= \frac{1}{2\beta C}\left\{\int_a^b (h' + \beta h)(g' + \beta g)\, dt + 2\beta h(a)g(a)\right\} \\ &= \frac{1}{2\beta C}\int_a^b (h'g' + \beta^2 hg)\, dt + \frac{1}{2C}\{h(a)g(a) + h(b)g(b)\}. \end{aligned} \tag{4.16}$$

The random variable $(h, X)_K$ in $L_2[X(t), a \leqq t \leqq b]$ corresponding to $h(\cdot)$ in $H(K)$ may be written

$$\begin{aligned} (h, X)_K = \frac{1}{2\beta C}&\left\{\beta^2 \int_a^b h(t)X(t)\, dt + \int_a^b h'(t)\, dX(t)\right\} \\ &+ \frac{1}{2C}\{h(a)X(a) + h(b)X(b)\}. \end{aligned} \tag{4.17}$$

Note that $X'(t)$ does not exist in any rigorous sense. Consequently we write $dX(t)$ where $X'(t)\, dt$ seems to be called for. It can be shown that (4.17) makes sense. In the case that $h(\cdot)$ is twice differentiable, one may integrate by parts and write

$$\int_a^b h'(t)\,dX(t) = h'(b)X(b) - h'(a)X(a) - \int_a^b X(t)h''(t)\,dt. \tag{4.18}$$

A stationary time series $X(t)$ is said to satisfy a second order autoregressive scheme if it is the solution of a second order linear differential equation whose input is white noise $\eta'(t)$

$$\frac{d^2X}{dt^2} + 2\alpha\frac{dX}{dt} + \gamma^2 X = \eta'(t), \qquad \alpha > 0. \tag{4.19}$$

If $\omega^2 = \gamma^2 - \alpha^2 > 0$, the covariance function of the time series is

$$R(t-u) = \frac{e^{-\alpha|u-t|}}{4\alpha\gamma^2}\left\{\cos\omega(u-t) + \frac{\alpha}{\omega}\sin\omega|u-t|\right\}. \tag{4.20}$$

The corresponding reproducing kernel Hilbert space contains all twice differentiable functions on the interval $a \leqq t \leqq b$ with inner product

$$\begin{aligned}(h, g)_K = \int_a^b (h'' + 2\alpha h' + \gamma^2 h)(g'' + 2\alpha g' + \gamma^2 g)\,dt \\ + 4\alpha\gamma^2 h(a)g(a) + 4\alpha h'(a)g'(a).\end{aligned} \tag{4.21}$$

To write an expression for $(h, X)_K$, one uses the same considerations as in (4.17).

## 5. The probability density functional of a normal time series

Given a normal time series $\{X(t), t \in T\}$ with known covariance function

$$K(s,t) = \text{Cov}\,[X(s), X(t)] \tag{5.1}$$

and mean value function $m(t) = E[X(t)]$, let $P_m$ be the probability measure induced on the space of sample functions of the time series. Next, let $m_1$ and $m_2$ be two functions, and let $P_1$ and $P_2$ be the probability measures induced by normal time series with the same covariance kernel $K$, and mean value functions equal to $m_1$ and $m_2$ respectively. By the Lebesgue decomposition theorem it follows that there is a set $N$ of $P_1$-measure 0 and a nonnegative $P_1$-integrable function, denoted by $dP_2/dP_1$, such that for every measurable set $B$ of sample functions

$$P_2(B) = \int_B \left(\frac{dP_2}{dP_1}\right) dP_1 + P_2(BN). \tag{5.2}$$

If $P_2(N) = 0$, then $P_2$ is absolutely continuous with respect to $P_1$, and $dP_2/dP_1$ is called the probability density function of $P_2$ with respect to $P_1$. Two measures which are absolutely continuous with respect to one another are called *equivalent*. Two measures $P_1$ and $P_2$ are said to be orthogonal if there is a set $N$ such that $P_1(N) = 0$ and $P_2(N) = 1$.

It has been proved, independently by various authors under various hypotheses, (see, [4], [5], [7], [9], [12], [13] and unpublished notes by L. Le Cam and

C. Stein) that two normal probability measures are either equivalent or orthogonal. From the point of view of obtaining an explicit formula for the probability density function, the following formulation of this theorem is useful (see also Striebel [15]).

THEOREM 5A. *Let $P_m$ be the probability measure induced on the space of sample functions of a time series $\{X(t), t \in T\}$ with covariance kernel $K$ and mean value function $m$. Let $P_0$ be the probability measure corresponding to the normal process with covariance kernel $K$ and mean value function $m(t) = 0$. If assumptions* 5A *hold, then $P_m$ and $P_0$ are either equivalent or orthogonal, depending on whether $m$ does or does not belong to the reproducing kernel Hilbert space $H(K)$. If $m \in H(K)$, then the probability density functional of $P_m$ with respect to $P_0$ is given by*

$$f(X, m) = \frac{dP_m}{dP_0} = \exp\left\{(X, m)_K - \left(\frac{1}{2}\right)(m, m)_K\right\}. \tag{5.3}$$

In order to develop a theory which applies simultaneously to discrete and continuous parameter (possibly multiple) time series throughout the paper the following assumptions are made.

*Assumptions* 5A. The index set $T$ of the family of random variables $\{X(t), t \in T\}$ under consideration is of the form for suitable sets $D$ and $S$

$$T = \{(j, t) : j \in D \quad \text{and} \quad t \in s\} = D \otimes S. \tag{5.4}$$

We assume that $D$ is finite. In regard to $S$, there are two possible assumptions, either that $S$ is a countable set or that $S$ is a separable metric space. In the latter case we assume that the covariance kernel

$$K(s, t) = \text{Cov}\,[X(s), X(t)] \tag{5.5}$$

has the following properties for each pair $i, j$ in $D$ and $t$ in $S$: $K[(i, s), (j, t)]$ is continuous as a function of $s$ in $S$, and $K[(i, s), (j, s)]$ is continuous as a function of $s$ in $S$ bounded in some neighborhood of $t$.

For $n = 1, 2, \cdots$ let

$$S_n = \{t_n t_2, \cdots, t_{N(n)}\} \tag{5.6}$$

be a sequence of monotone increasing finite subsets of $S$ such that the union $S_\infty = \bigcup_{n=1}^{\infty} S_n$ is either equal to $S$ or dense in $S$, depending on whether $S$ is countable or a separable metric space.

Next let

$$T_n = D \otimes S_n, \qquad T_\infty = D \otimes S_\infty. \tag{5.7}$$

For each integer $n$, let $\mathbf{B}_n$ be the smallest sigma-field containing, for any $t$ in $T_n$ and real number $a$, the set $[X(t) < a]$. The sequence of sigma-fields $\mathbf{B}_n$ is monotone nondecreasing. Let $\mathbf{B}_\infty$ be the smallest sigma-field containing each $\mathbf{B}_n$. It is assumed that for every $t$ in $T$, $X(t)$ is measurable with respect to $\mathbf{B}_\infty$. In the continuous parameter case, a sufficient condition for this to hold is that the stochastic process $\{X(t), t \in T\}$ is separable in the sense defined by Doob ([3], chapter 2).

PROOF. Let $P_m^{(n)}[P_0^{(n)}]$ be the restriction of $P_m[P_0]$ to $\mathbf{B}_n$. Let $K_n$ be the restriction of the covariance kernel $K$ to $T_n \otimes T_n$. Denote by $(h, g)_n$ the inner product between two functions $h$ and $g$ in the reproducing kernel Hilbert space $H(K_n)$. Theorem 5A is a consequence of the following lemmas, whose proofs we omit in order not to overload the present paper (for details, see Parzen [12]).

LEMMA 5a. *If for some integer $n$, the restriction of $m$ to $T_n$ does not belong to $H(K_n)$ then there is a random variable $U$ in $L_2[X(t), t \in T_n]$ such that $P_0^{(n)}[U = 0] = 1$, $P_m^{(n)}[U = 0] = 0$. Consequently $P_0$ and $P_m$ are orthogonal.*

LEMMA 5b. *If, for every integer $n$, the restriction of $m$ to $T_n$ belongs to $H(K_n)$, then $P_m^{(n)}$ is absolutely continuous with respect to $P_0^{(n)}$ with probability density function*

$$\frac{dP_m^{(n)}}{dP_0^{(n)}} = \exp\left\{(X, m)_n - \frac{1}{2}(m, m)_n\right\}. \tag{5.8}$$

*Further, for any $\alpha > 0$,*

$$\Delta_n(\alpha) = E_{P_0}\left[\left(\frac{dP_m^{(n)}}{dP_0^{(n)}}\right)^\alpha\right] = \exp\left\{\frac{\alpha(\alpha - 1)}{2}(m, m)_n\right\}. \tag{5.9}$$

LEMMA 5c. *If, for every integer $n$, the restriction of $m$ to $T_n$ belongs to $H(K_n)$, then*

$$\lim_{n\to\infty} (m, m)_n = \infty \qquad \text{if} \quad m \notin H(K) \tag{5.10}$$

$$\lim_{n\to\infty} (m, m)_n = (m, m)_K < \infty \qquad \text{if} \quad m \in H(K). \tag{5.11}$$

LEMMA 5d. *If $m$ belongs to $H(K)$, then $\lim_{n\to\infty}\Delta_n(2) < \infty$. By martingale theory, it follows that $P_m$ is absolutely continuous with respect to $P_0$, with probability density function*

$$\frac{dP_m}{dP_0} = \lim_{n\to\infty} \frac{dP_m^{(n)}}{dP_0^{(n)}} = \exp\left\{(X, m)_K - \frac{1}{2}(m, m)_K\right\} \tag{5.12}$$

*where the limit exists with probability one. Further, $P_0$ and $P_m$ are equivalent.*

LEMMA 5e. *If $m$ does not belong to $H(K)$, then $\lim_{n\to\infty}\Delta_n(1/2) = \infty$. Consequently, $P_0$ and $P_m$ are orthogonal* (by a theorem of Kraft [9]).

Theorem 5A has the following important consequence. Let $\{X(t), t \in T\}$ be a normal time series with known proper covariance function $K$ and unknown mean value function $m(t) = E[X(t)]$ belonging to a known class $M$ of functions. We have defined $L_2[X(t), t \in T]$ to be the Hilbert space consisting of all random variables $U$ which may be represented either as a finite linear combination

$$U = \sum_{i=1}^{n} c_i X(t_i) \tag{5.13}$$

for some integer $n$, points $t_1, \cdots, t_n$ in $T$, and real numbers $c_1, \cdots, c_n$ or as a limit in quadratic mean of such finite linear combinations under the inner product $(U, V)$ defined by

$$(U, V) = E[UV] = \text{Cov}[U, V] + E_m[U]E_m[V]. \tag{5.14}$$

The subscript $m$ on an expectation operator $E$ is written to indicate that the expectation is computed under the assumption that $m(\cdot)$ is the true mean value function.

The fact that the inner product in (5.14) depends on the true value of $m(\cdot)$ has the following consequence: the random variables belonging to the Hilbert space $L_2[X(t), t \in T]$ may not be the same for all values of $m(\cdot)$. This difficulty does not arise if $T$ is a finite set, for then $L_2[X(t), t \in T]$ consists of all random variables $U$ of the form

$$U = \sum_{t \in T} c_t X(t) \tag{5.15}$$

for some real constants $c_t$. However if $T$ is infinite, it has to be assumed that the space of random variables constituting $L_2[X(t), t \in T]$ is the same for all values of $m(\cdot)$ in the space $M$ of admissible mean value functions. If it is assumed that $M$ is a subset of $H(K)$, and that assumption 5A holds, it then follows from theorem 5A that the Hilbert space $L_2[X(t), t \in T]$, regarded as a space of random variables, is the same for all $m$ in $M$. Further, one can define a one-one correspondence between $L_2[X(t), t \in T]$ and $H(K)$ so that if $(X, g)_K$ denotes the random variable corresponding to $g$ in $K$, then for every $t$ in $T$ and $h, g$ in $H(K)$

$$[X, K(\cdot, t)]_K = X(t), \tag{5.16}$$

$$E_m[(X, g)_K] = (m, g)_K \qquad \text{for all } m \text{ in } M, \tag{5.17}$$

$$\text{Cov}\,[(X, g)_K, (X, h)_K] = (g, h)_K. \tag{5.18}$$

## 6. Regression analysis of normal time series

Using the concrete formula for the probability density functional of a normal process provided by (5.3) there is no difficulty in applying the concepts of classical statistical methodology to problems of inference on normal time series. In particular, let us consider the problem of regression analysis.

Let $\{X(t), t \in T\}$ be a normal time series with known proper covariance kernel $K(s, t) = \text{Cov}\,[X(s), X(t)]$ and whose mean value function is only assumed to belong to a known class $M$. If it is assumed that $M$ is a subset of the reproducing kernel space $H(K)$, then the probability measures $P_m$ are all equivalent.

The maximum likelihood estimate $m^*(\cdot)$ is defined as that estimate in the space $M$ of admissible mean value functions such that

$$f(X, m^*) = \max_{m \in M} f(X, m). \tag{6.1}$$

We assume that $M$ is a closed subspace of $H(K)$. Now under assumptions 5A, $H(K)$ is a separable Hilbert space. Consequently, let $\{w_j, j \in Q\}$ be a finite or countably infinite set of functions in $H(K)$ which are orthonormal and which span $M$; in symbols,

$$(w_i, w_j)_K = \delta(i, j), \tag{6.2}$$

and every function $m(\cdot)$ in $M$ may be written

$$m = \sum_{j \in Q} \beta_j w_j, \qquad \beta_j = (m, w_j)_K. \tag{6.3}$$

Consequently we may write

$$\log f(X, m) = \sum_{j \in Q} \beta_j (X, w_j)_K - \frac{1}{2} \sum_{j \in Q} \beta_j^2. \tag{6.4}$$

Differentiating with respect to $\beta_j$, we find that the values $\{\beta_j^*, j \in Q\}$ minimizing (6.4) as a function of $\{\beta_j, j \in Q\}$ are

$$\beta_j^* = (X, w_j)_K, \qquad j \in Q. \tag{6.5}$$

Consequently, the maximum likelihood estimates of $m(\cdot)$ is given by

$$m^*(\cdot) = \sum_{j \in Q} (X, w_j)_K w_j(\cdot). \tag{6.6}$$

This argument is easily justified if $Q$ is a finite set, as it often will be. If $Q$ is an infinite set, the situation is more complicated. The random variable $m^*(t)$, defined for each $t$ in $T$ by

$$m^*(t) = \sum_{j \in Q} (X, w_j)_K w_j(t) \tag{6.7}$$

is well defined because

$$E[|m^*(t)|^2] = \sum_{j \in Q} w_j^2(t) < \infty. \tag{6.8}$$

However, regarded as a function in $H(K)$, $m^*(\cdot)$ has, for almost all sample functions $X(\cdot)$, infinite norm since (using theorem B, p. 251, of Loève [11])

$$P_0[\|m^*(\cdot)\|_K^2 = \sum_{j \in Q} |(X, w_j)_K|^2 = \infty] = 1. \tag{6.9}$$

Thus $m^*(\cdot)$ does not belong to $H(K)$ and a maximum likelihood estimate can not be said to exist.

Nevertheless the estimate defined by (6.7) is still a desirable estimate, since it may be interpreted as the uniformly minimum variance unbiased estimate of the value $m(t)$ at a particular time $t$ of the mean value function $m(\cdot)$. We omit the proof of this fact which is shown in [12]. In the next section we treat the simpler problem of showing that $m^*(t)$ is the uniformly minimum variance unbiased *linear* estimate of $m(t)$, and give other formulas for $m^*(t)$ and $\operatorname{Var}[m^*(t)]$.

## 7. Minimum variance linear unbiased estimation of the mean value function

Let $\{X(t), t \in T\}$ be a time series whose proper covariance kernel

$$K(s, t) = \operatorname{Cov}[X(s), X(t)] \tag{7.1}$$

is known. The mean value function

$$m(t) = E[X(t)] \tag{7.2}$$

is only assumed to belong to a known class $M$. One case of particular importance is when $M$ consists of all finite linear combinations of $q$ known functions $w_1(t), \cdots, w_q(t)$, so that the mean value function is of the form

$$m(t) = \beta_1 w_1(t) + \cdots + \beta_q w_q(t) \tag{7.3}$$

for unknowns $\beta_1, \cdots, \beta_q$ to be estimated.

In this section we consider the problem of estimating various functionals $\psi(m)$ of the true mean value function $m(\cdot)$ by estimates which (i) are linear in the observations $\{X(t), t \in T\}$ in the sense that they belong to $L_2[X(t), t \in T]$, (ii) are *unbiased*, in a sense to be defined, and (iii) have *minimum variance* among all linear unbiased estimates.

We assume that $M$ is a subset of $H(K)$. Further, we assume that between $L_2[X(t), t \in T]$ and $H(K)$ there exists a one-one linear mapping with the following properties: if $(h, X)_K$ denotes the random variable in $L_2[X(t), t \in T]$ which corresponds under the mapping to the function in $H(K)$, then for every $t$ in $T$, and $h$ and $g$ in $H(K)$,

$$[K(\cdot, t), X]_K = X(t), \tag{7.4}$$

$$E_m[(h, X)_K] = (h, m)_K \qquad \text{for all } m \text{ in } M, \tag{7.5}$$

$$\text{Cov}\,[(h, X)_K, (g, X)_K] = (h, g)_K. \tag{7.6}$$

The subscript $m$ on an expectation operator is written to indicate that the expectation is computed under the assumption that $m(\cdot)$ is the true mean value function.

A functional $\psi(m)$ is said to be linearly estimable if it possesses an unbiased linear estimate $(g, X)_K$. Since

$$E_m[(g, X)_K] = (g, m)_K = \psi(m) \qquad \text{for all } m \text{ in } M \tag{7.7}$$

it follows that $\psi(m)$ is linearly estimable if and only if there exists a function $g$ in $H(K)$ satisfying (7.7). Now the variance of a linear estimate is given by

$$\text{Var}\,[(g, X)_K] = (g, g)_K. \tag{7.8}$$

Consequently finding the minimum variance unbiased linear estimate $\psi^* = (g^*, X)_K$ of $\psi(m)$ is equivalent to finding that function $g^*$ in $H(K)$ which has minimum norm among all functions $g$ satisfying the restraint (7.7). The solution to this problem is given by the projection theorem in Hilbert space.

PROJECTION THEOREM. *Let $H$ be an abstract Hilbert space, let $M$ be a Hilbert subspace of $H$, let $v$ be a vector in $H$, and let $v^*$ be a vector in $M$. A necessary and sufficient condition that $v^*$ be the unique vector in $M$ satisfying*

$$\|v^* - v\| = \min_{u \text{ in } M} \|u - v\| \tag{7.9}$$

*is that*

$$(v^*, u) = (v, u) \qquad \textit{for every } u \textit{ in } M. \tag{7.10}$$

The vector $v^*$ satisfying (7.10) is called the projection of $v$ onto $M$, and is also written $E^*[v|M]$. The vector $E^*[v|M]$ may also be characterized as the vector $v^*$

in $H$ satisfying (7.10) which has minimum norm $\|v^*\|$ among all vectors $v^*$ satisfying (7.10).

THEOREM 7A. *The uniformly minimum variance unbiased linear estimate* $\psi^*$ *of a linearly estimable function* $\psi(m)$ *is given by*

$$\psi^* = [E^*(g|\overline{M}), X]_K \tag{7.11}$$

*with variance*

$$\text{Var}(\psi^*) = \|E^*(g|\overline{M})\|_K^2, \tag{7.12}$$

*where* $g$ *is any function satisfying* (7.7), $\overline{M}$ *is the smallest Hilbert subspace of* $H(K)$ *containing* $M$, *and* $E^*(g|\overline{M})$ *denotes the projection onto* $\overline{M}$ *of* $g$.

In particular the uniformly minimum variance unbiased linear estimate $m^*(t)$ of the value $m(t)$ at a particular point $t$ of the mean value function $m(\cdot)$ is given by

$$m^*(t) = \{E^*[K(\cdot, t)|\overline{M}], X\}_K \tag{7.13}$$

since

$$m(t) = [K(\cdot, t), m]_K. \tag{7.14}$$

It may be verified that (7.13) and (6.7) coincide.

We next consider the special case that $M$ consists of all functions of the form of (7.3). Given an estimable linear function $\psi(\beta)$ of the parameters $\beta_1, \cdots, \beta_q$

$$\psi(\beta) = \psi_1\beta_1 + \cdots + \psi_q\beta_q \tag{7.15}$$

where the constants $\psi_1, \cdots, \psi_q$ are known, the minimum variance unbiased linear estimate of $\psi(\cdot)$ is

$$\psi^* = \psi_1\beta_1^* + \cdots + \psi_q\beta_q^* \tag{7.16}$$

where $\beta_1^*, \cdots, \beta_q^*$ are any solution of the set of normal equations

$$\begin{bmatrix} (w_1, w_1)_K & \cdots & (w_1, w_q)_K \\ \cdot & & \\ \cdot & \cdots & \\ \cdot & & \\ (w_q, w_1)_K & \cdots & (w_q, w_q)_K \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \cdot \\ \cdot \\ \cdot \\ \beta_q^* \end{bmatrix} = \begin{bmatrix} (w_1, X)_K \\ \cdot \\ \cdot \\ \cdot \\ (w_q, X)_K \end{bmatrix}. \tag{7.17}$$

If the functions $w_1, \cdots, w_q$ are linearly independent as functions in $H(K)$, then one may write explicitly

$$m^*(t) = -\frac{1}{W} \begin{vmatrix} (w_1, w_1)_K & \cdots & (w_1, w_q)_K & (X, w_1)_K \\ \cdot & & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ (w_q, w_1)_K & \cdots & (w_q, w_q)_K & (X, w_q)_K \\ w_1(t) & \cdots & w_q(t) & 0 \end{vmatrix}, \tag{7.18}$$

$$(7.19)\qquad \operatorname{Var}[m^*(t)] = -\frac{1}{W}\begin{vmatrix} (w_1, w_1)_K & \cdots & (w_1, w_q)_K & w_1(t) \\ \cdot & & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ (w_q, w_1)_K & \cdots & (w_q, w_q)_K & w_q(t) \\ w_1(t) & \cdots & w_q(t) & 0 \end{vmatrix},$$

where

$$(7.20)\qquad W = \begin{vmatrix} (w_1, w_1)_K & \cdots & (w_1, w_q)_K \\ \cdot & & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & & \cdot \\ (w_q, w_1)_K & \cdots & (w_q, w_q)_K \end{vmatrix}.$$

## 8. Hypothesis testing and simultaneous confidence bands for mean value functions

If the time series $X(t)$ is assumed to be normal, or if all linear functionals $(h, X)_K$ may be assumed to be approximately normally distributed, then one may state a confidence band for the entire mean value function $m(\cdot)$ as follows. Given a confidence level $\alpha$, let $C_q(\alpha)$ denote the $\alpha$ percentile of the $\chi^2$ distribution with $q$ degrees of freedom, that is,

$$(8.1)\qquad P[\chi_q^2 \geqq C_q(\alpha)] = \alpha.$$

In particular, for $q = 2$ and $\alpha = 0.95$, $C_q(\alpha)$ is approximately 6.

We now show that if the space $M$ of possible mean value functions has finite dimension $q$, then

$$(8.2)\qquad m^*(t) - [C_q(\alpha)]^{1/2}\sigma[m^*(t)] \leqq m(t) \leqq m^*(t) + [C_q(\alpha)]^{1/2}\sigma[m^*(t)],$$

for all $t$ in $-\infty < t < \infty$, is a simultaneous confidence band for all values of the mean value function with a level of significance not less than $\alpha$, that is, if $m(\cdot)$ is the true mean value function then (8.2) holds with a probability greater than or equal to $\alpha$.

To prove (8.2) we prove a more general theorem.

THEOREM 8A. *Simultaneous confidence interval of significance level $\alpha$ for all estimable functions $(m, g)$: for all $m$ in $M$*

$$(8.3)\qquad P_m\left(\sup_{g\in H(K)} \frac{|[X, E^*(g|\overline{M})]_K - (m, g)_K|^2}{\operatorname{Var}\{[X, E^*(g|\overline{M})]_K\}} \leqq C_q(\alpha)\right) = \alpha.$$

PROOF. Let $w_1, \cdots, w_q$ be linearly independent functions which span $M$. Then we may write $m = \beta_1 w_1 + \cdots + \beta_q w_q$ where $\beta_1, \cdots, \beta_q$ are functions of $m$. Further,

$$(8.4)\qquad \begin{aligned} (m, g)_K &= \alpha_1\beta_1 + \cdots + \alpha_q\beta_q \\ [X, E^*(g|\overline{M})]_K &= \alpha_1\beta_1^* + \cdots + \alpha_q\beta_q^*, \end{aligned}$$

where $\alpha_j = (w_j, g)_K$ for $j = 1, \cdots, q$, and $\beta_1^*, \cdots, \beta_q^*$ are the solution of the normal equations

$$\sum_{k=1}^{q} W_{jk}\beta_k^* = (X, w_j)_K, \qquad j = 1, \cdots, q \tag{8.5}$$

in which $W_{jk} = (w_j, w_k)_K$. Next the random variable appearing in (8.3) is equal to, letting $\{W^{jk}\}$ denote the inverse matrix of $\{W_{jk}\}$,

$$\sup_{-\infty<\alpha_1,\cdots,\alpha_q<\infty} \frac{\left|\sum_{j=1}^{q} \alpha_j(\beta_j^* - \beta_j)\right|^2}{\sum_{j,k=1}^{q} \alpha_j W^{jk}\alpha_k} = \sum_{j,k=1}^{q} (\beta_j^* - \beta_j)W_{jk}(\beta_k^* - \beta_k) \tag{8.6}$$

which is distributed as $\chi_q^2$ (which is immediate if one takes $W_{jk} = \delta(j, k)$; compare Scheffé [14], p. 416).

From the foregoing proof we obtain immediately the useful fact that for every $m$ in $M$

$$P_m[||m^*(t) - m(t)||_K^2 \leqq C_q(\alpha)] = \alpha. \tag{8.7}$$

To prove (8.7) one need only note that the random variable in (8.7) is equal to the right side of (8.6), and consequently, is distributed as $\chi_q^2$. Using (8.7) one may construct a test of the hypothesis that the mean value function $m(t)$ is identically 0 against the alternative that it belongs to the $q$-dimensional subspace $M$.

More generally, given a $q$-dimensional subspace $M$ of $H(K)$, and a $q'$-dimensional subspace $M'$ of $M$, to test the composite null hypothesis

$$H_0: m(\cdot) \in M' \tag{8.8}$$

against the composite alternative hypothesis

$$H_1: m(\cdot) \in M \tag{8.9}$$

one may use the statistic

$$\Delta = ||m_M^*(t) - m_{M'}^*(t)||_K^2 \tag{8.10}$$

where $m_M^*(t)[m_{M'}^*(t)]$ denotes the minimum variance unbiased linear estimate of $m(t)$ under the hypothesis $H_1[H_0]$. It may be shown that, under $H_0$, $\Delta$ is distributed as $\chi^2$ with $q - q'$ degrees of freedom.

## 9. Iterative evaluation of reproducing kernel inner products

In this section we give an iterative method of evaluating the reproducing kernel inner product $(h, h)_K$ and corresponding random variable $(h, X)_K$ for time series $\{X(t), a \leqq t \leqq b\}$. The method given makes possible the approximate synthesis of minimum variance unbiased linear estimates, assuming a known covariance kernel $K$ which can be of any form and can be known either analytically or numerically. The method to be described is a gradient method and can be extensively generalized. It is described here in order to establish the feasibility of iterative methods.

Let $K(s, t)$ be a covariance kernel, defined for $a \leqq s, t \leqq b$. Let $H(K)$ be the corresponding reproducing kernel Hilbert space. Let $C(a, b)$ be the space of continuous functions on the interval $a$ to $b$.

Given a function $h$ in $H(K)$, it is of interest to develop methods of generating sequences $\{H_n\}$ of functions in $C(a, b)$ having the properties that

$$\lim_{n\to\infty} E\left[\left|(X, h)_K - \int_a^b H_n(t)X(t)\,dt\right|^2\right] = 0 \tag{9.1}$$

$$(h, h)_K = \lim_{n\to\infty} \int_a^b \int_a^b H_n(s)K(s, t)H_n(t)\,ds\,dt. \tag{9.2}$$

It is easily shown that sequences $\{H_n\}$ satisfying (9.1) and (9.2) exist. As in section 4, let $\lambda_n$ be the eigenvalues, arranged in decreasing order, $\lambda_1 \geqq \lambda_2 \geqq \cdots$, and let $\varphi_n(\cdot)$ be the corresponding eigenfunctions of the kernel $K(s, t)$. Then a function $h$ belongs to $H(K)$ if and only if

$$(h, h)_K = \sum_{n=1}^{n} \frac{1}{\lambda_n}\left|\int_a^b h(t)\varphi_n(t)\,dt\right|^2 < \infty. \tag{9.3}$$

Consequently, define

$$H_n(t) = \sum_{k=1}^{n} \varphi_k(s)\,\frac{1}{\lambda_k}\int_a^b h(s)\varphi_k(s)\,ds. \tag{9.4}$$

Clearly $H_n(\cdot)$ belongs to $C(a, b)$.

It may be verified that

$$\int_a^b \int_a^b H_n(s)K(s, t)H_n(t)\,ds\,dt = \sum_{k=1}^{n} \frac{1}{\lambda_k}\left|\int_a^b h(t)\varphi_k(t)\,dt\right|^2 \tag{9.5}$$

and

$$\int_a^b H_n(t)X(t)\,dt = \sum_{k=1}^{n} \frac{1}{\lambda_k}\int_a^b h(s)\varphi_k(s)\,ds \int_a^b X(t)\varphi_k(t)\,dt. \tag{9.6}$$

Therefore the sequence defined by (9.4) satisfies (9.1) and (9.2). However, it is not computationally convenient to use (9.4), inasmuch as it involves the calculation of eigenvalues and eigenfunctions.

Define a transformation $T$ on functions $H$ in $C(a, b)$ as follows:

$$TH(t) = \int_a^b H(s)K(s, t)\,ds, \qquad a \leqq t \leqq b. \tag{9.7}$$

We may then write

$$\int_a^b H(t)X(t)\,dt = (TH, X)_K \tag{9.8}$$

$$\int_b^a \int_a^b H(s)K(s, t)H(t)\,ds\,dt = (TH, TH)_K. \tag{9.9}$$

Next, define a sequence of functions $H_n$ as follows. Let $\alpha$ be a constant to be specified. Let $H_0(t) = 1$, or some other function in $C(a, b)$. For $n \geqq 1$, let

$$H_{n+1} = H_n - \alpha(TH_n - h). \tag{9.10}$$

We claim that if $\alpha$ is chosen in an interval specified by (9.18) or (9.21), then the sequence $H_n$ defined by (9.10) satisfies (9.1) and (9.2). To prove this assertion it suffices to prove that

$$E[|(h, X)_K - (TH_n, X)_K|^2] = ||(h - TH_n)||_K^2 \to 0 \qquad \text{as} \quad n \to \infty. \tag{9.11}$$

From (9.10) we may write

$$\begin{aligned} TH_{n+1} - h &= (TH_n - h) - \alpha T(TH_n - h) \\ &= (I - \alpha T)(TH_n - h) \end{aligned} \tag{9.12}$$

where $I$ is the identity operator, $Ih(t) = h(t)$. From (9.12) it follows that for $n \geqq 0$

$$TH_n - h = (I - \alpha T)^n(TH_0 - h). \tag{9.13}$$

We next note that for any function $g$ in $H(K)$,

$$g(t) = \sum_{n=1}^{\infty} \varphi_n(t) \int_a^b \varphi_n(s)g(s)\, dt, \tag{9.14}$$

$$Tg(t) = \sum_{n=1}^{\infty} \varphi_n(t)\lambda_n \int_a^b \varphi_n(s)g(s)\, ds, \tag{9.15}$$

$$||(I - \alpha T)g||_K^2 = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \left\{\int_a^b \varphi_n(s)g(s)\, ds\right\}^2 \{1 - \alpha\lambda_n\}^2. \tag{9.16}$$

From (9.13) and (9.16) it follows that, defining $g = TH_0 - h$ and $\gamma_n = \int_a^b \varphi_n(s)g(s)\, ds$,

$$||TH_n - h||_K^2 = \sum_{m=1}^{\infty} \frac{1}{\lambda_m} \gamma_m^2 \{1 - \alpha\lambda_m\}^{2n}. \tag{9.17}$$

Let $\alpha$ be chosen so that, for every integer $m$,

$$-1 < 1 - \alpha\lambda_m < 1 \quad \text{or} \quad 0 < \alpha < \frac{2}{\lambda_m}. \tag{9.18}$$

If (9.18) holds, then for any integer $M$

$$||TH_n - h||_K^2 \leqq \sum_{m=1}^{M} \frac{1}{\lambda_m} \gamma_m^2 \{1 - \alpha\lambda_m\}^{2n} + \sum_{m>M} \frac{1}{\lambda_m} \gamma_m^2 \tag{9.19}$$

which tends to 0 as one first lets $n$ tend to $\infty$, and then lets $M$ tend to $\infty$ (note that the last term in (9.19) is the remainder term of a convergent series). We have thus showed that if (9.18) is satisfied then (9.11) holds. Further the procedure converges monotonically, in the sense that

$$||TH_{n+1} - h||_K \leqq ||TH_n - h||_K. \tag{9.20}$$

If $M$ is a constant such that $\max_m \lambda_m < M$, then (9.18) is satisfied if one chooses $\alpha$ so that

$$0 < \alpha \leqq 2/M. \tag{9.21}$$

A convenient choice for $M$ is

$$M = \sum_{m=1}^{\infty} \lambda_m = \int_a^b K(t, t)\, dt. \tag{9.22}$$

It should be remarked that (9.19) implies that

$$\lim_{n \to \infty} \int_a^b |(TH_n - h)(t)|^2\, dt = 0 \tag{9.23}$$

since for any $g$ in $H(K)$

$$\begin{aligned} |g(t)|^2 &\leqq \|g\|_K^2\, K(t, t) \\ \int_a^b |g(t)|^2 &\leqq \|g\|_K^2 \int_a^b K(t, t)\, dt. \end{aligned} \tag{9.24}$$

## REFERENCES

[1] N. Aronszajn, "La théorie générale des noyaux réproduisants et ses applications," *Proc. Cambridge Philos. Soc.*, Vol. 39 (1944), pp. 133–153.

[2] ———, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, Vol. 68 (1950), pp. 337–404.

[3] J. L. Doob, *Stochastic Processes*, New York, Wiley, 1953.

[4] J. Feldman, "Equivalence and perpendicularity of Gaussian processes," *Pacific J. Math.*, Vol. 8 (1958), pp. 699–708; correction, Vol. 9 (1959), pp. 1295–1296.

[5] U. Grenander, "Stochastic processes and statistical inference," *Ark. Mat.*, Vol. 1 (1950), pp. 195–277.

[6] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, Berkeley and Los Angeles, University of California Press, 1958.

[7] J. Hájek, "On a property of normal distribution of any stochastic process," *Czechoslovak Math. J.*, Vol. 8 (1958), pp. 610–618. (In Russian.)

[8] K. Karhunen, "Uber linear Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fenn., Ser. A I.*, No. 37 (1947), 79 pp.

[9] C. Kraft, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. California Publ. Statist.*, Vol. 2 (1955), pp. 125–142.

[10] M. Loève, "Fonctions aléatoires du second ordre," supplement to P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Paris, Gauthier-Villars, 1948.

[11] ———, *Probability Theory*, New York, Van Nostrand, 1955.

[12] E. Parzen, "Statistical inference on time series by Hilbert space methods, I," Department of Statistics, Stanford University, Technical Report No. 23, January 2, 1959.

[13] T. S. Pitcher, "Likelihood ratios of Gaussian process," *Ark. Mat.*, Vol. 4 (1960), pp. 35–44.

[14] H. Scheffé, *The Analysis of Variance*, New York, Wiley, 1959.

[15] C. Striebel, "Densities for stochastic processes," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 559–576.

[16] P. Whittle, *Hypothesis Testing in Time Series Analysis*, Uppsala, Almqvist & Wiksell, 1951.

[17] ———, "The analysis of multiple stationary time series," *J. Roy. Statist. Soc., Ser. B*, Vol. 15 (1952), pp. 125–139.