The Annals of Probability
1976, Vol. 4, No. 1, 98–101

PREDICTION THEORY AND ERGODIC SPECTRAL DECOMPOSITIONS[1]

BY BENNETT EISENBERG

Lehigh University

Linear prediction theory for a stationary sequence X_n ordinarily begins with the assumption that the covariance $R(n)=E(X_{m+n}\bar{X}_m)=\int_{-\pi}^{\pi} e^{i\lambda n}\,dF(\lambda)$ is known. The best linear predictor of X_0 given the past $X_{-1}, X_{-2}, \cdots$ is then the projection ψ of X_0 on the span of $X_{-1}, X_{-2}, \cdots$. The prediction error is $E(|X_0 - \psi|^2)$.

In practice R is not known but is estimated from the past. If the process is ergodic and the entire past is known this causes no problem since then the estimate $\hat{R}$ of R must equal R. But if the process is not ergodic then $\hat{R}$ does not equal R. In this paper we consider the relationship between prediction using $\hat{R}$ and R. One conclusion is that if the process is Gaussian, it doesn't matter whether $\hat{R}$ or R is used in constructing the best linear predictor. The predictor is the same and the prediction error is the same.

1. The ergodic spectral decomposition. Assume $X_n(\omega)$ is a second order strictly stationary process on $(\Omega, \mathscr{F}, P)$ with $TX_n(\omega) = X_n(\tau\omega) = X_{n+1}(\omega)$. Let $\mathscr{I}$ be the σ field of invariant sets modulo a null set. Define the estimate $\hat{R}(k)$ of $R(k)$ as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X_{-n}(\omega)\bar{X}_{-n-k}(\omega)\,,$$

which exists a.s. by the ergodic theorem. The first theorem describes the ergodic spectral decomposition.

THEOREM 1.

1. $\hat{R}(k) = E(X_{n+k}\bar{X}_n \mid \mathscr{I})$.
2. $\hat{R}(k)$ *is a positive definite sequence* a.s.
3. $E(X_{n+k}\bar{X}_n \mid \hat{R}(j), j = 0, \pm 1, \pm 2, \cdots) = \hat{R}(k)$.
4. $F(\lambda) = E(\hat{F}(\lambda, \omega))$, *where* $\hat{R}(k, \omega) = \int_{-\pi}^{\pi} e^{i\lambda k}\,d\hat{F}(\lambda, \omega)$.

PROOF. (1) This is a version of Theorem 6.28 in Breiman (1968) obtained by substituting $X_{n+k}\bar{X}_n$ for X_k.

(2) This follows from the positivity of conditional expectation. Namely,

$$\sum_j \sum_k C_k \bar{C}_j \hat{R}(k-j) = E(|\sum C_k X_k|^2 \mid \hat{R}) \geqq 0 \quad \text{a.s.}$$

This holds simultaneously for all finite sets of rational $C_1, \cdots, C_n$ and then by continuity for all $C_1, \cdots, C_n$.

Received August 26, 1974; revised May 15, 1975.

[1] Research supported in part by N.S.F. grant GP-37771.

AMS 1970 *subject classifications.* 60G10, 60G15, 60G25.

Key words and phrases. Spectral decomposition, linear prediction, stationary sequences.

(3) $\hat{R}(j)$ for $j = 0, \pm 1, \pm 2, \cdots$ are measurable with respect to $\mathscr{I}$. Hence

$$\begin{aligned} E(X_{n+k}\bar{X}_n \mid \hat{R}(j), j = 0, \pm 1, \pm 2, \cdots) \\ &= E(E(X_{n+k}\bar{X}_n \mid \mathscr{I}) \mid \hat{R}(j), j = 0, \pm 1, \pm 2, \cdots) \\ &= E(\hat{R}(k) \mid \hat{R}(j), j = 0, \pm 1, \pm 2, \cdots) = \hat{R}(k). \end{aligned}$$

(4) $\int e^{i\lambda k}\, d\hat{F}(\lambda, \omega) = \hat{R}(k, \omega)$ is measurable for each k and since $E(\hat{R}(k, \omega)) = R(k)$ we have

$$\int \left(\int e^{i\lambda k}\, d\hat{F}(\lambda, \omega)\right) dP(\omega) = \int e^{i\lambda k}\, dF(\lambda)\,.$$

Taking limits of linear combinations $\sum C_k e^{i\lambda k}$ we get

$$\iint g(\lambda)\, d\hat{F}(\lambda, \omega)\, dP(\omega) = \int g(\lambda)\, dF(\lambda)$$

for bounded measurable $g(\lambda)$. In particular letting $g(\lambda)$ be the characteristic function of $[-\pi, \lambda_0]$ we have

$$\int \hat{F}(\lambda_0, \omega)\, dP(\omega) = F(\lambda_0)\,. \qquad \square$$

Theorem 1 is closely related to the decomposition theorems of von Neumann–Choquet, Blum–Hanson, and Varadarajan under varying conditions on the measure space. In the language of Varadarajan (1963), pages 203–205, we have that for f in $L^1[dP]$, $\int f\, dP = \iint f(\omega')\, d\beta_\omega(\omega')\, dP(\omega)$, where β_ω are ergodic measures and $\int f(\omega')\, d\beta_\omega(\omega') = E(f \mid \mathscr{I})$. For $f = X_k\bar{X}_0$ this says $R(k) = \iint X_k\bar{X}_0\, d\beta_\omega\, dP$, where $\int X_k\bar{X}_0\, d\beta_\omega = \hat{R}(k, \omega)$.

Example. Let X_n have covariance $R(n) = \sum |A_\lambda|^2 e^{in\lambda}$ where the sum is over a finite set of λ. The spectral representation of X_n then becomes

$$X_n(\omega) = \sum A_\lambda e^{in\lambda}\phi_\lambda(\omega) \qquad \text{and}$$

$$\frac{1}{N}\sum_{n=1}^N X_{n+k}\bar{X}_n = \sum_\lambda \sum_{\lambda'} \frac{1}{N}\sum_{n=1}^N A_\lambda \overline{A_{\lambda'}} e^{in(\lambda-\lambda')}\phi_\lambda(\omega)\overline{\phi_{\lambda'}(\omega)} e^{i\lambda k}$$

which approaches $\sum_\lambda |A_\lambda|^2|\phi_\lambda(\omega)|^2 e^{i\lambda k} = \hat{R}(k, \omega)$. Thus $\hat{F}$ has point masses of magnitude $|A_\lambda|^2|\phi_\lambda(\omega)|^2$ at the same set of λ as F. In fact, from the ergodic spectral decomposition $F(\lambda) = \int \hat{F}(\lambda, \omega)\, dP(\omega)$, we see this is the only possible type of decomposition for a process with discrete spectrum.

2. Applications to prediction. The best linear predictor of X_0 based on X_{-1}, $X_{-2}, \cdots$ is the ψ in the closure of the span of $X_{-1}, X_{-2}, \cdots$ such that $E(|\psi - X_0|^2) = \inf E(|\varphi - X_0|^2)$, φ of the form $\sum_{k=1}^M a_k X_{-k}$. The prediction error $E(|\psi - X_0|^2)$ is known to be $\exp(\int_{-\pi}^{\pi} \log F'(\lambda)\, d\lambda)$ where F' is the derivative of the absolutely continuous component of F.

Proposition. $\exp(\int \log F'\, d\lambda) \geqq E(\exp(\int \log \hat{F}'\, d\lambda))$.

Proof.

$$\begin{aligned} \inf E(|\varphi - X_0|^2) &= \inf E(E(|\varphi - X_0|^2 \mid \hat{R})) \\ &\geqq E\,(\inf\nolimits_\varphi E(|\varphi - X_0|^2 \mid \hat{R}))\,, \end{aligned}$$

where the inf is over φ of the form $\sum_{k=1}^M a_k X_{-k}$. $\square$

COROLLARY. *If* $\exp(\int_{-\pi}^{\pi} \log \hat{F}' \, d\lambda) = 0$ *then for almost every* $\hat{F}$ *in the ergodic spectral decomposition of* F, $\exp(\int_{-\pi}^{\pi} \log \hat{F} \, d\lambda) = 0$.

If X_n is assumed to be Gaussian much more can be said.

THEOREM 2. *If* X_n *is a mean zero stationary Gaussian sequence with covariance* R, *then*

(1) *For almost every* $\hat{R}$ *the best linear predictor of* X_0 *is the same as the best linear predictor under* R.

(2) *The prediction errors are the same under* R *and* $\hat{R}$ *for almost every* $\hat{R}$.

The proof relies on the following lemma.

LEMMA. *Assume* T *is an operator on a Hilbert space* H *of functions* $f(\omega)$ *induced by a measure preserving transformation having discrete spectrum. (*T *need not be ergodic). Assume* U *is an operator on a Hilbert space* K *of functions* $g(\omega')$ *induced by a weakly mixing transformation. Then the functions invariant under* $T \otimes U$ *are of the form* $f(\omega)$ *where* $f(\omega)$ *is invariant under* T.

PROOF. Let $Tf_j = f_j$ for j in M and $Tf_j = e^{i\lambda j} f_j$ for j not in M. Let g_k be an arbitrary basis for K. Then $H \otimes K$ consists of functions $\sum a_{jk} f_j(\omega) g_k(\omega')$, $\sum |a_{jk}|^2 < \infty$. If $T \otimes U \sum a_{jk} f_j(\omega) g_k(\omega') = \sum a_{jk} f_j(\omega) g_k(\omega')$ then

$$\sum_{j \text{ in } M} a_{jk} f_j(\omega) U g_k(\omega') + \sum_{j \text{ not in } M} a_{jk} e^{i\lambda j} f_j(\omega) U g_k(\omega')$$
$$= \sum_{j \text{ in } M} a_{jk} f_j(\omega) g_k(\omega') + \sum_{j \text{ not in } M} a_{jk} f_j(\omega) g_k(\omega') .$$

Since the coefficient of $f_j(\omega)$ must be the same on both sides we have $U(\sum_k a_{jk} g_k(\omega')) = \sum_k a_{jk} g_k(\omega')$ for j in M, $e^{i\lambda j} U(\sum_k a_{jk} g_k(\omega')$ for j not in M. Since U comes from a weakly mixing transformation this says $\sum_k a_{jk} g_k(\omega') =$ constant for j in M and $\sum_k a_{jk} g_k(\omega') = 0$ for j not in M. In other words any h invariant under $T \otimes U$ is of the form $\sum_{j \text{ in } M} b_j f_j(\omega)$. □

PROOF OF THEOREM 2. Write $F(\lambda) = F_c(\lambda) + F_d(\lambda)$ where F_c is continuous and F_d discrete. Then $X_n = T_n + U_n$ where U_n has spectral distribution F_c and T_n is independent of U_n with spectral distribution F_d. $U_n(\omega')$ is weakly mixing and by results in Blum–Eisenberg (1974), the underlying transformation T of T_n has discrete spectrum. Hence by the lemma the subspace of functions $h(\omega, \omega')$ invariant under the transformation $T \otimes U$ consists of functions $f(\omega)$ invariant under T. The U_n process is thus independent of the σ-field generated by the T_n process and the invariant sets $\mathscr{I}$. Hence

$$\begin{aligned} &E((T_k + U_k)(\bar{T}_0 + \bar{U}_0) \,|\, \mathscr{I}) \\ &\quad = E(T_k \bar{T}_0 \,|\, \mathscr{I}) + E(T_k \bar{U}_0 \,|\, \mathscr{I}) + E(U_k \bar{T}_0 \,|\, \mathscr{I}) + E(U_k \bar{U}_0 \,|\, \mathscr{I}) \\ &\quad = \hat{R}_T(k) + E(\bar{U}_0) E(T_k \,|\, \mathscr{I}) + E(U_k) E(\bar{T}_0 \,|\, \mathscr{I}) + R_U(k) \\ &\quad = \hat{R}_T(k) + R_U(k) . \end{aligned}$$

That is, the components of the ergodic decomposition of R_X are $\hat{R}_T(k)$ and $R_U(k)$. But T_n has discrete spectrum so $\hat{R}_T$ must have spectral measure concentrated on the same points as R_T. Now the best predictor of a process with discrete spectrum and continuous spectrum can be found by filtering out the discrete spectrum component and predicting its future perfectly and then adding the predictor for the continuous spectrum component. But the best predictor for a process with covariance R_T or $\hat{R}_T$ is the same since it depends only on the location of the atoms in the spectral measure. The prediction error will be zero. Since $R_U = \hat{R}_U$ the predictors for the continuous components will be the same and have the same prediction error. $\square$

REFERENCES

[1] BLUM, J. and EISENBERG, B. (1974). Spectral properties of processes derived from stationary Gaussian sequences. *Stochastic Processes Appl.* **2** 177–181.

[2] BLUM, J. and HANSON, D. (1960). On invariant probability measures I. *Pacific J. Math.* **10** 1125–1129.

[3] BREIMAN, L. (1968). *Probability*. Addison-Wesley Reading.

[4] MACKEY, G. (1974). Ergodic theory and its significance for probability and statistical mechanics. *Advances in Math.* **12** 178–268.

[5] VARADARAJAN, V. S. (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** 191–220.

DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015