The Annals of Probability
1978, Vol. 6, No. 4, 574-584

ROBUST LINEAR EXTRAPOLATIONS OF SECOND-ORDER STATIONARY PROCESSES

BY YUZO HOSOYA

Tohoku University

This paper considers the problems of linear prediction under the condition that the spectral structures of second-order stationary processes are vaguely specified. The approach adopted is closely in line with the theory of robust estimation due to Peter Huber. The paper shows that there exists a minimax one-step ahead predictor for the set of spectral distributions given as $\{H: H = (1-\varepsilon)F + \varepsilon G, G \in \mathscr{D}_1\}$, where F is a fixed probability distribution function and $\mathscr{D}_1$ is the set of all absolutely continuous probability distribution functions. That predictor turns out to be the optimal linear predictor for a spectral distribution which is derived by a suitable modification applied to F. Though there generally exists no minimax predictor for the set $\{H: H = (1-\varepsilon)F + \varepsilon G, G \in \mathscr{D}_0\}$ ($\mathscr{D}_0$ is the set of all probability distribution functions), a linear predictor is explicitly constructed so that its maximal prediction error is arbitrarily close to the lower bound of the maximal prediction errors of possible linear predictors. The results obtained in this paper would have an important application in the errors-in-variable models.

0. Introduction. This paper considers the linear prediction of second-order stationary processes under the condition that their spectral structure is vaguely specified. For the usual theory of linear prediction, due mainly to Kolmogorov and Wiener, to be effective, it is essential that the spectrum of a process concerned is completely known. The present paper intends to replace that assumption by a weaker one; namely, it assumes only that a certain neighborhood of a spectrum is given and constructs an optimal (in a sense specified below) linear predictor, based on the knowledge that the spectrum of a process belongs to that neighborhood. For this purpose, this paper adopts an approach that is almost in line with the theory of robust estimation due to Huber (1964), especially with his theory of minimax estimation of a location parameter for a class of contaminated normal distributions. In other words, the main purpose of the paper is the construction of a minimax predictor for a class of ε-contaminated spectral distributions.

To be more specific, assumed that the complex-valued stationary process $\{x_t: t = 0, \pm 1, \pm 2, \cdots\}$ has a spectral distribution H of the form $H(\omega) = (1-\varepsilon)F(\omega) + \varepsilon G(\omega)$, $-\pi < \omega \leqq \pi$, for a constant ε $(0 < \varepsilon < 1)$ where F is a known probability distribution function and G denotes a variable function which ranges over a certain class of probability distributions (F and G are supported

Received February 23, 1977; revised August 1977.

AMS 1970 *subject classifications.* Primary 60G25; Secondary 62G35.

Key words and phrases. Second-order stationary processes, linear prediction, robust estimation, prediction errors, Poisson kernel, boundary values.

by the interval $(-\pi, \pi]$). The assumption that $F(\pi) = G(\pi) = 1$ will not limit the generality of the following arguments, since otherwise the ratio $\varepsilon/(1-\varepsilon)$ could be suitably modified. A typical example pertaining to this model might be that the observed process $\{x_t\}$ is given as a mixture of two processes $\{y_t\}$ and $\{z_t\}$; the spectrum of $\{y_t\}$ is specified as F, whereas that of the z_t is unknown. Now the following notations are used throughout: Let $\mathscr{D}_0$ be the class of all probability distribution functions supported by the interval $(-\pi, \pi]$; let $\mathscr{D}_1 (\subset \mathscr{D}_0)$ be the class of probability distribution functions which are absolutely continuous with respect to the Lebesgue measure. In addition, denote by $\mathscr{D}_0(F)$ a subset of $\mathscr{D}_0$ such that any member H of $\mathscr{G}_0(F)$ is of the form: $H = (1-\varepsilon)F + \varepsilon G$ for some $G \in \mathscr{D}_0$. Similarly, $\mathscr{F}_1(F) = \{H \in \mathscr{D}_0 \colon H = (1-\varepsilon)F + \varepsilon G;\ G \in \mathscr{D}_1\}$.

For the purpose of determining the appropriate classes of linear predictors, denote by $\mathscr{L}(H)$ the completion of the linear hull of the set $\{e^{it\omega} \colon t \leqq -1\}$ with respect to the mean-square norm in the space of the square-integrable complex-valued functions defined on the measure space $\{(-\pi, \pi], \mathscr{B}, H\}$, where $\mathscr{B}$ is the Borel subsets of the interval $(-\pi, \pi]$ and H is a probability distribution. Though for a fixed H any element of $\mathscr{L}(H)$ may be regarded as a linear predictor, in the framework of the present problem it will be more natural to restrict the class of linear predictors to $\bigcap_{H \in \mathscr{F}_0(F)} \mathscr{L}(H)$ or to $\bigcap_{H \in \mathscr{F}_1(F)} \mathscr{L}(H)$. For the sake of brevity, write $\bigcap_{H \in \mathscr{F}_0(F)} \mathscr{L}(H)$ and $\bigcap_{H \in \mathscr{F}_1(F)} \mathscr{L}(H)$ respectively as $\mathscr{L}_0(F)$ and $\mathscr{L}_1(F)$. Call any element of $\mathscr{L}_0(F)$ or $\mathscr{L}_1(F)$ a linear predictor (it depends on contexts which class is to be referred to).

In order to predict x_ν (ν is a nonnegative integer) by the data $\{x_t \colon t \leqq -1\}$, the usual theory of prediction searches for the linear combination y of the data which makes the mean-square error $E|x_\nu - y|^2$ minimal. As is well known, this is equivalent to looking for $\phi \in \mathscr{L}(F)$ which minimizes $V(\phi, F) = \int_{-\pi}^{\pi} |e^{i\nu\omega} - \phi(\omega)|^2 \, dF(\omega)$ if the spectrum of the process $\{x_t\}$ is given as F. Suppose, however, that only partial information concerning the spectrum H of the process is available; formally, suppose that H varies over the range $\mathscr{F}_0(F)$ or over $\mathscr{F}_1(F)$. Then a set of mean-square prediction errors $\{V(\phi, H) \colon H \in \mathscr{F}_i(F)\}$ corresponds to each predictor $\phi \in \mathscr{L}_i(F)$, $i = 0, 1$. This paper adopts the minimax principle as the criterion of optimality, defining a minimax linear predictor as one which minimizes the maximal prediction error $\max_{H \in \mathscr{F}_i(F)} V(\phi, H)$. Consequently, the problem of robust prediction is reduced to the construction of a predictor $\phi^* \in \mathscr{L}_i(F)$ such that

$$(1) \qquad \max_{H \in \mathscr{F}_i(F)} V(\phi^*, H) = \min_{\phi \in \mathscr{L}_i(F)} \max_{H \in \mathscr{F}_i(F)} V(\phi, H) , \qquad i = 0, 1 .$$

As for the scope of the paper, it exclusively considers the prediction of one-step ahead, that is, the case where $\nu = 0$. There seems to be no straightforward generalization of the proceeding results, though in the case of the usual theory of linear prediction there arises no special difficulty in considering general ν-steps ahead prediction. Sections 1 and 2 are for establishing Theorems 1 and 2: Theorem 1 asserts that for $\mathscr{F}_1(F)$ with an absolutely continuous F there exists a minimax predictor ϕ_m which is the optimal linear predictor for a spectral density f_m obtained from f, density of F, after a suitable modification, while Theorem 2 deals with more general F. The results of Sections 1 and 2, however, do not apply to the class $\mathscr{F}_0(F)$. The difficulty is mainly that the robust predictor is given only as an a.e. point-wise limit so that there is no guarantee that $\phi_m \in \mathscr{L}_0(F)$. Section 3 considers robust prediction for the class $\mathscr{F}_0(F)$. Theorem 3 of that section shows that there is an approximate predictor to ϕ_m in $\mathscr{L}_0(F)$ such that its maximal prediction error is arbitrarily close to $\min_{\phi \in \mathscr{L}_1(F)} \max_{H \in \mathscr{F}_1(F)} V(\phi, H)$.

1. Minimax problem. Suppose first that the probability distribution function F is absolutely continuous with respect to the Lebesgue measure and has a density f. Let m be a positive number and let $E_m(f)$ and $F_m(f)$ be respectively subsets of the interval $(-\pi, \pi]$ such that $m \geqq (1-\varepsilon)f(\omega)$ for $\omega \in E_m(f)$ and $m < (1-\varepsilon)f(\omega)$ for $\omega \in F_m(f)$. Furthermore, define $P(f, m)$ as $P(f, m) = \int_{E_m(f)} \{m - (1-\varepsilon)f(\omega)\}\, d\omega$.

LEMMA 1. *There exists a unique m which satisfies the equation*:

$$(2) \qquad (1-\varepsilon) \int_{F_m(f)} f(\omega)\, d\omega + m \int_{E_m(f)} d\omega = 1\,.$$

PROOF. Since $\int_{-\pi}^{\pi} f(\omega) = 1$, equation (2) can be rewritten as $P(f, m) = \varepsilon$. However $P(f, m)$ is a monotone nondecreasing continuous function of m such that it tends to 0 as $m \to 0$ and increases to infinity as $m \to \infty$; consequently, this shows that there exists a unique solution of equation (2). □

Henceforth, the notation $m(f)$ refers to the positive number satisfying (2) and write $E_{m(f)}(f)$ and $F_{m(f)}(f)$ respectively as $E_{m(f)}$ and $F_{m(f)}$. As the next step, define f_m as follows: $f_m(\omega) = (1-\varepsilon)f(\omega)$ for $\omega \in F_{m(f)}$; $f_m(\omega) = m(f)$ for $\omega \in E_{m(f)}$.

It is obvious in view of the lemma above that the function f_m is a probability density whose distribution belongs to $\mathscr{F}_1(F)$. Let ϕ_m be the optimal linear one step ahead predictor when the spectral density of the process is given as f_m; namely, $\phi_m \in \mathscr{L}(H_m)$ and it holds that

$$\int_{-\pi}^{\pi} |1 - \phi_m(\omega)|^2 f_m(\omega)\, d\omega = \min_{\phi \in \mathscr{L}(H_m)} \int_{-\pi}^{\pi} |1 - \phi(\omega)|^2 f_m(\omega)\, d\omega\,,$$

where H_m is the distribution function of f_m. This function ϕ_m turns out to belong to $\mathscr{L}_0(F)$ and to be a minimax predictor; that is,

THEOREM 1. *Suppose the spectral distribution F has a density f. Among the class $\mathscr{L}_1$ of linear predictors, there exists a predictor which minimizes the maximal error of one-step ahead prediction for the set $\mathscr{F}_1(F)$ of spectral distributions. That predictor ϕ_m is given as the optimal linear predictor for the spectral density f_m such that $f_m(\omega) = (1-\varepsilon)f(\omega)$ for $\varepsilon \in F_{m(f)}$ and $f_m(\omega) = m(f)$ for $\varepsilon \in E_{m(f)}$. Its maximal prediction error* $\max_{H \in \mathscr{F}_1(F)} V(\phi_m, H)$ *is equal to* $2\pi \exp\{1/2\pi \int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\}$.

The theorem consists of two separate propositions: one is that $\max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) = \min_{\phi \in \mathscr{L}_1(F)} \max_{H \in \mathscr{F}_1(F)} V(\phi, H)$ (Proposition 1); the other is that $\phi_m \in \mathscr{L}_1(F)$ (Proposition 2). This section demonstrates only the first proposition, leaving the proof of Proposition 2 as well as an extension of Theorem 1 to the next section.

Proposition 1 is proved by steps through Lemmas 2 to 5 in the following. For any $H \in \mathscr{F}_1(F)$, let h be the density of H; namely, $h = (1 - \varepsilon)f + \varepsilon g$ for a certain density g. Now the next lemma is self-evident in view of the usual theory of prediction.

LEMMA 2. $\min_{\phi \in \mathscr{L}_1(F)} V(\phi, H) \geqq 2\pi \exp\{1/2\pi \int_{-\pi}^{\pi} \log h(\omega)\, d\omega\}$.

LEMMA 3. $\int_{-\pi}^{\pi} \log h(\omega)\, d\omega \leqq \int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega$.

PROOF. Let k_1 and k_2 be functions defined on $(-\pi, \pi]$ as: $k_1(\omega) = h(\omega) - f_m(\omega)$ and $k_2(\omega) = 0$ for $\omega \in F_{m(f)}$; $k_1(\omega) = 0$ and $k_2(\omega) = f_m(\omega) - h(\omega)$ for $\omega \in E_{m(f)}$. Then k_1 is nonnegative, and it holds that

$$\int_{-\pi}^{\pi} (k_1(\omega) - k_2(\omega))\, d\omega = 0 \quad \text{and} \quad \int_{-\pi}^{\pi} k_1(\omega)k_2(\omega)\, d\omega = 0\,. \tag{3}$$

Now,

$$\begin{aligned} &\int_{-\pi}^{\pi} \log h(\omega)\, d\omega - \int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega \\ &\qquad = \int_{F_{m(f)}} \log\left(1 + \frac{h(\omega) - f_m(\omega)}{f_m(\omega)}\right) d\omega \\ &\qquad\quad + \int_{E_{m(f)}} \log\left(1 - \frac{f_m(\omega) - h(\omega)}{f_m(\omega)}\right) d\omega \\ &\qquad \leqq \int_{-\pi}^{\pi} \log\left(1 + \frac{k_1(\omega)}{m(f)}\right)\left(1 - \frac{k_2(\omega)}{m(f)}\right) d\omega\,. \end{aligned} \tag{4}$$

On the other hand, in view of the concavity of the logarithm,

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(1 + \frac{k_1(\omega)}{m(f)}\right)\left(1 - \frac{k_2(\omega)}{m(f)}\right) d\omega \\ &\qquad \leqq \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1 + \frac{k_1(\omega) - k_2(\omega)}{m(f)} - \frac{k_1(\omega)k_2(\omega)}{m(f)^2}\right) d\omega\,. \end{aligned} \tag{5}$$

In view of the relations in (3) it follows from (4) and (5) that

$$\int_{-\pi}^{\pi} \log h(\omega)\, d\omega \leqq \int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\,. \qquad \square$$

It can be concluded from Lemmas 2 and 3 that

$$\max_{H \in \mathscr{F}_1(F)} \min_{\phi \in \mathscr{L}_1(F)} V(\phi, H) \geqq 2\pi \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\right\}. \tag{6}$$

The next lemma is a straightforward consequence of the definition of maximum and minimum.

LEMMA 4. $\max_{H \in \mathscr{F}_1(F)} \min_{\phi \in \mathscr{L}_1(F)} V(\phi, H) \leqq \min_{\phi \in \mathscr{L}_1(F)} \max_{H \in \mathscr{F}_1(F)} V(\phi, H)$.

Let $\phi_m(\omega)$ be the optimal linear one-step ahead predictor corresponding to the spectral density $f_m(\omega)$; that is, $\phi_m(\omega) \in \mathscr{L}(H_m)$ $\int_{-\pi}^{\pi} |1 - \phi_m(\omega)|^2 f_m(\omega)\, d\omega = \min_{\phi \in \mathscr{L}(H_m)} \int_{-\pi}^{\pi} |1 - \phi(\omega)|^2 f_m(\omega)\, d\omega$, where H_m is the spectral distribution induced by the density $f_m(\omega)$.

LEMMA 5. $\max_{H\in\mathscr{F}_1(F)} V(\phi_m, H) = 2\pi \exp\{1/2\pi \int_{-\pi}^{\pi} \log f(\omega)\, d\omega\}$.

PROOF. Since, in view of the theory of linear predictions,

$$(7)\qquad |1-\phi_m(\omega)|^2 = \frac{2\pi}{f_m(\omega)} \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\right\} \quad \text{a.e.},$$

(see, e.g., Grenander and Rosenblatt (1957) pages 68–69; also the next section gives a brief account of the construction of the optimal linear predictor), it follows that, for $H \in \mathscr{F}_1(F)$,

$$(8)\qquad V(\phi_m, H) = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\right\} \cdot \int_{-\pi}^{\pi} \frac{1}{f_m(\omega)}\, dH(\omega),$$

whereas

$$(9)\qquad \int_{-\pi}^{\pi} \frac{1}{f_m(\omega)}\, dH(\omega) = \int_{-\pi}^{\pi} \frac{1-\varepsilon}{f_m(\omega)} f(\omega)\, d\omega + \varepsilon \int_{-\pi}^{\pi} \frac{1}{f_m(\omega)} g(\omega)\, d\omega$$

for some density $g(\omega)$. From (8) and (9), it follows that $V(\phi_m, H)$ attains its maximum for a spectral distribution H' such that, if the density H' is $(1-\varepsilon)f + \varepsilon g'$, g' is the uniform distribution on the subset $E_{m(f)}$. For such H', it is easy to derive that $\int_{-\pi}^{\pi} (1/f_m(\omega))\, dH'(\omega) = 1$; hence,

$$V(\phi_m, H') = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\right\}. \qquad \square$$

Finally, the preceding lemmas lead to the following conclusion.

PROPOSITION 1.

$$\begin{aligned}\max_{H\in\mathscr{F}_1(F)} V(\phi_m, H) &= \min_{\phi\in\mathscr{L}_1(F)} \max_{H\in\mathscr{F}_1(F)} V(\phi, H)\\ &= 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_m(\omega)\, d\omega\right\}.\end{aligned}$$

2. The existence of optimal robust predictor. The purpose of this section is to prove Proposition 2 and also to extend Theorem 1 to the case where the the spectral distribution F has point masses. Since, in the ordinary framework of linear predictions, the optimal predictor ϕ for a spectrum G is constructed as the projection of the constant function 1 on the closed subspace $\mathscr{L}(G)$, it is not necessary to prove anew that $\phi \in \mathscr{L}(G)$. However, this is not the case for the present situation; in other words, ϕ is not usually a projection with respect to other spectra. Therefore, the fact that $\phi_m \in \mathscr{L}_1(F)$ must be established independently from its construction. Lemma 9 below shows that the function $l(\lambda - \lambda_0)$ which is equal to 1 at $\lambda = \lambda_0$ and zero elsewhere belongs to $\mathscr{L}_1(F)$ even if F has a jump at $\lambda = \lambda_0$: that lemma enables the extension of Theorem 1 in a straightforward manner.

It is convenient to review, at first, the construction of ϕ_m mainly for the purpose of introducing notation necessary for further arguments. Let a_n, $n = 0, \pm 1, \pm 2, \cdots$, be the Fourier coefficients of the integrable function $\log f_m(\omega)$. Using these a_n, define a function $g(z)$ in the open unit disc of the complex plane as $g(z) = a_0 + 2\sum_{n=1}^{\infty} a_n z^n$ ($|z| < 1$); moreover, let c_j be complex coefficients such that $(2\pi)^{\frac{1}{2}} \exp\{g(z)\} = \sum_{j=0}^{\infty} \bar{c}_j z^j$ ($\bar{c}_j$ is the conjugate of c_j). This expansion makes sense for the domain ($|z| < 1\}$, since $g(z)$ is analytic there. Now let $c(z)$ be a function defined in the open unit disc as: $c(z) = \sum_{j=0}^{\infty} c_j z^j$. Then it is known that it has the boundary value $c(e^{-i\lambda})$ almost everywhere such that $c(e^{-i\lambda}) = \sum_{j=0}^{\infty} c_j e^{-i\lambda j} = \lim_{r\to 1-} c(re^{-i\lambda})$, where $\sum_{j=0}^{\infty} |c_j|^2 < \infty$. Furthermore, it is true that $f_m(\omega) = (1/2\pi)|c(e^{-i\omega})|^2$ almost everywhere with respect to the Lebesgue measure. Finally, the optimal predictor ϕ_m is given by $\phi_m(\omega) = \sum_{j=1}^{\infty} c_j e^{-i\lambda j}/c(e^{-i\lambda})$. Let $\xi(z) = \sum_{j=1}^{\infty} c_j z^j/c(z)$. Then since $1/c(z)$ is analytic in the open unit disc, $\xi(z)$ has the expansion: $\xi(z) = \sum_{j=1}^{\infty} b_j z^j$, $|z| < 1$. Obviously, as $r \to 1-$, $\xi(re^{-i\omega})$ converges to $\xi(e^{-i\omega}) = \phi_m(\omega)$ a.e. For fixed r, write $\xi_r(\omega) = \xi(re^{-i\omega})$.

The proof of the proposition that $\phi_m \in \mathscr{L}_1(F)$ proceeds as follows: Lemma 6 states the fact that $\xi_r \in \mathscr{L}_1(F)$, $0 < r < 1$. Proposition 2 shows that ξ_r converges to $\phi_m(\omega)$, as $r \to 1-$, in the mean-square with respect to every spectral distribution belonging to $\mathscr{F}_1(F)$. Lemma 7 establishes a preliminary result which is useful for Proposition 2.

LEMMA 6. $\xi_r \in \mathscr{L}_1(F)$, $0 < r < 1$.

PROOF. For any $H \in \mathscr{F}_1(F)$, let $\gamma(k)$, $k = 0, \pm 1, \pm 2, \cdots$, be the covariance function of the spectrum H (namely, $\gamma(k) = \int_{-\pi}^{\pi} e^{ik\omega}\, dH(\omega)$). Since $|\gamma(k)| \leqq \gamma(0)$ for every integer k and $|b_k| < M$ for some positive M,

$$\int_{-\pi}^{\pi} |\textstyle\sum_{j=n}^{\infty} b_j r^j e^{-ij\omega}|^2\, dH(\omega) \leqq M^2\gamma(0) r^{2n}/(1-r)^2\,. \tag{10}$$

Since $\sum_{j=1}^{n-1} b_j r^j e^{-ij\omega} \in \mathscr{L}(H)$ and right-hand side of (10) converges to 0 as $n \to \infty$, it follows that $\xi_r \in \mathscr{L}(H)$. □

Functions defined on the interval $(-\pi, \pi]$ are naturally extended to periodic functions on the real line by means of the convention: $f(\omega + 2\pi) = f(\omega)$. This convention will be used where necessary in the proceeding discussions without notice. Write the shifts of a periodic function $k(\omega)$ as $k_\lambda(\omega) = k(\omega - \lambda)$; thus, for example, $\phi_{m,\lambda}(\omega) = \phi_m(\omega - \lambda)$. Furthermore, write the L^2-norm of $k(\omega)$ with respect to a density $h(\omega)$ as $\|k\|_h$; namely, $\|k\|_h^2 = \int_{-\pi}^{\pi} |k(\omega)|^2 h(\omega)\, d\omega$.

LEMMA 7. *For any* $\varepsilon > 0$, *there exists a positive* δ_0 *such that* $\sup\{\|\phi_m - \phi_{m,\lambda}\|_h^2 : |\lambda| < \delta_0\} < \varepsilon$.

PROOF. ϕ_m is essentially bounded, since

$$|1 - \phi_m(\omega)|^2 = 1|c(e^{-i\omega}) \leqq 1/(2\pi f_m(\omega))\,, \quad \text{a.e.}$$

For this essentially bounded ϕ_m, there exists a continuous periodic function η such that $\|\phi_m - \eta\|_h^2 < \varepsilon/(4\pi + 6)$ (cf. e.g., Hewitt and Stromberg (1969), page 197). Because of the uniform continuity of η, there exists a positive δ_1 such that $|\eta(\omega) - \eta(\omega - \lambda)| < \varepsilon/(4\pi + 6)$ for $|\lambda| < \delta_1$. Then, it holds that

$$\begin{aligned} \|\phi_m - \phi_{m,\lambda}\|_h^2 &\leqq 2\{\|\phi_m - \eta\|_h^2 + \|\eta_\lambda - \eta\|_h^2 + \|\eta_\lambda - \phi_{m,\lambda}\|_h^2\} \\ &\leqq (4\pi + 2)\varepsilon/(4\pi + 6) + 2\|\eta_\lambda - \phi_{m,\lambda}\|_h^2\,. \end{aligned} \tag{11}$$

On the other hand,

$$(12)\qquad \big|\|\phi_m - \eta\|_h^2 - \|\phi_{m,\lambda} - \eta_\lambda\|_h^2\big| < \textstyle\int_{-\pi}^{\pi} |\phi_m(\omega) - \eta(\omega)|^2 |h(\omega) - h(\omega + \lambda)|\, d\omega \le N \textstyle\int_{-\pi}^{\pi} |h(\omega) - h(\omega + \lambda)|\, d\omega\,,$$

where N is a certain positive constant. Now in view of the fact that the shift of an integrable function is continuous with respect to the L^1-norm for the Lebesgue measure, the term in the right-hand side of the second inequality of (12) can be made smaller than $\varepsilon/(4\pi + 6)$ for $|\lambda| < \delta_2$. Set $\delta_0 = \min(\delta_1, \delta_2)$; then for $|\lambda| < \delta_0$

$$\|\phi_m - \phi_{m,\lambda}\|_h^2 < (4\pi + 2)\varepsilon/(4\pi + 6) + 4\varepsilon/(4\pi + 6) = \varepsilon\,. \qquad \square$$

PROPOSITION 2. $\phi_m \in \mathscr{L}_1(F)$.

PROOF. By virtue of Lemma 6, it suffices to show that the mean-square $\int_{-\pi}^{\pi} |\xi_r(\omega) - \phi_m(\omega)|^2 h(\omega)\, d\omega$ tends to 0 as $r \to 1-$ for any density h, since ϕ_m is square-integrable with respect to h. Write the Poisson kernel as $P_r(\lambda)$; namely $P_r(\lambda) = (1 - r^2)/(1 - 2r\cos\omega + r^2)$ (for the Poisson kernel see Hoffman (1962)). Then by the usual theory of that kernel, it holds that

$$(13)\qquad \|\phi_m - \xi_r\|_h \le \textstyle\int_{-\pi}^{\pi} \|\phi_m - \phi_{m,\lambda}\|_h P_r(\lambda)\, d\lambda\,.$$

On the other hand, for $\delta > 0$,

$$(14)\qquad \textstyle\int_{-\pi}^{\pi} \|\phi_m - \phi_{m,\lambda}\|_h P_r(\lambda)\, d\lambda \le \sup_{|\lambda|\le\delta} \|\phi_m - \phi_{m,\lambda}\|_h + 2K \sup_{|\lambda|>\delta} P_r(\lambda)\,,$$

where K is the essential supremum of the function ϕ_m. Then by Lemma 7, the first term in the right-hand side of the inequality (14) can be made arbitrarily small by choosing δ sufficiently small, and at the same time the second term is made arbitrarily small by choosing r sufficiently close to 1. This proves that $\|\phi_m - \xi_r\|_h \to 0$ as $r \to 1-$. $\square$

REMARK. Since ϕ_m is defined only as an a.e. limit, it is generally not valid that $\phi_m \in \mathscr{L}_0(F)$. If ϕ_m is continuous and $\phi_m(-\pi+) = \phi_m(\pi)$, the convergence $\xi_r \to \phi_m$ is uniform. But it is not certain whether the continuity of f_m implies that of ϕ_m.

In the rest of this section, suppose that F has jumps at countable points λ_i, $i = 1, 2, \cdots$, with the corresponding saltuses $\Delta F(\lambda_i)$; namely $F(\lambda) = \int_{-\pi}^{\lambda} f(\omega)\, d\omega + \sum_{\lambda_i \le \lambda} \Delta F(\lambda_i)$. In this case, however, $\int_{-\pi}^{\pi} f(\omega)\, d\omega = 1 - \beta < 1$, where $\beta = \sum_{i=1}^{\infty} \Delta F(\lambda_i)$. Denote by G the set $\{\lambda_i : i = 1, 2, \cdots\}$; let G' be the complement of G in $(-\pi, \pi]$. As in Lemma 1, there exists a unique m for which

$$(15)\qquad (1 - \varepsilon) \textstyle\int_{F_m(f)} f(\omega)\, d\omega + m \textstyle\int_{E_m(f)} d\omega = 1 - \beta + \beta\varepsilon\,.$$

Denote $n(f)$ the unique m and define f_n as this: $f_n(\omega) = (1 - \varepsilon) f(\omega)$ for $\omega \in F_{n(f)}$ and $f_n(\omega) = n(f)$ for $\omega \in E_{n(f)}$. Furthermore, define ϕ_n as follows: ϕ_n is equal to the optimal linear predictor corresponding to the density f_n for $\omega \in G'$ and is identically equal to 1 on G. Then it is straightforward to see that

$$(16)\qquad \min_{\phi\in\mathscr{L}_1(F)}\max_{H\in\mathscr{F}_1(F)} V(\phi, H) = \max_{H\in\mathscr{F}_1(F)} V(\phi, H) = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_n(\omega)\,d\omega\right\}.$$

In order to show that $\phi_n \in \mathscr{L}_1(F)$, it suffices to establish the next lemma.

LEMMA 8. *Let $\delta(\lambda)$ be a function defined on $(-\pi, \pi]$ such that $\delta(0) = 1$ and $\delta(\omega) = 0$ for $\omega \neq 0$. Then, for each i, $\delta(\lambda - \lambda_i)$ belongs to $\mathscr{L}_1(F)$.*

PROOF. Let m be a positive integer and set $\delta_m(\lambda-\lambda_i) = \sum_{t=-(2m+1)}^{-1} e^{it(\lambda-\lambda_i)}/(2m+1)$. It is obvious that δ_m thus defined belongs to $\mathscr{L}(H)$ for any $H \in \mathscr{F}_1(F)$. Moreover it holds that

$$(17)\qquad \int_{-\pi}^{\pi}|\delta(\lambda-\lambda_i) - \delta_m(\lambda-\lambda_i)|^2\,dH(\lambda) = \int_{(-\pi,\pi]\cap\{\lambda_i\}'}\left|\sin\frac{(m-1)(\lambda-\lambda_i)}{2}\Big/\left\{(2m+1)\sin\frac{(\lambda-\lambda_i)}{2}\right\}\right|^2 dH(\lambda),$$

where $(-\pi, \pi] \cap \{\lambda_i\}'$ is the set obtained by the point λ_i eliminated from $(-\pi, \pi]$. Since the integrand in the right-hand side of (17) converges point-wise to 0 as $m \to \infty$, it follows from the Lebesgue bounded convergence theorem that the mean square in (17) converges to 0. □

THEOREM 2. *Suppose that the spectral distribution F has jumps on the countable set $G = \{\lambda_i\}$ and has the representation $F(\lambda) = \int_{-\pi}^{\lambda} f(\omega)\,d\omega + \sum_{\lambda_i\leq\lambda}\Delta F(\lambda_i)$. The function ϕ_n which is equal to 1 on G and equal elsewhere to the optimal linear predictor for the spectral density f_n belongs to $\mathscr{L}_1(F)$ and is a minimax predictor. Its maximal prediction error is given by $2\pi\exp\{1/2\pi \int_{-\pi}^{\pi} f_n(\omega)\,d\omega\}$.*

3. Robust prediction in the presence of the class $\mathscr{F}_0(F)$. The minimax theory developed so far does not, in general, apply to $\mathscr{F}_0(F)$. This section aims at constructing a predictor which is robust and approximates ϕ_m in the sense that its maximal prediction error for $\mathscr{F}_0(F)$ is sufficiently close to $\min_{\phi\in\mathscr{L}_1(F)}\max_{H\in\mathscr{F}_1(F)} V(\phi, H)$. The construction will proceed as follows: At first, the density f_m is approximated by a continuous function; Proposition 3 states that there exists a continuous density k_m such that ψ_m, the optimal linear predictor for k_m, satisfies the inequality $\max_{H\in\mathscr{F}_1(F)} V(\psi_m, H) < \min_{\phi\in\mathscr{L}_1(F)}\max_{H\in\mathscr{F}_1(H)} V(\phi, H) + \delta$, for any given δ (>0). Secondly, the density k_m is approximated by a two-times continuously differentiable Stekloff function the optimal linear predictor for which is shown to belonging to $\mathscr{L}_0(F)$ (Lemma 12). Finally, Theorem 3 shows that, for any $\delta > 0$, there exists a linear predictor $\xi \in \mathscr{L}_0(F)$ such that $\max_{H\in\mathscr{F}_0(F)} V(\xi, H) < \min_{H\in\mathscr{L}_1(F)}\max_{\phi\in\mathscr{F}_1(F)} V(\phi, H) + \delta$. The term "continuous function" below refers exclusively to a continuous function defined on $(-\pi, \pi]$ whose natural extension to a periodic function is again continuous. The spectrum F is assumed to possess a density f; Theorem 3 below would be easily extended so as to apply to general F by virtue of Theorem 2, though the paper does not try it.

The following lemmas (9 through 11) establish preliminary results for Proposition 3. The next one is a well-known fact.

LEMMA 9. *For any* $\delta > 0$, *there exists a continuous density* g *such that* $\int_{-\pi}^{\pi} |f(\omega) - g(\omega)| \, d\omega < \delta$.

Denote by $\mathscr{G}_\delta$ the set of densities such that, for $g \in \mathscr{G}_\delta$, $\int_{-\pi}^{\pi} |g(\omega) - f(\omega)| \, d\omega < \delta$. Recall the definition of $m(f)$, f_m and $P(f, m)$ given in the first paragraph of Section 1; then,

LEMMA 10. *For any* $\delta > 0$, *there exists* $\lambda (>0)$ *such that for any* $g \in \mathscr{G}_\lambda$ *the inequalities* $|1/m(g) - 1/m(f)| < \delta$ *and* $|m(g) - m(f)| < \delta$ *hold simultaneously.*

PROOF. The lemma follows from the fact that $P(f, m)$ is a continuous monotone function of m and from the inequality:

$$|P(f, m(g)) - P(f, m(f))| < \int_{-\pi}^{\pi} |f(\omega) - g(\omega)| \, d\omega \, . \qquad \square$$

LEMMA 11. *For any* $\delta > 0$, *there exists* $\lambda (>0)$ *such that for any* $g \in \mathscr{G}_\lambda$ *it holds that* $1/m(g) \int_{-\pi}^{\pi} |g_m(\omega) - f_m(\omega)| \, d\omega < \delta$, *where* g_m *is defined in the same way as* f_m.

PROOF. In view of the relation:

$$\frac{1}{m(g)} \int_{-\pi}^{\pi} |g_m(\omega) - f_m(\omega)| \, d\omega \leqq \frac{1-\varepsilon}{m(g)} \int_{-\pi}^{\pi} |g(\omega) - f(\omega)| \, d\omega + \frac{|m(g) - m(f)|}{m(g)} \, ,$$

the lemma is a straightforward consequence of Lemmas 9 and 10. $\square$

PROPOSITION 3. *Suppose that the spectral distribution* F *has a density* f; *then for any* $\delta > 0$ *there exists a continuous density* k_m *such that, to denote by* ϕ_m *the optimal linear predictor for* k_m, *it holds that*

$$\text{(18)} \qquad \max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) < \min_{\phi \in \mathscr{L}_1(F)} \max_{H \in \mathscr{F}_1(F)} V(\phi, H) + \delta \, .$$

PROOF. Choose η to be a positive number satisfying the inequality:

$$\text{(19)} \qquad (1+\eta)\{\max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) + 2\eta\} < \max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) + \delta \, .$$

Then, in view of Lemmas 9 through 11, it is seen that there exists a continuous density k for which the following four inequalities hold simultaneously (where k_m is defined in the same way as f_m):

$$\text{(20)} \qquad |1/m(k) - 1/m(f)| < \eta \, ;$$

$$\text{(21)} \qquad |m(k) - m(f)| < \eta \, ;$$

$$\text{(22)} \qquad \frac{1}{m(k)} \int_{-\pi}^{\pi} |k_m(\omega) - f_m(\omega)| \, d\omega < \eta \, ;$$

$$\text{(23)} \qquad \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\omega) - k(\omega)| \, d\omega < \eta \, .$$

It evidently follows from (22) and (23) that $1/2\pi \int_{-\pi}^{\pi} |f_m(\omega) - k_m(\omega)| \, d\omega < \eta$. Since ϕ_m is the optimal linear predictor for k_m,

$$\text{(24)} \qquad \max_{H \in \mathscr{F}_1(F)} \int_{-\pi}^{\pi} |1 - \phi_m|(\omega)^2 \, dH(\omega)$$
$$= 2\pi \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log k_m(\omega) \, d\omega\right\} \left\{(1-\varepsilon) \int_{-\pi}^{\pi} \frac{f(\omega)}{k_m(\omega)} \, d\omega + \frac{\varepsilon}{m(k)}\right\} .$$

On the other hand,

$$(25)\qquad \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log k_m(\omega)\,d\omega\right\}\Big/\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_m(\omega)\,d\omega\right\}$$
$$\leqq 1+\frac{1}{2\pi}\int_{-\pi}^{\pi}|f_m(\omega)-k_m(\omega)|\,d\omega \leqq 1+\eta\,.$$

Furthermore, in view of (22),

$$(26)\qquad \left|\int_{-\pi}^{\pi} f(\omega)/k_m(\omega)\,d\omega - \int_{-\pi}^{\pi} f(\omega)/f_m(\omega)\,d\omega\right|$$
$$\leqq \int_{-\pi}^{\pi}|k_m(\omega)-f_m(\omega)|\,d\omega/m(k) < \eta\,.$$

By means of (24), (25) and (26), it holds that

$$\max_{H\in\mathscr{F}_1(F)}\int_{-\pi}^{\pi}|1-\phi_m(\omega)|^2\,dH(\omega)$$
$$\leqq 2\pi(1+\eta)\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_m(\omega)\,d\omega\right\}$$
$$\times\left\{(1-\varepsilon)\int_{-\pi}^{\pi} f(\omega)/f_m(\omega)\,d\omega+\varepsilon/m(f)+2\eta\right\}$$
$$\leqq (1+\eta)\{\max_{H\in\mathscr{F}_1(F)} V(\phi_m, H)+2\eta\}$$
$$\leqq \min_{\phi\in\mathscr{L}_1(F)}\max_{H\in\mathscr{F}_1(F)} V(\phi, H)+\delta\,. \qquad \square$$

For an integrable function g, the Stekloff functions g_h and g_{hh} $(h>0)$ are defined respectively as:

$$(27)\qquad g_h(\omega)=\frac{1}{2h}\int_{-h}^{h} g(\omega+\lambda)\,d\lambda\,, \qquad \text{and} \qquad g_{hh}(\omega)=\frac{1}{2h}\int_{-h}^{h} g_h(\omega+\lambda)\,d\lambda\,.$$

Then the next lemma is a straightforward consequence of Achieser (1956), pages 174–175.

LEMMA 12. *If a function g is continuous, g_{hh} is twice continuously differentiable; moreover, there exists δ_0 such that, for $h<\delta_0$,* $\sup|g(\omega)-g_{hh}(\omega)|<\varepsilon$ *for any* $\varepsilon>0$.

Let $k_{m,l}$ and $k_{m,ll}$ $(l>0)$ be Stekloff functions derived from spectral density k_m. Denote by ξ_l the optimal linear predictor when the spectral density is given as $k_{m,ll}$.

LEMMA 13. *$\xi_l\in\mathscr{L}_0(F)$ for any $l>0$.*

PROOF. Since $k_{m,ll}$ is twice continuously differentiable and bounded away from 0, $\log k_{m,ll}$ is also two-times continuously differentiable. Let a_n^* be the nth Fourier coefficient of $\frac{1}{2}\log k_{m,ll}$; then,

$$(28)\qquad \tfrac{1}{2}\log k_{m,ll}(\omega)=\sum_{-\infty}^{\infty} a_n^*\,e^{in\omega}\,,$$

where the sum converges uniformly and $\sum_{-\infty}^{\infty}|a_n^*|<\infty$. As in Section 2, let $g^*(z)=a_0^*+2\sum_{n=1}^{\infty}a_n^*z^n$ and define $c^*(z)$ by $c^*(z)=\sum_{j=0}^{\infty}c_j^*z^j$, where $\sum_{j=0}^{\infty}\bar{c}_j^*z^j=(2\pi)^{\frac{1}{2}}\exp\{g(z)\}$. Then it is evident that $c^*(z)$ thus defined is analytic in the open unit disc and continuous on the unit circle. The optimal linear predictor ξ_l for $k_{m,ll}$ is defined as: $\xi_l(\omega)=\sum_{j=1}^{\infty}c_j^*e^{-ij\omega}/c^*(e^{-i\omega})$, where $c^*(e^{-i\omega})=\sum_{j=0}^{\infty}c_j^*e^{-ij\omega}$. Let $\lambda_r(\omega)=\sum_{j=1}^{\infty}c_j^*(re^{-i\omega})^j/c^*(re^{-i\omega})$ for any fixed r $(0<r<1)$.

Then, since ξ_l is a continuous function of ω, $\xi_l(\omega)$ is the boundary value of $\lambda_r(\omega)$ $r \to 1-$ and the convergence is uniform. (See Hoffman (1962), page 33.) Now it is evident that $\int_{-\pi}^{\pi} |\xi_l(\omega) - \lambda_r(\omega)|^2 \, dH(\omega)$ converges to 0 as $r \to 1-$ for any $H \in \mathscr{F}_0(F)$. Since $\lambda_r(\omega) \in \mathscr{L}(H)$ and $\int_{-\pi}^{\pi} |\xi_l(\omega)|^2 \, dH(\omega) < \infty$, the proof is complete. □

THEOREM 3. *For any* $\delta > 0$, *there exists* l_0 (>0) *such that*

$$\max_{H \in \mathscr{F}_0(F)} V(\xi_l, H) < \max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) + \delta \qquad \textit{for all positive} \quad l < l_0 .$$

PROOF. In view of the relations

$$(29) \qquad |1 - \phi_m(\omega)|^2 = 2\pi \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log k_m(\omega) \, d\omega\right\} \Big/ k_m(\omega) , \quad \text{a.e.} ,$$

$$|1 - \xi_l(\omega)|^2 = 2\pi \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log k_{m,ll}(\omega) \, d\omega\right\} \Big/ k_{m,ll}(\omega) ,$$

it follows from Lemma 12 that for sufficiently small l $|1 - \xi_l|^2 < |1 - \phi_m|^2 + \delta/2$, a.e.; $\sup |1 - \xi_l|^2 <$ ess. $\sup |1 - \phi_m|^2 + \delta/2$. Then it holds that

$$\begin{aligned} \max_{H \in \mathscr{F}_0(F)} V(\xi_l, H) &= (1 - \varepsilon) \int_{-\pi}^{\pi} |1 - \xi_l(\omega)|^2 f(\omega) \, d\omega + \varepsilon \sup |1 - \xi_l(\omega)|^2 \\ &< (1 - \varepsilon) \int_{-\pi}^{\pi} |1 - \phi_m(\omega)|^2 f(\omega) \, d\omega \\ &\quad + \varepsilon \text{ ess. sup} |1 - \phi_m(\omega)^2 + \frac{\delta}{2} \\ &= \max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) + \frac{\delta}{2} . \end{aligned}$$

Since, by Proposition 3, $\max_{H \in \mathscr{F}_1(F)} V(\psi_m, H) < \max_{H \in \mathscr{F}_1(F)} V(\phi_m, H) + \delta/2$ for sufficiently small l, the proof is complete. □

REFERENCES

ACHIESER, N. I. (1956). *Theory of Approximation.* Frederick Unger, New York.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

HEWITT, E. and STROMBERG, K. (1969). *Real and Abstract Analysis.* Springer-Verlag, New York.

HOFFMAN, K. (1962). *Banach Spaces of Analytic Functions.* Prentice-Hall, New Jersey.

HUBER, P. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

FACULTY OF ECONOMICS
TOHOKU UNIVERSITY
KAWAUCHI, SENDAI 980
JAPAN