

On empirical spectral analysis of stochastic processes

By ULF GRENANDER

1. Introduction

Two main cases of empirical determination of the spectrum of a stationary stochastic process will be treated in this paper.

To a given frequency, which all the time will be chosen as zero, there corresponds a discrete line in the spectrum. One wants to determine its mean amplitude when a realization of the process has been observed. Among all possible ways of estimating the mean amplitude there is one which gives maximum precision. If the rest of the spectrum is known *a priori* this estimate can be explicitly constructed. It is shown that this construction is related to the problem of prediction. For the construction of the asymptotically best estimate the knowledge of the rest of the spectrum is not necessary, at least not for purely non-deterministic processes.

The other case arises when the spectrum is absolutely continuous and one wants to estimate the spectral intensity or, what is equivalent, the covariance function. A class of estimates is given and studied in relation to periodogram analysis and to an estimate proposed by BARTLETT [2]. A principle of uncertainty is stated.

In passing, some simple properties of linear processes are studied.

2. Estimation and prediction

2.1. Let $x(t)$, $-\infty < t < \infty$, be a real valued process with finite variance. The mean value m and the covariance function $\varrho(s, t)$ are defined as

$$\begin{cases} m = E x(t) \\ \varrho(s, t) = E [x(s) - m][x(t) - m]. \end{cases}$$

The process is supposed to be continuous in the mean. Consider $x(t)$ in an interval $T = (a, b)$. Introduce the inner product

$$(y, z) = E y z$$

and the norm

$$\|y\| = \sqrt{(y, y)},$$

for any stochastic variables y and z with finite variance. The Hilbert space generated by $x(t)$ when t runs through T is denoted $L_2(T)$. When T is the whole real axis we shall write $L_2(T) = L_2$.

Suppose now that m is unknown, that a realization of $x(t)$ is observed in T and that we want to estimate m by a linear, unbiased estimate. In other words we want to form a linear combination

$$\begin{cases} \mu^* = \sum_1^n c_i x(t_i) \\ \sum_1^n c_i = 1, t_i \in T, \end{cases}$$

or a limit of such combinations. The c 's shall not depend on m . In the set M consisting of all such unbiased linear estimates we want to choose one with small standard deviation. In a recent publication [7] the author has shown the existence of one and but for equivalence only one estimate m^* for which

$$\begin{cases} m^* \in M \\ D m^* = \inf_{\mu^* \in M} D \mu^*. \end{cases}$$

Further m^* is the unique solution of the integral equation

$$E m^* [x(t) - m] \equiv c, t \in T, \tag{1}$$

where the constant c is equal to the minimum variance

$$D^2 m^* = c.$$

2.2. We shall now make a detour to the problem of prediction. Under the same conditions as above, except that we now put $m = 0$, we want to find an element in $L_2(T)$ which can be used as a prognosis to the time $b + h$, $h > 0$. The *best linear prognosis* $\hat{x}(b + h)$ is defined as the element in $L_2(T)$ which makes $\|\hat{x}(b + h) - x(b + h)\|$ a minimum. Then

$$\hat{x}(b + h) = P_{L_2(T)} x(b + h) \in L_2(T)$$

satisfying the equation

$$E [\hat{x}(b + h) - x(b + h)] x(t) \equiv 0, t \in T. \tag{2}$$

The "prediction" (or extrapolation) backwards in time is defined in the analogous way. When T is a finite interval KARHUNEN has given an explicit solution in terms of the eigen-elements of the kernel $r(s, t)$. When $a = -\infty$ and $x(t)$ is a stationary process WIENER [14] has constructed the solution.

Let us now return to the case when m is an unknown constant. Then it seems advantageous to introduce one more condition on the predictor, that

$$E x^*(b + h) = E x(b + h) = m$$

irrespective of the value of m . In order to define the best of these predictors consider the set M defined above. M is a convex set and then it can be

shown in the same way as in [7] p. 241 that there is a unique element $x^*(b+h) \in M$ such that

$$\|x^*(b+h) - x(b+h)\| = \inf_{z \in M} \|z - x(b+h)\|$$

and this element satisfies the equation

$$E[x^*(b+h) - x(b+h)]x(t) = k \text{ for all } t \in T. \tag{3}$$

The constant is

$$k = \|x^*(b+h)\|^2 - (x(b+h), x^*(b+h)).$$

$x^*(b+h)$ is called the *best unbiased prediction*. This concept seems to have been introduced first by DOLPH and WOODBURY [5]. These authors have constructed x^* in the case when the covariance function is a Green function of a second order differential operator.

2.3. We shall show that for stationary processes the best unbiased estimate of the mean value can be expressed in terms of the best unbiased prediction. To do this we shall first study an operator P^* associated with the prediction. P^* is defined in the following way. For any element $x \in L_2$ we define P^*x as the unique element in M which satisfies

$$\|x - P^*x\| = \inf_{z \in M} \|x - z\|.$$

For $x = x(b+h)$ we clearly get $P^*x(b+h) = x^*(b+h)$, i.e. the best unbiased prediction to the time $b+h$.

If P^* were a linear bounded operator we would have

$$\begin{cases} \|P^*x\| \leq c\|x\| \\ P^*(ax + by) = aP^*x + bP^*y, \end{cases}$$

but these relations do not hold in general for P^* . Some similar relations can however be easily obtained. We have

$$E[x - P^*x]x(t) \equiv \text{constant for } t \in T \tag{4}$$

and hence

$$x - P^*x \perp x(\alpha) - x(\beta) \text{ for all } \alpha, \beta \in T.$$

This implies that if x and y are two arbitrary elements of L_2 we have

$$x - y = P^*x - P^*y + x - P^*x - y + P^*y = P^*x - P^*y + u, u \perp P^*x - P^*y$$

so that

$$\|P^*x - P^*y\| \leq \|x - y\|. \tag{5}$$

From this we infer that the process $P^*x(t)$ is continuous in the mean and hence Cramér integrable (see [3]) which fact will be used below. — Suppose that $\sum_1^n c_\nu = 1$ and consider the element $P^* \left[\sum_1^n c_\nu x_\nu \right]$ where $x_\nu \in L_2$. Clearly

$$\sum_1^n c_\nu P^* x_\nu \in M$$

and

$$E \left\{ \sum_1^n c_\nu P^* x_\nu - P^* \left[\sum_1^n c_\nu x_\nu \right] \right\} x(t) = \sum_1^n c_\nu E P^* x_\nu \cdot x(t) - E \sum_1^n c_\nu x_\nu x(t) - c.$$

But $E[P^*x_\nu - x_\nu]x(t)$ is a constant so that the above expression is also a constant. As P^*x is uniquely determined by (4) we have

$$P^* \left[\sum_1^n c_\nu x_\nu \right] = \sum_1^n c_\nu P^* x_\nu. \tag{6}$$

If

$$I = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t) dt$$

we can approximate I by sums of the type $S_n = \frac{1}{n} \sum_1^n x(\tau_\nu^{(n)})$, and we get using (5) and (6)

$$\| P^* I - \frac{1}{n} \sum_1^n P^* x(\tau_\nu^{(n)}) \| = \| P^* I - P^* S_n \| \leq \| I - S_n \| \rightarrow 0$$

as n tends to infinity. Thus we have as $P^*x(t)$ is integrable

$$P^* \left[\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t) dt \right] = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} P^* x(t) dt. \tag{7}$$

Let $x(t)$ be a stationary process. Then the limit

$$\text{l.i.m.}_{A \rightarrow \infty} \frac{1}{2A} \int_{-A}^A x(t) dt$$

exists and is an unbiased estimate of m . It is easily seen that this estimate has minimum variance among all linear, unbiased estimates of m which can be formed using values of the process from the entire time axis. When we have observed a realization in the finite time interval T and want to obtain the best estimate of m , the above integral cannot be calculated as $x(t)$ is not known for $t \notin T$. It is then natural to extrapolate the realization and use this to determine the above integral. We shall now make this argument rigorous.

When A and B tend to infinity we have

$$\lim_{A, B \rightarrow \infty} \left\| \frac{1}{2A} \int_{-A}^A x(t) dt - \frac{1}{2B} \int_{-B}^B x(t) dt \right\| = 0.$$

But then according to (7) and (5)

$$\overline{\lim}_{A, B \rightarrow \infty} \left\| \frac{1}{2A} \int_{-A}^A x^*(t) dt - \frac{1}{2B} \int_{-B}^B x^*(t) dt \right\| \leq \overline{\lim}_{A, B \rightarrow \infty} \left\| \frac{1}{2A} \int_{-A}^A x(t) dt - \frac{1}{2B} \int_{-B}^B x(t) dt \right\| = 0$$

so that

$$z = \text{l.i.m.}_{A \rightarrow \infty} \frac{1}{2A} \int_{-A}^A x^*(t) dt$$

exists. Further $z \in M$ and as (if $s_1, s_2 \in T$) because of (4)

$$\begin{aligned} E z [x(s_1) - x(s_2)] &= \lim_{A \rightarrow \infty} \frac{1}{2A} \int_{-A}^A E x^*(t) [x(s_1) - x(s_2)] dt = \\ &= \lim_{A \rightarrow \infty} \frac{1}{2A} \int_{-A}^A E x(t) [x(s_1) - x(s_2)] dt = \\ &= \lim_{A \rightarrow \infty} \frac{1}{2A} \int_{-A-s_1}^{A-s_1} r(t) dt - \lim_{A \rightarrow \infty} \frac{1}{2A} \int_{-A-s_2}^{A-s_2} r(t) dt = 0, \end{aligned}$$

we have according to (1) the wanted representation

$$m^* = \text{l.i.m.}_{A \rightarrow \infty} \frac{1}{2A} \int_{-A}^A x^*(t) dt. \tag{8}$$

2.4. We shall now show that there is a similar relation between the best unbiased estimate m^* and the best prognosis $\hat{x}(t)$, at least under some regularity conditions. As the prediction operator is a projection and hence linear and bounded, $\hat{x}(t)$ is continuous in the mean. We shall consider stationary processes for which the integral

$$I = \int_{-\infty}^{\infty} \hat{x}(t) dt \tag{9}$$

exists. If $x(t)$ is purely non-deterministic (see [8] and [10]) it can be represented by means of a homogeneous orthogonal process $\xi(u)$ as

$$x(t) = \int_{-\infty}^t g(t-u) d\xi(u)$$

where

$$\int_0^{\infty} |g(u)|^2 du < \infty.$$

The convergence of

$$\int_0^{\infty} u^2 |g(u)|^2 du < \infty \tag{10}$$

is a sufficient condition for the existence of (9). To show this we consider the quantity

$$\|\hat{x}(b+h) - x(b+h)\|^2, \quad h > 0.$$

Introducing the prediction $\hat{x}_{\infty}(t) = P_{L_2(-\infty, b)} x(t)$, which corresponds to a realization observed in $(-\infty, b)$, we get

$$\|\hat{x}_{\infty}(b+h) - x(b+h)\| \leq \|\hat{x}(b+h) - x(b+h)\|.$$

But we have (see [11] p. 65)

$$\|\hat{x}_{\infty}(b+h) - x(b+h)\|^2 = \int_0^h |g(u)|^2 du$$

and thus

$$\|\hat{x}(b+h) - x(b+h)\|^2 \geq \int_0^h |g(u)|^2 du.$$

But

$$x(b+h) = \hat{x}(b+h) + [x(b+h) - \hat{x}(b+h)],$$

where the second term is orthogonal to the first one, from which follows

$$\int_0^{\infty} |g(u)|^2 du = \|x(b+h)\|^2 = \|\hat{x}(b+h)\|^2 + \|x(b+h) - \hat{x}(b+h)\|^2,$$

and

$$\|\hat{x}(b+h)\|^2 \leq \int_h^{\infty} |g(u)|^2 du.$$

Because of the symmetry of the covariance function we get a similar inequality for the extrapolation backwards in time. Now we get using (10)

$$\int_0^{\infty} \|\hat{x}(b+h)\|^2 dh \leq \int_0^{\infty} dh \int_h^{\infty} |g(u)|^2 du = \int_0^{\infty} u |g(u)|^2 du < \infty$$

so that according to Theorem 7 in [9] the integral (9) exists.

We have

$$r(t) = \int_{-\infty}^{\infty} g(u) \overline{g(u+t)} du; \quad g(u) = 0 \text{ for } u < 0;$$

where $g(u)$ is absolutely integrable because of

$$\left[\int_{-\infty}^{\infty} |g(u)| du \right]^2 \leq \int_{-\infty}^{\infty} \frac{du}{1+u^2} \int_{-\infty}^{\infty} (1+u^2) |g(u)|^2 du < \infty.$$

Hence

$$\int_A^B |r(t)| dt \leq \int_{-\infty}^{\infty} |g(u)| du \int_A^B |g(u+t)| dt \leq \left[\int_{-\infty}^{\infty} |g(u)| du \right]^2 < \infty$$

so that $r(t)$ is absolutely integrable. Using this and (2) we get

$$\begin{aligned} EI[x(s) - m] &= \int_{-\infty}^{\infty} E \hat{x}(t) [x(s) - m] dt = \int_{-\infty}^{\infty} E x(t) [x(s) - m] dt = \\ &= \int_{-\infty}^{\infty} r(t) dt \equiv \text{constant for } s \in T. \end{aligned}$$

Supposing that $\int_{-\infty}^{\infty} r(t) dt \neq 0$, i.e. that the spectral intensity corresponding to the frequency $\lambda = 0$ is positive, there is a constant γ , such that

$$\gamma \int_{-\infty}^{\infty} \hat{x}(t) dt = m^*.$$

Let us apply the above to a process with the covariance function $r(t) = e^{-\beta|t|}$. As is well known (see [14]) the best prediction has the form

$$\left. \begin{aligned} \hat{x}(b + \alpha) &= e^{-\beta\alpha} x(b), \quad \alpha > 0 \\ \hat{x}(t) &= x(t), \quad t \in (a, b) \\ \hat{x}(a - \alpha) &= e^{-\beta\alpha} x(a), \quad \alpha > 0 \end{aligned} \right\}.$$

We get

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \hat{x}(t) dt = \int_{-\infty}^a e^{-\beta(a-t)} x(a) dt + \int_a^b x(t) dt + \int_b^{\infty} e^{-\beta(t-b)} x(b) dt = \\ &= \frac{x(a)}{\beta} + \int_a^b x(t) dt + \frac{x(b)}{\beta}. \end{aligned}$$

If $\gamma = \frac{\beta}{2 + \beta(b-a)}$ the estimate γI is unbiased and the best unbiased estimate has the form

$$m^* = \frac{x(a) + x(b) + \beta \int_a^b x(t) dt}{2 + \beta(b-a)}$$

which was proved in another way in [7]. For such processes for which the best prediction depends only upon the last observed value of the realization and a finite number N of its derivatives, the best unbiased estimate is the sum of terms of the type

$$\int_a^b x(t) dt, x(a), x(b), x'(a), \dots, x^{(N)}(b).$$

2.5. It deserves mentioning that it is not essential to the above that T is an interval. T may be any bounded measurable set on the real axis and the results still hold. Consider e.g. the following problem. The process has again the covariance function $e^{-\beta|t|}$ and the mean value m shall be estimated. It is possible to observe the process in a given interval (A, B) but it is required that the set of all time points of observation shall consist of at most n disjoint interval situated in (A, B) and of total given length $L < B - A$. How are these intervals to be chosen in order to get maximum information about m ?

Suppose that the intervals are fixed. We can then construct the best unbiased estimate of m in the manner described above, because for the given covariance function the problem of extrapolation and interpolation is solved. The variance of this best estimate is a function of the location of the intervals of observation. One of the solutions (there are several) realizing the minimum of this function locates the intervals symmetrically in (A, B) .

3. Asymptotic properties of the equidistributed estimate

3.1. In 2.4. we showed how m could be expressed in terms of the best prediction

$$m^* = \gamma \left[\int_{-\infty}^a \hat{x}(t) dt + \int_a^b x(t) dt + \int_b^{\infty} \hat{x}(t) dt \right]. \tag{11}$$

The second integral can be used as an estimate of m because of its simple form and especially because it does not depend upon the covariance function. Let us form *the equidistributed estimate* (putting $T = (-T, T)$)

$$m_E^* = \frac{1}{2T} \int_{-T}^T x(t) dt. \tag{12}$$

If the process is purely non-deterministic the first and third terms in (11) will be small compared to the second one. More precisely, we have shown in [7] that if $x(t)$ is purely non-deterministic, so that there is at least one $\varphi(\lambda)$ whose Fourier transform vanishes for negative arguments and such that $F'(\lambda) = |\varphi(\lambda)|^2$, and if there is one $\varphi(\lambda)$ satisfying this condition and

then
$$\varphi(\lambda) = \varphi_0 + \varphi_1 \cdot \lambda(1 + o(1)), \quad \lambda \rightarrow 0, \quad \varphi_0 \neq 0, \tag{13}$$

$$\lim_{T \rightarrow \infty} \frac{D(m^*)}{D(m_E^*)} = 1, \tag{14}$$

i.e. the equidistributed estimate is asymptotically efficient (in this linear sense).

Can it happen¹ that the equidistributed estimate is not only asymptotically efficient but that already for some finite T

$$D m_E^* = D m^* ?$$

As the best unbiased estimate is unique we must have $m_E^* = m^*$ and because of (1)

$$E m_E^* [x(t) - m] \equiv c, \quad -T \leq t \leq T.$$

If the spectral distribution function of the process is denoted by $F(\lambda)$ we get

$$\int_{-T}^T r(s-t) ds = 2 \int_{-\infty}^{\infty} e^{-it\lambda} \frac{\sin T\lambda}{\lambda} dF(\lambda) \equiv c_1, \quad -T \leq t \leq T.$$

Derivation of this equation with respect to t may be performed under the sign of integration (see e.g. [4] p. 68) and gives

$$\int_{-\infty}^{\infty} e^{-it\lambda} \sin T\lambda dF(\lambda) = 0, \quad -T \leq t \leq T.$$

Putting $t = -T$ and $t = T$ and subtracting the two resulting equations we obtain

$$\int_{-\infty}^{\infty} \sin^2 T\lambda dF(\lambda) = 0.$$

Hence the spectrum must be empty except perhaps for the frequencies

$$\lambda = \frac{n\pi}{T}, \quad n = 0, \pm 1, \pm 2, \dots$$

To show that this condition is also sufficient we have only to consider the expression

$$E m_E^* [x(t) - m] = \frac{1}{2T} \int_{-T}^T r(s-t) ds.$$

¹ This question was raised by Prof. CRAMÉR at a seminar on stochastic processes in March 1950.

But

$$r(t) = \sum_{n=-\infty}^{\infty} \Delta F_n e^{i \frac{n\pi}{T} t}$$

where ΔF_n are the saltuses of $F(\lambda)$ in the points $\lambda = \frac{n\pi}{T}$. Hence

$$E m_E^* [x(t) - m] = \Delta F_0, \quad -T \leq t \leq T,$$

which implies $m_E^* = m^*$. One has to observe however that if m_E^* is efficient for some value of T it does not follow that it is efficient for all larger values of T .

3.2. The condition (13) can be replaced by another which is given directly in terms of the spectral intensity $F'(\lambda) = f(\lambda)$. Although it is presumably far from necessary it is sufficient for our purpose.

Suppose that we consider a purely non-deterministic process with an integral valued time parameter; further let $f(\lambda)$ be positive and have continuous derivatives of the first and second order. Then (13) is satisfied. To show this we use Wiener's construction (see [14]) of $\varphi(\lambda)$. The function $\log f(\lambda)$ is continuous together with its derivatives of the first and second order. We can develop in $(-\pi, \pi)$

$$\log f(\lambda) = \sum_{-\infty}^{\infty} \gamma_\nu e^{i\nu\lambda}$$

and we have

$$\sum_{-\infty}^{\infty} \nu^4 |\gamma_\nu|^2 < \infty.$$

Putting

$$h(\lambda) = \frac{\gamma_0}{2} + \sum_1^{\infty} \gamma_\nu e^{i\nu\lambda}$$

we have

$$Re h(\lambda) = \frac{1}{2} \log f(\lambda).$$

If we put

$$\varphi(\lambda) = e^{h(\lambda)}$$

which can be developed in non-negative powers of $e^{i\lambda}$ we have

$$|\varphi(\lambda)|^2 = e^{\log f(\lambda)} = f(\lambda).$$

As

$$\sum_1^{\infty} \nu |\gamma_\nu| \leq \sqrt{\sum_1^{\infty} \frac{1}{\nu^2} \sum_1^{\infty} \nu^4 |\gamma_\nu|^2} < \infty$$

the functions $h(\lambda)$ and $\varphi(\lambda)$ have a continuous first derivative as stated.

4. The linear processes

4.1. Any wide sense stationary process, which is continuous in the mean, can be written as the sum of two wide sense stationary processes

$$x(t) = x_1(t) + x_2(t)$$

where

$$\begin{cases} x_1(t) = \int_{-\infty}^t g(t-u) d\xi(u) \\ x_2(t) \perp x_1(s) \text{ for all } t \text{ and } s. \end{cases} \quad (15)$$

$\xi(u)$ is orthogonal and $\|d\xi(u)\|^2 = du, \int_0^\infty |g(u)|^2 du < \infty, x_1(t)$ is purely non-deterministic and $x_2(t)$ deterministic (see [8] and [10]). In applications $x_2(t)$ either vanishes or is of some simple form, e.g. a finite trigonometrical sum with random amplitudes.

It is possible to give (15) a simple physical interpretation. Let us consider $\Delta\xi(u)$ as a series of random shocks entering a linear filter characterized by the transient function $g(t)$. $x_1(t)$ is then the output of the filter. In many cases it is natural to assume that the increments of $\xi(u)$ over disjoint intervals are not only uncorrelated but also independent. Using the terminology of Bartlett $x_1(t)$ is then called a *linear process*.

In this paper we shall from now on deal with the case of an integral valued parameter. With some obvious modifications the results of 2.1.—3.1. hold in this case also. The processes shall be linear

$$x_n = \sum_{\nu=0}^{\infty} a_\nu \xi_{n-\nu}, \quad -\infty < n < \infty, \quad (16)$$

where the ξ 's are independent and identically distributed with finite variance. Let

$$\sum_{\nu=0}^{\infty} |a_\nu| < \infty.$$

According to a theorem of Kolmogoroff the sum in (16) converges then almost certainly. We shall have to introduce a more stringent condition later on.

4.2.1. Let x_n be a stationary (strict sense) process whose realizations we denote by $\omega = (\dots x_{-1}, x_0, x_1 \dots)$. The sample space consisting of all ω will be denoted by Ω . If $F(\omega)$ is a function which is measurable and quadratically integrable with respect to the probability measure defined on Ω and if the usual operator of translation is T , then

$$y_n = F[T^n \omega]$$

is evidently a stationary (strict sense) process.

4.2.2. Let x_n be ergodic, i.e. for every $f(\omega) \in L_1(\Omega)$ the limit

$$\hat{f}(\omega) = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{-N}^N f[T^n \omega],$$

which exists almost certainly according to the ergodic theorem, reduces to a constant

$$\hat{f}(\omega) = E f(\omega).$$

Then it is self-evident that also y_n is ergodic.

4.2.3. Suppose now that $F(\omega)$ depends only upon the values of $\dots x_{-2}, x_{-1}, x_0$.

$$F(\omega) = F(\dots x_{-2}, x_{-1}, x_0).$$

Further let x_n be a linear process. It is then automatically purely non-deterministic. The process

$$y_n = F[T^{-n}\omega] - EF[T^{-n}\omega]$$

is also purely non-deterministic. To show this let us assume the contrary. The Hilbert space spanned by y_ν for $\nu \leq n$ will be called $L_2(Y; n)$ and $\lim_{n \rightarrow -\infty} L_2(Y; n) = L_2(Y; -\infty)$. Under our assumption there exists a non-vanishing element $z(\omega) \in L_2(Y; -\infty) \subset L_2(Y; n)$. The last relation holds by definition for all n . Now it is known (see [6]) that

$$\lim_{n \rightarrow -\infty} E[z(\omega) | \xi_n, \xi_{n+1}, \dots] = z(\omega)$$

almost certainly. But as $z(\omega) \in L_2(Y; n-1)$, it is the limit of functions not depending upon $\xi_n, \xi_{n+1}, \xi_{n+2}, \dots$. Then

$$z(\omega) = E z(\omega) = 0$$

contrary to our hypothesis.

4.2.4. Starting from a purely random process ξ_n and defining the function F as

$$F[\omega] = \sum_{\nu=0}^{\infty} a_\nu \xi_{-\nu}, \quad \sum_{\nu=0}^{\infty} |a_\nu| < \infty$$

we get a linear process

$$x_n = F[T^{-n}\omega] = \sum_{\nu=0}^{\infty} a_\nu \xi_{n-\nu}.$$

As a simple application of **4.2.1.**—**4.2.3.** this process is strictly stationary, ergodic and purely non-deterministic.

4.2.5. Now we shall consider a special functional of the realization, viz. the equidistributed estimate

$$m_E^* = \frac{1}{n} \sum_1^n x_\nu$$

which has been considered above.

This estimate is asymptotically normally distributed according to a theorem of MORAN, who has shown [13] that the central limit law applies to the linear processes if

$$\left| \sum_0^{\infty} a_r \right|^2 = 2 \pi f(0) \neq 0.$$

MORAN gives an example of a linear process for which $f(0) = 0$ and for which the central limit law is not valid.

5. Introductory remarks on the estimation of the covariance sequence and the spectral intensity

Suppose that we have observed the process x_n in a finite number of points and have got a sample $(x_{-N}, x_{-N+1}, \dots, x_{N-1}, x_N)$. To simplify the problem let $E x_v \equiv 0$. It is required to form a function of the sample and of n in such a way that this function can be used as an estimate of the covariance sequence r_n . Because of the Fourier representation of r_n this is equivalent to estimating the spectral intensity $f(\lambda)$. Some special cases, e.g. when x_n is an autoregressive process and the problem is reduced to estimating the parameters in the characteristic equation, have been studied a great deal. It seems important, however, to consider this problem of estimation more generally without restricting the type of process too much. This has been pointed out especially by BARTLETT [2].

As $r_n = E x_v x_{v+n}$ it seems natural to consider some quadratic form of the sample values

$$r_n^* = q_n(x_{-N}, \dots, x_N) \tag{17}$$

as an estimate of the covariance sequence. In the same way we can choose another quadratic form

$$f^*(\lambda) = Q(x_{-N}, \dots, x_N; \lambda) \tag{18}$$

as an estimate of the spectral intensity. As

$$\begin{cases} r_n = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda \\ f(\lambda) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} r_n e^{-in\lambda}, \end{cases} \tag{19}$$

where the second formula holds in the cases we are going to consider, it seems natural to connect the estimates in the corresponding way. Further the estimate $f^*(\lambda)$ of the spectral intensity shall be non-negative; r_n^* is then an admissible covariance sequence because of (19).

In addition to these conditions we want some criterion in order to be able to choose between all existing estimates of the above type. To do this we shall study the mean value, and, though in general no unbiased estimate of $f(\lambda)$ exists, we can study the deviation $E f^*(\lambda) - f(\lambda)$ and try to make it small

in some sense. As the estimates introduced above are quadratic forms in the sample values, the knowledge of the covariances of the process is enough for the determination of $E f^*(\lambda)$. To study the sampling fluctuations of the estimates we consider their variances. But to be able to determine these we must in general know the moments of the process up to the 4th order. In order to avoid complications we shall now introduce the assumption that the process is normal, so that its distribution is completely specified by the mean value and the covariance function. It must be admitted that this condition is of a rather restrictive character, but we hope that, if these processes can be treated satisfactorily, the general case will be considerably clarified.

In addition we shall from now on suppose that $f(\lambda)$ is 1) positive, 2) continuous together with its two first derivatives. This regularity condition which is presumably far from necessary will be denoted by (R) .

6. Estimation of the correlation sequence

Having observed the sample (x_{-N}, \dots, x_N) we want to estimate $r_n = E x_\nu x_{\nu+n}$. We possess the following $2N + 1 - |n|$ unbiased estimates of r_n

$$x_{-N} x_{-N+|n|}, x_{-N+1} x_{-N+1+|n|}, \dots, x_{N-|n|} x_N. \tag{20}$$

If $|n| > 2N$ it is evidently impossible to give an unbiased estimate of r_n . Combining the estimates (20) linearly we get

$$r_n^* = \sum_{\nu=-N}^{N-|n|} c_\nu x_\nu x_{\nu+|n|} \tag{21}$$

and to make this estimate unbiased we have to put

$$\sum_{\nu=-N}^N c_\nu = 1. \tag{22}$$

The estimates are then of the following type

$$\begin{cases} r_n^* = \sum_{\nu=-N}^{N-n} c_\nu^{(n)} x_\nu x_{\nu+n}; & n \geq 0 \\ r_{-n}^* = r_n^*; & \sum_{\nu} c_\nu^{(n)} = 1 \\ r_n^* = 0; & |n| \geq 2N + 1. \end{cases} \tag{23}$$

For a fixed n we want to choose the c 's to minimize the variance $D^2 r_n^*$. Introducing the process $y_\nu = x_\nu x_{\nu+n}$ which has mean value r_n , the problem is formally reduced to that studied in 2.1.—3.2. and there is a unique solution. One has to observe, however, that in the present case the unknown parameters r_n enter also into the expression for the covariance sequence of the y -process. The asymptotically best estimate can, however, be explicitly constructed. We shall use the identity

$$E x_1 x_2 x_3 x_4 = E x_1 x_2 E x_3 x_4 + E x_1 x_3 E x_2 x_4 + E x_1 x_4 E x_2 x_3 \quad (24)$$

valid for normally distributed variables with mean zero. The y -process has then the covariance sequence

$$r_v(y) = r_v^2 + r_{v+n} r_{v-n}. \quad (25)$$

The corresponding spectral intensity is then

$$f_y(\lambda) = \int_{-\pi}^{\pi} f(\mu) f(\lambda - \mu) [1 + \cos n(\lambda - 2\mu)] d\mu. \quad (26)$$

From (26) follows that 1) $f_y(\lambda) > 0$ and 2) $f_y(\lambda)$ is continuous together with its two first derivatives, so that $f_y(\lambda)$ satisfies (R). Then the equidistributed estimate

$$r_n^* = \frac{1}{2N + 1} \sum_{v=-N}^{N-n} x_v x_{v+n} \quad (27)$$

is an asymptotically efficient estimate of r_n as N tends to infinity. Its variance is asymptotically

$$D^2\{r_n^*\} \sim \frac{2\pi}{N} \int_{-\pi}^{\pi} f^2(\mu) [1 + \cos 2n\mu] d\mu. \quad (28)$$

When N tends to infinity the estimate

$$\varrho_n^* = \frac{1}{2N + 1} \sum_{v=-N}^{N-n} x_v x_{v-n} \quad (29)$$

is asymptotically equivalent to r_n^* .

7. Some properties of the periodogram

When we turn to the problem of estimating the spectral intensity a rather natural way to proceed would be the following. In 6 we have shown that ϱ_n^* is in an asymptotic sense the best estimate of r_n . To estimate $f(\lambda)$ we could then put ϱ_n^* instead of r_n in (19) and obtain

$$\begin{aligned} f^*(\lambda) &= \frac{1}{2\pi} \sum_v \varrho_v^* e^{-iv\lambda} = \frac{1}{2\pi(2N + 1)} \sum_{v, \mu=-N}^N x_v x_\mu e^{-i(v-\mu)\lambda} = \\ &= \frac{1}{2\pi(2N + 1)} \left| \sum_{v=-N}^N x_v e^{-iv\lambda} \right|^2. \end{aligned}$$

This is nothing else than the periodogram which was originally devised to detect hidden periodicities. It is usually used in practice also as an estimate of the spectral intensity.

Let us introduce the notation

$$I_N(\lambda) = \frac{1}{2\pi(2N+1)} \left| \sum_{\nu=-N}^N x_\nu e^{-i\nu\lambda} \right|^2.$$

If we have the spectral representation

$$x_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda} dZ(\lambda)$$

we get

$$\sum_{\nu=-N}^N x_\nu e^{-i\nu\lambda} = \int_{-\pi}^{\pi} \frac{\sin(N + \frac{1}{2})(l - \lambda)}{\sin \frac{l - \lambda}{2}} dZ(l). \tag{30}$$

Hence

$$EI_N(\lambda) = \frac{1}{2\pi(2N+1)} \int_{-\pi}^{\pi} \frac{\sin^2(N + \frac{1}{2})(l - \lambda)}{\sin^2 \frac{l - \lambda}{2}} f(l) dl,$$

which tends, according to a property of the Féjer kernel and because of (R), to $f(\lambda)$ as N tends to infinity. This fact that $I_N(\lambda)$ is an asymptotically unbiased estimate of $f(\lambda)$ has been made an argument in favour of using $I_N(\lambda)$ as an estimate of $f(\lambda)$.

To compute the variance we note that (24) can be extended (e.g. by analytic continuation of a polynomial from the real axis to the point $z = i$) to the case when the variables can take complex values. We have using (24)

$$EI_N^2(\lambda) = 2[E|\Sigma|^2]^2 + |E\Sigma^2|^2,$$

where

$$\Sigma = \frac{1}{2\pi(2N+1)} \sum_{\nu=-N}^N x_\nu e^{-i\nu\lambda},$$

and hence

$$\begin{aligned} D^2\{I_N(\lambda)\} &= [E|\Sigma|^2]^2 + |E\Sigma^2|^2 = \\ &= \left[\frac{1}{2\pi(2N+1)} \int_{-\pi}^{\pi} \frac{\sin^2(N + \frac{1}{2})(l - \lambda)}{\sin^2 \frac{l - \lambda}{2}} f(l) dl \right]^2 + \\ &+ \left[\frac{1}{2\pi(2N+1)} \int_{-\pi}^{\pi} \frac{\sin(N + \frac{1}{2})(l + \lambda)}{\sin \frac{l + \lambda}{2}} \frac{\sin(N + \frac{1}{2})(l - \lambda)}{\sin \frac{l - \lambda}{2}} f(l) dl \right]^2. \tag{31} \end{aligned}$$

As $f(\lambda)$ is a symmetric function $D^2\{I_N(\lambda)\}$ tends to $2f^2(\lambda) > 0$ resp. $f^2(\lambda)$ as N tends to infinity when $\lambda = 0$ resp. $\lambda \neq 0$. Hence the periodogram is not even a consistent estimate of the spectral intensity.

To see whether $I_N(\lambda)$ converges in the mean to some stochastic variable I , which then would have positive variance, we consider the expression

$$E I_N I_M = E I_N E I_M + |E \Sigma_N \bar{\Sigma}_M|^2 + |E \Sigma_N \Sigma_M|^2.$$

But

$$E \Sigma_N \Sigma_M = \frac{1}{2\pi V(2N+1)(2M+1)} \int_{-\pi}^{\pi} \frac{\sin(N+\frac{1}{2})(l-\lambda)}{\sin\frac{l-\lambda}{2}} \frac{\sin(M+\frac{1}{2})(l+\lambda)}{\sin\frac{l+\lambda}{2}} f(l) dl.$$

Fixing N and letting M tend to infinity we obtain using a property of the Dirichlet kernel

$$\lim_{M \rightarrow \infty} E \Sigma_N \Sigma_M = 0,$$

and analogously

$$\lim_{M \rightarrow \infty} E \Sigma_N \bar{\Sigma}_M = 0.$$

Hence

$$\lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} E I_N I_M = f^2(\lambda). \tag{32}$$

But we should have

$$\lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} E I_N I_M = \lim_{N \rightarrow \infty} E I_N I = E I^2 = \lim_{N \rightarrow \infty} E I_N^2 \geq 2 f^2(\lambda) > f^2(\lambda).$$

Hence the suite $I_N(\lambda)$ does not converge. That $I_N(\lambda)$ shows no regular behaviour when N increases has also been observed in practice (see e.g. [2]).

If λ is considered as a parameter we get a new stochastic process by forming $I_N(\lambda)$. We shall introduce its covariance function

$$R_N(\lambda, \mu) = \text{cov} [I_N(\lambda), I_N(\mu)]$$

and get

$$R_N(\lambda, \mu) = \left[\frac{1}{2\pi(2N+1)} \int_{-\pi}^{\pi} \frac{\sin(N+\frac{1}{2})(l-\lambda)}{\sin\frac{l-\lambda}{2}} \frac{\sin(N+\frac{1}{2})(l-\mu)}{\sin\frac{l-\mu}{2}} f(l) dl \right]^2 + \left[\frac{1}{2\pi(2N+1)} \int_{-\pi}^{\pi} \frac{\sin(N+\frac{1}{2})(l-\lambda)}{\sin\frac{l-\lambda}{2}} \frac{\sin(N+\frac{1}{2})(l+\mu)}{\sin\frac{l+\mu}{2}} f(l) dl \right]^2. \tag{33}$$

DANIELL has shown that the correlation tends to zero when N tends to infinity ([1] p. 90).

Summing up we can describe the behaviour of the periodogram for large values of N in the following way: $I_N(\lambda)$ will fluctuate about $f(\lambda)$ but will in general not tend to this or some other definite value when the sample number increases. As $R_N(\lambda, \mu)$ tends to zero, $I_N(\lambda)$ and $I_N(\mu)$ will be uncorrelated (and even independent, see [12]) in the limit. The curve will hence show a very irregular appearance.

There is a simple paradox here. Although we have used the estimate ϱ_n^* , which in 6 was shown to be asymptotically efficient, we have got a very poor estimate $I_N(\lambda)$. This can be explained in the following way. In 6 we were mainly interested in estimating r_n for n fixed, i.e. a single covariance. But when estimating $f(\lambda)$ we have to estimate *all* the covariances at the same time and this constitutes an entirely different problem.

8. A class of estimates of the spectral intensity

In this section we shall introduce a class of estimates of the spectral intensity. In (19) we shall insert some estimates of the covariances instead of r_n . Although we know that ϱ_n^* is an asymptotically best estimate of r_n , it is clear that for a fixed sample number N the estimates ϱ_n^* will be formed by a smaller and smaller number of the observed values when n increases. It is then natural to introduce some weighting factors $w_\nu(\lambda)$

$$f^*(\lambda) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} w_\nu(\lambda) \varrho_\nu^* e^{-i\nu\lambda} \quad (34)$$

and choose the w 's in order to get a good estimate of $f(\lambda)$. The w 's may depend on λ . It has to be observed that they cannot be chosen completely arbitrarily because $f^*(\lambda)$ should be non-negative. Hence

$$f^*(\lambda) = \frac{1}{2\pi(2N+1)} \sum_{\mu, \nu=-N}^N w_{\nu-\mu}(\lambda) e^{-i(\nu-\mu)\lambda} x_\nu x_\mu \geq 0.$$

Separating the real and imaginary parts of $w_\nu e^{-i\nu\lambda}$ we get

$$w_\nu e^{-i\nu\lambda} = \alpha_\nu + i\beta_\nu$$

and

$$\sum_{\nu, \mu=-N}^N \alpha_{\nu-\mu} x_\nu x_\mu \geq 0; \quad \sum_{\nu, \mu=-N}^N \beta_\nu x_\nu x_\mu = 0. \quad (35)$$

(35) implies that α_ν shall be a positive semidefinite sequence. Then there exists a bounded non-decreasing function $W_\lambda(l)$ so that

$$\alpha_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda} dW_\lambda(l). \quad (36)$$

We thus get the following representation for the estimates under consideration

$$f^*(\lambda) = \frac{1}{2\pi(2N+1)} \sum_{\nu, \mu=-N}^N x_\nu x_\mu \int_{-\pi}^{\pi} e^{i(\nu-\mu)l} dW_\lambda(l) = \int_{-\pi}^{\pi} I_N(l) dW_\lambda(l). \quad (37)$$

$W_\lambda(l)$ can be chosen so that $\Delta W(l) = \Delta W(-l)$, which is supposed in the variance formulae.

Especially if we choose $W_\lambda(l) = \varepsilon(l-\lambda)$ we get the periodogram

$$f^*(\lambda) = I_N(\lambda)$$

which, as we have seen in 6, is not a satisfactory estimate. It seems more promising to choose $W_\lambda(l)$ as an absolutely continuous function with a derivative

$$\frac{dW_\lambda(l)}{dl} = w_\lambda(l). \tag{38}$$

If for different values of λ we just translate the weighting function we get

$$W_\lambda(l) = W(l - \lambda) \tag{39}$$

and

$$w_\nu(\lambda) e^{-i\nu\lambda} = \int_{-\pi}^{\pi} e^{i\nu l} dW(l - \lambda) = e^{i\nu\lambda} \int_{-\pi}^{\pi} e^{i\nu l} dW(u) = w_\nu e^{i\nu\lambda}.$$

Thus

$$f^*(\lambda) = \frac{1}{2\pi} \sum_{\nu} w_\nu \varrho_\nu^* e^{-i\nu\lambda}, \tag{40}$$

and the weighting factors do not depend upon λ when the different weighting functions have been generated by translation.

In [1] p. 89 DANIELL proposed (starting from another point of view) averaging the periodogram over neighbouring frequencies. That is a special case of (37) which will be studied in 10.1.

For a given estimate we can calculate two coefficients ε_1 and ε_2 measuring the errors in the following way

$$\begin{cases} \varepsilon_1 = \int_{-\pi}^{\pi} [E f^*(\lambda) - f(\lambda)]^2 d\lambda \\ \varepsilon_2 = \int_{-\pi}^{\pi} D^2 \{f^*(\lambda)\} d\lambda. \end{cases} \tag{41}$$

From (40) one gets

$$\begin{cases} \varepsilon_1 = \sum_{\nu=-2N-1}^{2N+1} r_\nu^2 \left[w_\nu \frac{2N+1-|\nu|}{2N+1} - 1 \right]^2 + \sum_{|\nu| > 2N+1} r_\nu^2 \\ \varepsilon_2 = \sum_{\nu=-2N-1}^{2N+1} w_\nu^2 D^2 \{\varrho_\nu^*\}. \end{cases}$$

We also have

$$\int_{-\pi}^{\pi} E [f^*(\lambda) - f(\lambda)]^2 d\lambda = \varepsilon_1 + \varepsilon_2.$$

9. Estimates with an absolutely continuous weighting function

9.1. Consider the estimate

$$f^*(\lambda) = \int_{-\pi}^{\pi} I_N(l) w_\lambda(l) dl$$

where $w_\lambda(l)$ is a quadratically integrable function in $(-\pi, \pi)$. We have

$$E f^*(\lambda) = \int_{-\pi}^{\pi} E I_N(l) w_{\lambda}(l) dl. \tag{42}$$

Using the representation of $E I_N(\lambda)$ in 7 and the condition (R) it is seen that $E I_N(\lambda)$ is uniformly bounded, and as it tends to $f(l)$ we get

$$\lim_{N \rightarrow \infty} E f^*(\lambda) = \int_{-\pi}^{\pi} f(l) w_{\lambda}(l) dl. \tag{43}$$

To get a simple and meaningful estimate we let $w(u)$ be periodic with the period 2π and put

$$\begin{cases} w_{\lambda}(l) = w(l - \lambda) \\ w(u) = 0, \quad |h| < |u| < \pi \\ \int_{-h}^h w(u) du = 1. \end{cases} \tag{44}$$

Then we get the estimate

$$f^*(\lambda) = \int_{\lambda-h}^{\lambda+h} I_N(l) w(l) dl \tag{45}$$

which is an asymptotically unbiased estimate of the spectral mass situated in the frequency band $(\lambda - h, \lambda + h)$. We can say that the resolving power of this estimate is limited to frequencies differing at least $2h$.

To evaluate the variance of estimates of the type under consideration we use (33). We have putting $w_{\lambda_0}(l) = w(l)$

$$D_1^2 f^*(\lambda_0) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} R_N^{(1)}(\lambda, \mu) w(\lambda) w(\mu) d\lambda d\mu$$

and then

$$\begin{aligned} 4\pi^2 (2N + 1)^2 D_1^2 f^*(\lambda_0) &= \\ &= \sum_{n, m, k, l = -N}^N \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{in(u-\lambda) + im(u-\mu) + ik(v-\lambda) + il(v-\mu)} \cdot f(u)f(v) w(\lambda) w(\mu) du dv d\mu d\lambda = \\ &= \sum_{n, m, k, l = -N}^N r_{n+m} r_{k+l} w_{n+k} w_{m+l}. \end{aligned} \tag{46}$$

Here we have denoted the first term of the right side of (33) with $R_N^{(1)}(\lambda, \mu)$. The second term gives also the sum in (46).

Introducing the quantity

$$S_v = \sum_{n=-\infty}^{\infty} |r_n w_{n+v}| \tag{47}$$

we get using (R)

$$S_v^2 \leq \sum_{n=-\infty}^{\infty} (1 + |n|)^2 r_n^2 \sum_{n=-\infty}^{\infty} \frac{|w_{n+v}|^2}{(1 + |n|)^2} \leq c \sum_{n=-\infty}^{\infty} \frac{|w_n|^2}{(1 + |n - v|)^2}, \quad (48)$$

and

$$\sum_{v=-\infty}^{\infty} S_v^2 \leq c \sum_{n=-\infty}^{\infty} |w_n|^2 \sum_{v=-\infty}^{\infty} \frac{1}{(1 + |n - v|)^2} < \infty.$$

Hence we have absolute convergence

$$\sum_{n,m,v=-\infty}^{\infty} |r_n w_{n+v} r_m w_{m+v}| < \infty. \quad (49)$$

Returning to (46) we have

$$2\pi^2 (2N + 1)^2 D^2 f^*(\lambda_0) = \sum_{\alpha, \beta, \gamma} c(\alpha, \beta, \gamma) r_\alpha r_\beta w_\gamma w_{\alpha+\beta-\gamma}$$

where

$$0 \leq c(\alpha, \beta, \gamma) \leq 2N + 1$$

and

$$\frac{c(\alpha, \beta, \gamma)}{2N + 1} \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

Because of (49) we get

$$\begin{aligned} \lim_{N \rightarrow \infty} 2\pi^2 (2N + 1) D^2 f^*(\lambda_0) &= \sum_{n,m,v=-\infty}^{\infty} r_n w_{n+v} r_m w_{m+v} = \\ &= \sum_{v=-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} r_n w_{n+v} \right]^2 = 8\pi^3 \int_{-\pi}^{\pi} w^2(l) f^2(l) dl. \end{aligned}$$

Hence we get the asymptotic relation as N tends infinity

$$D^2 f^*(\lambda_0) \sim \frac{2\pi}{N} \int_{-\pi}^{\pi} w^2(l) f^2(l) dl. \quad (50)$$

As this expression tends to zero for large values of N the stochastic variable $f^*(\lambda_0)$ converges in probability to $f(\lambda)$ so that the estimate is consistent.

9.2. If we consider only the asymptotic behaviour of $f^*(\lambda)$ its variance can be minimized subject to a condition of unbiasedness

$$\int_{\lambda_0-h}^{\lambda_0+h} w_{\lambda_0}(l) dl = 1$$

and we then get

$$w_{\lambda_0}(l) = \frac{1}{f^2(l)} \frac{1}{\int_{\lambda_0-h}^{\lambda_0+h} \frac{1}{f^2(l)} dl}$$

and

$$\text{Min} \int_{-\pi}^{\pi} w_{\lambda_0}^2(l) f^2(l) dl = \frac{1}{\int_{\lambda_0-h}^{\lambda_0+h} \frac{1}{f^2(l)} dl}.$$

If the spectral intensity does not vary too much in $(\lambda_0 - h, \lambda_0 + h)$ so that $f(\lambda)$ can be considered as approximately a constant we can take a constant weighting function as the best estimate in the class considered.

Usually we do not know anything about the form of $f(\lambda)$ so that the best weighting function cannot be determined. Then it seems still natural to choose a constant $w(l)$, at least from the point of view of getting simple computations.

9.3. The greater h is chosen the smaller resolvability do we get. Let us call this uncertainty Δ_1 and give it a quantitative meaning by putting $\Delta_1 = h$.

The sampling variability gives rise to another uncertainty, say Δ_2 , and we shall take Δ_2 as proportional (at least asymptotically) to the variance of the estimate:

$$\Delta_2 = \int_{-\pi}^{\pi} w^2(l) f^2(l) dl.$$

Consider the product of the two uncertainties

$$\Delta = \Delta_1 \Delta_2 = h \int_{-\pi}^{\pi} w^2(l) f^2(l) dl.$$

Irrespective of the choice of $w(l)$ we have because of (51)

$$\Delta \geq \frac{1}{\frac{1}{h} \int_{\lambda_0-h}^{\lambda_0+h} \frac{1}{f^2(l)} dl}.$$

The right side is a positive continuous function of h so that there is a positive number $\Delta_0(f)$ depending only upon f satisfying

$$\Delta \geq \Delta_0(f). \tag{52}$$

From this "uncertainty principle" follows that if we want high resolvability we have to sacrifice some precision of the estimate and *vice versa*. As an extreme case we can consider the periodogram which is asymptotically unbiased but not even consistent.

It is possible to extend (52) to estimates whose weighting functions do not necessarily vanish outside some interval $(\lambda_0 - h, \lambda_0 + h) \subset (-\pi, \pi)$. In order that the estimate shall still possess a sort of asymptotic unbiasedness we shall demand that $\int_{-\pi}^{\pi} w(l) dl = 1$ (see (37)). To measure the resolvability we introduce the quantities

$$\begin{cases} \bar{\lambda} = \int_{-\pi}^{\pi} l w(l) dl \\ \Delta'_1 = + \sqrt{\int_{-\pi}^{\pi} (l - \bar{\lambda})^2 w(l) dl} \end{cases}$$

We can consider $w(l)$ as a frequency function for a stochastic variable l taking values in $(-\pi, \pi)$. The above quantities are then the mean and the standard deviation of this variable. Using the inequalities of Schwarz and Techebycheff we get

$$\int_{-\pi}^{\pi} w^2(l) dl \geq \int_{\bar{\lambda}-2\Delta'_1}^{\bar{\lambda}+2\Delta'_1} w^2(l) dl \geq \frac{P^2[|l - \bar{\lambda}| < 2\Delta'_1]}{4\Delta'_1} \geq \frac{9}{64\Delta'_1}$$

Hence

$$\Delta' = \Delta'_1 \Delta_2 \geq F \Delta'_1 \int_{-\pi}^{\pi} w^2(l) dl \geq \frac{9}{64} F > 0,$$

so that (52) is replaced by

$$\Delta' \geq \Delta'_0(f). \tag{52 a}$$

10. Some special estimates with absolutely continuous weighting functions

10.1. A simple form of the estimate is obtained when $w(\lambda)$ is chosen as a constant in $(\lambda - h, \lambda + h)$. We then get (supposing $0 < \lambda - h < \lambda + h < \pi$)

$$f^*(\lambda) = \frac{1}{2h} \int_{\lambda-h}^{\lambda+h} I_N(l) dl \tag{53}$$

with

$$\begin{cases} \lim_{N \rightarrow \infty} E f^*(\lambda) = \frac{1}{2h} \int_{\lambda-h}^{\lambda+h} f(l) dl \\ D^2 f^*(\lambda) \sim \frac{\pi}{4h^2 N} \int_{\lambda-h}^{\lambda+h} f^2(l) dl. \end{cases} \tag{54}$$

The corresponding weighting factors are according to (39) and (40)

$$w_\nu = \frac{1}{2h} \int_{-h}^h e^{i\nu l} dl = \frac{1}{h} \frac{\sin \nu h}{\nu}$$

so that

$$f^*(\lambda) = \frac{1}{2\pi} \sum_{-2N-1}^{2N+1} \frac{1}{h\nu} \sin \nu h \varrho_\nu^* e^{-i\nu\lambda}.$$

10.2. Bartlett has proposed an estimate which we shall denote $f_B^*(\lambda)$. For the motivation we refer to [2], where this estimate has been applied to Kendall's artificial series. Let $2N + 1 = nm$ where n and m are positive integers. Then

$$f_B^*(\lambda) = \frac{1}{2\pi} \sum_{\nu=-n+1}^{n-1} \left(1 - \frac{|\nu|}{n}\right) r_\nu^* e^{-i\nu\lambda} \tag{55}$$

so that

$$f_B^*(\lambda) = \frac{1}{2\pi} \sum_{\nu=-n+1}^{n-1} \frac{1 - \frac{|\nu|}{n}}{1 - \frac{|\nu|}{nm}} e_\nu^* e^{-i\nu\lambda}. \tag{56}$$

To study $f_B^*(\lambda)$ we introduce a modified form of this estimate

$$f_m^*(\lambda) = \frac{1}{2\pi} \sum_{\nu=-n+1}^{n-1} \left(1 - \frac{|\nu|}{n}\right) e_\nu^* e^{-i\nu\lambda}. \tag{57}$$

The weighting factors are

$$w_\nu^{(n)} = \begin{cases} 1 - \frac{|\nu|}{n}, & |\nu| \leq n \\ 0, & |\nu| > n \end{cases} \tag{58}$$

and the weighting function

$$w_n(l) = \frac{1}{2\pi n} \frac{\sin^2 \frac{n}{2} l}{\sin^2 \frac{l}{2}}. \tag{59}$$

As $E I_N(\lambda)$ tends uniformly to $f(\lambda)$ we get from (37) after a simple reasoning

$$\lim_{m, n \rightarrow \infty} E f_m^*(\lambda) = f(\lambda). \tag{60}$$

To get an asymptotic expression for the variance we have to modify the deduction in 9.1. if n and m are tending to infinity in some arbitrary manner, because in the present case $w_n(\lambda)$ depends upon n and varies as n and m tend to infinity.

Consider the sum

$$S = \frac{1}{n} \sum_{\alpha=\alpha_1}^{\alpha_2} \sum_{\beta=\beta_1}^{\beta_2} \sum_{\gamma=\gamma_1}^{\gamma_2} |r_\alpha w_{\alpha+\gamma}^{(n)} r_\beta w_{\beta+\gamma}^{(n)}|. \tag{61}$$

As $|w_\nu^{(n)}| \leq 1$ and as at most $2n + 1$ weighting factors appearing in the sum are different from zero we get immediately

$$S \leq 3 \sum_{\alpha=\alpha_1}^{\alpha_2} |r_\alpha| \sum_{\beta=\beta_1}^{\beta_2} |r_\beta|.$$

Hence the sum (49) divided with n converges uniformly in n . But then

$$2\pi^2 m D^2 f_m^*(\lambda) = \sum_{\alpha, \beta, \gamma} \frac{c(\alpha, \beta, \gamma)}{2N+1} \frac{r_\alpha r_\beta w_\gamma w_{\alpha+\beta-\gamma}}{n} =$$

$$= \frac{1}{n} \sum_{\alpha, \beta, \gamma} r_\alpha r_\beta w_\gamma w_{\alpha+\beta-\gamma} - \sum_{\alpha, \beta, \gamma} \left(1 - \frac{c(\alpha, \beta, \gamma)}{2N+1}\right) \frac{r_\alpha r_\beta w_\gamma w_{\alpha+\beta-\gamma}}{n},$$

and because of the uniform convergence, which has just been shown, the second term tends to zero. Hence, after some calculations of Fourier type,

$$\lim_{n, m \rightarrow \infty} m D^2 \{f_m^*(\lambda)\} = \begin{cases} \frac{1}{3} f^2(0), & \lambda = 0. \\ \frac{2}{3} f^2(\lambda), & \lambda \neq 0. \end{cases}$$

Putting

$$w_\nu^{(n)} = \begin{cases} 1 - \frac{\nu}{n}, & |\nu| \leq n \\ 1 - \frac{\nu}{nm} & \\ 0 & , \quad |\nu| > n \end{cases}$$

and

$$v_\nu = u_\nu^{(n)} - w_\nu^{(n)}$$

we see that to $f_B^*(\lambda) - f_m^*(\lambda)$ correspond the weighting factors v_ν . But if $|\nu| < n$

$$|v_\nu| = |u_\nu^{(n)} - w_\nu^{(n)}| = \frac{1}{m} \frac{|\nu|}{n} \frac{1 - \frac{|\nu|}{n}}{1 - \frac{|\nu|}{nm}} \leq \frac{2}{m} w_\nu^{(n)};$$

otherwise $v_\nu = 0$. (We have supposed $m \geq 2$.) From this follows that

$$|E[f_B^*(\lambda) - f_m^*(\lambda)]| < \frac{k}{m}$$

and using (60) we get

$$\lim_{m, n \rightarrow \infty} E f_B^*(\lambda) = f(\lambda).$$

We have

$$\frac{1}{n} \sum_{\alpha, \beta, \gamma=-\infty}^{\infty} |r_\alpha v_{\alpha+\gamma} r_\beta v_{\beta+\gamma}| \leq \frac{8}{m^2} \left[\sum_{-\infty}^{\infty} |r_\alpha| \right]^2$$

so that

$$\overline{\lim}_{n, m \rightarrow \infty} m^3 D^2 \{f_B^*(\lambda) - f_m^*(\lambda)\} < \infty.$$

From the triangular inequality we get

$$\left| \|f_B^* - E f_B^*\| - \|f_m^* - E f_m^*\| \right| \leq \|f_B^* - f_m^* - E(f_B^* - f_m^*)\| = D(f_B^* - f_m^*)$$

which is at most of the order $m^{-\frac{3}{2}}$. But $\|f_m^* - E f_m^*\| = D(f_m^*)$ is of the order $m^{-\frac{1}{2}}$. Hence

$$\lim_{n, m \rightarrow \infty} m D^2 \{f_B^*(\lambda)\} = \lim_{n, m \rightarrow \infty} m D^2 \{f_m^*(\lambda)\} = \begin{cases} \frac{4}{3} f^2(0), & \lambda = 0. \\ \frac{2}{3} f^2(\lambda), & \lambda \neq 0. \end{cases} \quad (62)$$

Bartlett's estimate is asymptotically unbiased contrary to those of the form (53). It would be possible to make these unbiased in the limit too by letting h tend to zero as N tends to infinity. Because of the uncertainty principle we must be ready, however, to sacrifice some precision of the estimates if we want higher resolvability. The variance of $f_B^*(\lambda)$ e.g., has a variance proportional to $m^{-1} = n \cdot N^{-1}$ which is infinitely large compared to the variance in (54) when n tends to infinity. — One could define the asymptotic efficiency of an estimate f_0^* e.g. by putting

$$e(f_0^*, f) = \frac{\min \Delta(f^*, f)}{\Delta(f_0^*, f)},$$

or one could use Δ' instead of Δ .

11. Some remarks on the case of a continuous time-parameter

11.1. In sections 5–10 we have only dealt with processes whose parameter takes integral values. There seems to be no essential difficulty to extend the methods used to the case of a continuous parameter. In this section we shall, however, only point out some possible ways of approximating a continuous parameter process with a discrete one and show what the consequences are for the problem of estimating the spectral intensity.

11.2. Let $x(t)$ be a real-valued stationary process with mean value zero and covariance function $r(t)$ with corresponding spectral intensity $f(\lambda)$

$$r(t) = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda) d\lambda.$$

The process is only observed in the equidistant time points $h\nu$, where $h > 0$ and ν runs through the integers. Call the observed process x_ν . Its covariance function is then

$$r_\nu = \int_{-\infty}^{\infty} e^{i\nu h \lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{i\nu \lambda} f_x(\lambda) d\lambda, \quad (63)$$

where we have put

$$f_x(\lambda) = \frac{1}{h} \sum_{-\infty}^{\infty} f\left(\frac{\lambda + 2n\pi}{h}\right). \quad (64)$$

The sum converges for almost all λ . As is easily seen it is no use trying to find a consistent estimate of $f(\lambda)$. Because even if we knew the probability

distribution of $\{x_\nu\}$ completely we could only calculate $f_h(\lambda)$ which does not determine $f(\lambda)$ uniquely. But if h is so small that the spectral mass outside the frequency band $\left(-\frac{\pi}{h}, \frac{\pi}{h}\right)$ can be neglected, then we need only consider the term corresponding to $n = 0$ in (64) and $f(\lambda)$ can be calculated (approximately).

11.3. Sometimes it may be more realistic to suppose that what we observe is not $x_\nu = x(\nu h)$ but

$$y_\nu = \frac{1}{h} \int_{(\nu-\frac{1}{2})h}^{(\nu+\frac{1}{2})h} x(t) dt.$$

(See [11] p. 387.)

The covariance sequence of y_ν is found to be

$$r_\nu(y) = \frac{4}{h^2} \int_{-\infty}^{\infty} e^{i\nu h \lambda} \frac{\sin^2 \frac{h \lambda}{2}}{\lambda^2} f(\lambda) d\lambda.$$

Hence its spectral intensity is (see (64))

$$f_y(\lambda) = \frac{4}{h} \sin^2 \sum_{-\infty}^{\infty} \frac{f\left(\frac{\lambda + 2n\pi}{h}\right)}{(\lambda + 2n\pi)^2}. \tag{65}$$

If we use a y -sample to estimate $f_x(\lambda)$ we make a systematic error because for $-\pi < \lambda < \pi$ we have

$$4 \frac{\sin^2 \frac{\lambda}{2}}{(\lambda + 2n\pi)^2} < 4 \frac{\sin^2 \frac{\lambda}{2}}{\lambda^2}$$

which is less than one if λ is different from zero. Hence we underestimate $f_x(\lambda)$ especially at the extreme ends of the spectral interval. Let us again suppose that h is sufficiently small. Then

$$f_y(\lambda) \simeq \frac{4}{h} \frac{\sin^2 \frac{\lambda}{2}}{\lambda^2} f\left(\frac{\lambda}{h}\right) \simeq \frac{4 \sin^2 \frac{\lambda}{2}}{\lambda^2} f_x(\lambda)$$

and

$$f_x(\lambda) \simeq \frac{\lambda^2}{4 \sin^2 \frac{\lambda}{2}} f_y(\lambda).$$

except perhaps in the neighbourhood of the points $\pm \pi$.

11.4. Suppose that we observe the z -process defined in the following way. Introduce a stochastic variable Δ whose characteristic function is $\varphi(\lambda) = E e^{i\Delta\lambda}$. On this variable we make independent observations $\Delta_\nu, \nu = 0, \pm 1, \pm 2, \dots$. Put

$$z_\nu = x(\nu + \Delta_\nu).$$

We get the covariance sequence

$$r_{\nu-\mu}(z) = E \int_{-\infty}^{\infty} e^{i(\nu-\mu)h\lambda + i(\Delta_\nu - \Delta_\mu)h\lambda} f(\lambda) d\lambda = \begin{cases} \int_{-\infty}^{\infty} e^{i(\nu-\mu)h\lambda} |\varphi(h\lambda)|^2 f(\lambda) d\lambda & \text{if } \nu \neq \mu \\ \int_{-\infty}^{\infty} f(\lambda) d\lambda & \text{if } \nu = \mu \end{cases}$$

The corresponding spectral intensity is

$$f_z(\lambda) = \frac{1}{h} \sum_{-\infty}^{\infty} |\varphi(\lambda + 2n\pi)|^2 f\left(\frac{\lambda + 2n\pi}{h}\right) + c, \tag{66}$$

where

$$c = \int_{-\infty}^{\infty} f(\lambda) [1 - |\varphi(\lambda)|^2] d\lambda.$$

As

$$\int_{-\infty}^{\infty} f_z(\lambda) d\lambda = \int_{-\infty}^{\infty} f(\lambda) d\lambda = \int_{-\pi}^{\pi} f_x(\lambda) d\lambda$$

the total spectral mass (the variance) is the same for the x - and z -processes, only the distributions of the mass differ. If we again suppose that h is sufficiently small one finds

$$f_z(\lambda) \simeq |\varphi(\lambda)|^2 f_x(\lambda) + c.$$

Hence by using the z -sample to estimate the spectrum of the x -process one overestimates the spectral intensity for frequencies where $|\varphi(\lambda)|$ is near 1 and underestimates it where $|\varphi(\lambda)|$ is small.

If we take Δ to be rectangularly distributed we get

$$\varphi(\lambda) = \frac{4}{\lambda} \sin \frac{\lambda}{2}.$$

12. We shall point out some questions in connection with empirical spectral analysis which have not been studied in this paper:

1. We have all the time supposed that the mean value of the process was zero. In practice this is not known. Is it then appropriate to base the calculations upon the values $x'_\nu = x_\nu - \bar{x}$?

2. Asymptotic expressions for the mean value and standard deviation of $f^*(\lambda)$ have been given. One should like to know also the asymptotic distribution of the estimate.

3. Is it possible to extend the results to non-normal linear processes, e.g. using 4.2.5?

4. In order to study $f^*(\lambda)$ for small and moderate values of N one could try to determine $E f^*(\lambda)$ and $D f^*(\lambda)$ exactly for some process of simple type. The result should be compared with the corresponding asymptotic expressions.

It would also be interesting to study various forms of estimates by calculating them from artificial time-series with known spectra.

REFERENCES. [1] **Bartlett, M. S.**: On the theoretical specification and sampling properties of autocorrelated time-series, Supplement to the Journal of the Royal Statistical Society, Vol. VIII, 1946. (See also Discussion on the papers p. 85—97.) — [2] —: Periodogram analysis and continuous spectra. *Biometrika*, Vol. 37, 1950. — [3] **Cramér, H.**: On the theory of stationary random processes, *Annals of Math.*, Vol. 41, 1940. — [4] —: *Mathematical Methods of Statistics*, Princeton 1946. — [5] **Dolph, C. L.** and **Woodbury, M. A.**: Optimal linear prediction of stochastic processes whose covariances are Green's functions, unpublished manuscript. — [6] **Doob, J. L.**: The law of large numbers for continuous stochastic processes, *Duke Math. Journal* 1940. — [7] **Grenander, U.**: Stochastic processes and statistical inference, *Arkiv för Mat.*, Band 1, 1950. — [8] **Hanner, O.**: Deterministic and non-deterministic stationary stochastic processes, *Arkiv för Mat.*, Band 1, 1949. — [9] **Karhunen, K.**: Über lineare Methoden in der Wahrscheinlichkeitsrechnung, *Ann. Ac. Sci. Fennicae, A I* 37, 1947. — [10] —: Über die Struktur stationärer zufälliger Funktionen, *Arkiv för Mat.*, Band 1, 1949. — [11] **Koopmans, T.**: Statistical inference in dynamic economic models, New York 1950. — [12] **Matérn, B.**: Independence of non-negative quadratic forms in normally correlated variables, *Annals of Math. Stat.*, Vol. XX, 1949. — [13] **Moran, P. A. P.**: Some theorems on time series. I. *Biometrika* XXXIV, 1947. — [14] **Wiener, N.**: *Extrapolation, interpolation and smoothing of stationary time-series*, New York 1949.

Tryckt den 31 maj 1951

Uppsala 1951. Almqvist & Wiksells Boktryckeri AB