

## On Toeplitz forms and stationary processes

By ULF GRENANDER

### 1. Introduction

This paper is intended to show that the theory of Toeplitz forms is applicable to some important linear statistical problems concerning stationary stochastic processes with a discrete time-parameter. Although the main results on Toeplitz forms have been known for more than 30 years they do not seem to have attracted the attention they deserve of the mathematical statisticians. This motivates a short summary of the most important results in this theory.

That summary is given in the next section. In 3 we study the problem of prediction in the light of what has been said in 2 about Toeplitz forms. Section 4 is devoted to the problem of finding asymptotic expressions for the distribution of certain quadratic forms which is of importance in the statistical analysis of stationary time series. In 5, finally, we study the estimation of the mean value of the process and a similar but more general problem of estimation which i.a. contains a theorem of KOLMOGOROFF on interpolation of stationary processes as a special case.

### 2. Some results from the theory of Toeplitz forms

2.1. Consider a Hermitian matrix  $M$  of the form

$$M = \begin{pmatrix} c_0 & c_{-1} & c_{-2} & \dots & c_{-n} \\ c_1 & c_0 & c_{-1} & \dots & c_{-n+1} \\ \dots & \dots & \dots & \dots & \dots \\ c_n & c_{n-1} & c_{n-2} & \dots & c_0 \end{pmatrix} = \{c_{\nu-\mu}\}; \quad \nu, \mu = 0, 1, \dots, n. \quad (1)$$

Then  $M$  is said to be a *Toeplitz matrix* and the adjoint quadratic form

$$T = \sum_{\nu, \mu=0}^n c_{\nu-\mu} x_\nu \bar{x}_\mu$$

is called a *Toeplitz form*. An important class of Toeplitz forms is defined in the following way. Let  $f(\lambda)$  be a non-negative Lebesgue-integrable function defined in the interval  $(-\pi, \pi)$ , and let the matrix elements be defined as

$$c_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda} f(\lambda) d\lambda. \tag{2}$$

Introducing the customary definition of inner product

$$(\varphi, \mu) = \int_{-\pi}^{\pi} \varphi(\lambda) \overline{\psi(\lambda)} f(\lambda) d\lambda,$$

we can apply Schmidt's well-known orthogonalization procedure to the set of functions  $1, e^{i\lambda}, e^{2i\lambda}, \dots$ . In this way we obtain an ON system

$$\varphi_0(\lambda), \varphi_1(\lambda), \varphi_2(\lambda), \dots$$

so that

$$(\varphi_\nu, \varphi_\mu) = \delta_{\nu\mu}.$$

2.2. SZEGÖ has studied (see [9], [10]) the following important problem. Let  $\alpha$  be a complex number. Which is the polynomial  $P_n(e^{i\lambda})$  of the  $n^{\text{th}}$  order in  $e^{i\lambda} = z$  that makes the integral

$$\int_{-\pi}^{\pi} |P_n(e^{i\lambda})|^2 f(\lambda) d\lambda \tag{3}$$

a minimum under the condition

$$P_n(\alpha) = 1.$$

The minimum of (3) is denoted by  $\mu_n(\alpha; f)$ .

Using Schwarz' inequality SZEGÖ shows that the minimum is obtained for

$$P_n(z) = \mu_n(\alpha; f) \{ \overline{\varphi_0(\alpha)} \varphi_0(z) + \overline{\varphi_1(\alpha)} \varphi_1(z) + \dots + \overline{\varphi_n(\alpha)} \varphi_n(z) \} = \mu_n(\alpha; f) s_n(\alpha, z), \tag{4}$$

where

$$\frac{1}{\mu_n(\alpha; f)} = |\varphi_0(\alpha)|^2 + |\varphi_1(\alpha)|^2 + \dots + |\varphi_n(\alpha)|^2. \tag{5}$$

2.3. In the interior of the unit circle in the  $\alpha$ -plane there is defined a function  $G(\alpha; f)$ . If the integral  $\int_{-\pi}^{\pi} \log f(\lambda) d\lambda$  diverges  $G(\alpha; f)$  is put identically equal to zero. Otherwise we shall have

$$G(\alpha; f) = e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) \frac{1-\rho^2}{1-2\rho \cos(\lambda-\varphi) + \rho^2} d\lambda},$$

where  $\alpha = \rho e^{i\varphi}$ ;  $|\rho| < 1$ .

Evidently the sequence  $\mu_n(\alpha; f)$  is non-increasing. SZEGÖ shows that

$$\mu(\alpha; f) = \lim_{n \rightarrow \infty} \mu_n(\alpha; f) = 2\pi G(\alpha; f)(1 - |\alpha|^2). \tag{6}$$

Further this relation holds already for a finite value of  $n$

$$\mu_n(\alpha; f) = 2\pi G(\alpha; f)(1 - |\alpha|^2)$$

if and only if

$$f(\lambda) = \frac{1}{|1 - \bar{\alpha}z|^2} \frac{1}{\varphi(\lambda)},$$

where  $\varphi(\lambda)$  denotes a positive trigonometric polynomial of order  $n$ .

2.4. An other auxiliary function  $D(z)$  is introduced in the following familiar way.

If  $\int_{-\pi}^{\pi} \log f(\lambda) d\lambda$  converges we use the coefficients of the formal Fourier series

$$\log f(\lambda) \sim k_0 + 2 \sum_1^{\infty} (k_n \cos n\lambda + l_n \sin n\lambda)$$

to form a function

$$g(z) = \frac{k_0}{2} + \sum_1^{\infty} (k_n - i l_n) z^n$$

and put

$$D(z) = e^{g(z)}.$$

Then  $D(z)$  is analytic and free from zeroes in  $|z| < 1$  and  $|D(z)|^2 = G(z; f)$ .

Then the series

$$\lim_{n \rightarrow \infty} s_n(\alpha, z) = \overline{\varphi_0(\alpha)} \varphi_0(z) + \overline{\varphi_1(\alpha)} \varphi_1(z) + \dots$$

converges for  $|\alpha| < 1$ ,  $|z| < 1$ , and uniformly for  $|\alpha|$  and  $|z| \leq r < 1$  to the sum

$$\frac{1}{2\pi} \frac{1}{1 - \bar{\alpha}z} \frac{1}{D(\alpha)} \frac{1}{D(z)}. \tag{7}$$

2.5. Introduce the *Toeplitz determinants*

$$D_n(f) = \begin{vmatrix} c_0 & c_{-1} & \dots & c_{-n} \\ c_1 & c_0 & \dots & c_{-n+1} \\ \dots & \dots & \dots & \dots \\ c_n & c_{n-1} & \dots & c_0 \end{vmatrix} > 0,$$

where the inequality sign follows because the matrix is positive definite. SZEGÖ shows (see [9] p. 192) that

$$\frac{D_n(f)}{D_{n-1}(f)} = \mu_n(0; f), \tag{8}$$

which combined with (6) gives

$$\lim_{n \rightarrow \infty} \frac{D_n(f)}{D_{n-1}(f)} = \lim_{n \rightarrow \infty} \frac{n+1}{\sqrt{D_n(f)}} = 2\pi e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda}. \tag{9}$$

2.6. To the matrix (1), (2) there corresponds  $n + 1$  eigen-values  $\lambda_0^{(n)}, \lambda_1^{(n)}, \dots, \lambda_n^{(n)}$ . SZEGÖ has shown that they coincide on the average with the equidistant ordinates of  $2\pi f(\lambda)$ . This is made a rigorous mathematical statement in the following way (see [9] p. 194): Suppose that  $-\infty < m \leq f(\lambda) \leq M < \infty$ ;  $f(\lambda)$  is not restricted to non-negative values. Then

$$2\pi m \leq \lambda_\nu^{(n)} \leq 2\pi M; \quad (\nu = 0, 1, \dots, n; n = 1, 2, \dots). \tag{10}$$

Further if  $F(x)$  is an arbitrary continuous function of  $x$  in the interval  $m \leq x \leq M$  we have

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} [F(\lambda_0^{(n)}) + F(\lambda_1^{(n)}) + \dots + F(\lambda_n^{(n)})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} F[2\pi f(\lambda)] d\lambda. \tag{11}$$

(11) is deduced from the only apparently more special formula (9).

For proofs and further developments of the theorems quoted in this section the reader is referred to SZEGÖ's original papers [9] and [10].

### 3. Extrapolation of stationary processes

3.1. Let  $\{x_n\}$  be a stationary (wide sense) process whose time parameter  $n$  takes all integral values. We assume that  $E x_n \equiv 0$ . According to a well-known theorem of BOCHNER its covariance sequence can be represented as a Fourier-Stieltjes integral

$$r_h = E x_n \bar{x}_{n+h} = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda)$$

where  $F(\lambda)$  is a bounded, non-decreasing function of  $\lambda$ . The process itself has the analogous representation

$$x_n = \int_{-\pi}^{\pi} e^{in\lambda} dz(\lambda), \tag{12}$$

where  $z(\lambda)$  is an orthogonal process corresponding to  $F(\lambda)$ . According to KARHUNEN [7] there is an isometry between the Hilbert spaces spanned by the process  $\{x_n\}$  resp. by the exponentials  $\{e^{in\lambda}\}$ .

WOLD has shown [15] that the process can be decomposed into two processes orthogonal to each other

$$x_n = y_n + z_n,$$

where  $\{y_n\}$  is purely indeterministic and  $\{z_n\}$  a deterministic process. The part  $\{y_n\}$  consists of the absolutely continuous part of the spectrum if the spectral intensity  $f(\lambda) = F'(\lambda)$  satisfies

$$\int_{-\pi}^{\pi} \log f(\lambda) d\lambda > -\infty; \tag{13}$$

otherwise  $\{y_n\}$  reduces to zero.

**3.2.** In 3.2–3.3 we shall suppose that the process is purely indeterministic so that the spectrum is absolutely continuous and (13) holds. Having observed a sample  $(x_{-n}, x_{-n+1}, \dots, x_{-1})$  we want to extrapolate the realization in the sense of least squares to the time 0. This problem has been solved for  $n = \infty$  (see [13]). Denote the extrapolation by

$$x_0^* = -c_1 x_{-1} - c_2 x_{-2} - \dots - c_n x_{-n}.$$

Then

$$E|x_0 - x_0^*|^2 = \int_{-\pi}^{\pi} |\pi_n(e^{-i\lambda})|^2 f(\lambda) d\lambda = \min$$

under the condition that the  $n^{\text{th}}$  order polynomial  $\pi_n(z) = \sum_0^n c_\nu z^\nu$  satisfies

$$\pi_n(0) = c_0 = 1.$$

Introducing the polynomial  $P_n(z) = \sum_0^n \bar{c}_\nu z^\nu$  we get instead

$$\left. \begin{aligned} \int_{-\pi}^{\pi} |P_n(e^{i\lambda})|^2 f(\lambda) d\lambda &= \min \\ P_n(0) &= 1 \end{aligned} \right\}$$

This is however a special case of SZEGÖ's minimum problem. The solution is (see 2.2)

$$\left. \begin{aligned} P_n(z) &= \mu_n(0; f) s_n(0, z) \\ E_{\min} |x_0 - x_0^*|^2 &= \mu_n(0; f) \end{aligned} \right\}$$

**3.3.** Let us see what happens when  $n$  tends to infinity. From (8) and (9) we get the following expression for the limit error

$$\lim_{n \rightarrow \infty} E_{\min} |x_0 - x_0^*|^2 = \mu(0; f) = 2\pi e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda} \quad (14)$$

This formula can also be found in [13].

The best (linear) extrapolator  $\lim_{n \rightarrow \infty} x_0^*$  can also be explicitly constructed. For a finite value of  $n$  we get from (12)

$$x_0^* = \int_{-\pi}^{\pi} [1 - P_n(e^{i\lambda})] dz(\lambda).$$

Introduce the function (see (4), (6) and (7))

$$\Phi(\lambda) = 1 - \frac{\overline{D(0)}}{D(e^{i\lambda})} = \frac{D(e^{i\lambda}) - \overline{D(0)}}{D(e^{i\lambda})}.$$

It is seen in the usual way that the stochastic variable

$$\int_{-\pi}^{\pi} \Phi(\lambda) dz(\lambda)$$

can be constructed as a limit in the mean of  $(\dots x_{-2}, x_{-1})$ . The simplest way to see that it is the best prediction is to compute the corresponding error

$$E \left| x_0 - \int_{-\pi}^{\pi} \Phi(\lambda) dz(\lambda) \right|^2 = \int_{-\pi}^{\pi} \left| \frac{D(0)}{D(e^{i\lambda})} \right|^2 f(\lambda) d\lambda = 2\pi e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda}$$

which equals (14). As the best prediction is unique (but for equivalence), being the projection of the element  $x_0$  upon the subspace  $L_2(X; \nu < 0)$ , the result follows.

It may happen that the best extrapolation can be obtained using only a finite number of the observed values of the process in the past. From 2.3 we see that this is true if and only if the spectral intensity is the reciprocal of a positive trigonometrical polynomial, i.e. in the case of an autoregressive process.

3.4. Suppose now that the process is of the general type. We form in the usual way the two spaces  $L_2(Y; \nu < 0)$  and  $L_2(Z; \nu < 0)$  which are orthogonal to each other. It can be shown that  $L_2(X; \nu < 0) = L_2(Y; \nu < 0) \oplus L_2(Z; \nu < 0)$  (see [4] p. 165). We then get the best extrapolation  $x_0^*$  as

$$\begin{aligned} x_0^* &= P_{L_2(X; \nu < 0)} x_0 = P_{L_2(X; \nu < 0)} y_0 + P_{L_2(X; \nu < 0)} z_0 = \\ &= P_{L_2(Y; \nu < 0)} y_0 + P_{L_2(Z; \nu < 0)} z_0 = y_0^* + z_0^*. \end{aligned}$$

But as  $\{z_n\}$  is deterministic we have  $z_0 = z_0^*$ , and hence the error

$$E|x_0 - x_0^*|^2 = E|y_0 - y_0^*|^2$$

depends only upon the purely non-deterministic part of the process and is then given by (14). This will be used below.

#### 4. Quadratic forms of stationary processes

4.1. Consider the moment matrix  $M_n$  belonging to the  $n+1$ -dimensional distribution of the variables  $x_0, x_1, \dots, x_n$  (we shall suppose the process to be real-valued from now on),

$$M_n = \{r_{\nu-\mu}\} = \begin{Bmatrix} r_0 & r_{-1} & \dots & r_{-n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_n & r_{n-1} & \dots & r_0 \end{Bmatrix}, \quad n = 1, 2, \dots,$$

and the corresponding determinant

$$M_n = |r_{\nu-\mu}| = \begin{vmatrix} r_0 & r_{-1} & \dots & r_{-n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_n & r_{n-1} & \dots & r_0 \end{vmatrix}, \quad n = 1, 2, \dots$$

If the spectrum is absolutely continuous the matrices  $M_n$  are of the same form as (1), (2). SZEGÖ's theorem on the asymptotic distribution of the eigenvalues of such matrices (see 2.6) can immediately be applied. Essentially the same result has been rediscovered by WHITTLE without using SZEGÖ's work (see [12]).

4.2. We now turn to SZEGÖ's asymptotic relation (9) for the Toeplitz determinants, also found by WHITTLE [12]. We shall give a proof using familiar statistical methods and which is valid under more general conditions.

Let  $\{x_n\}$  be an arbitrary process of the type described in 3.1 (not necessarily purely indeterministic). We form the best extrapolation  $x_0^*$  using  $(x_{-n}, x_{-n+1}, \dots, x_{-1})$ . As shown in 3.4 the corresponding error  $E|x_0 - x_0^*|^2$  tends to the error obtained if we had only considered the purely non-deterministic part of the process as  $n$  tends to infinity:

$$\lim_{n \rightarrow \infty} E|x_0 - x_0^*|^2 = 2\pi e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda}. \tag{15}$$

On the other hand the well-known formula for the residual variance (see [1] p. 305) gives for a finite value of  $n$

$$E|x_0 - x_0^*|^2 = \sigma_{0, -1, -2, \dots, -n}^2 = \frac{M_n}{M_{n-1}}. \tag{16}$$

Combining (15) and (16) we get the wanted relation

$$\lim_{n \rightarrow \infty} \frac{\log M_n}{n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda + \log 2\pi,$$

which is valid as soon as the purely indeterministic part of the process is not absent.

4.3. Using the asymptotic distribution of the eigen-values WHITTLE [12] has studied the distribution of certain quadratic forms and ratios of quadratic forms in the variables  $x_1, x_2, \dots, x_n$  which are supposed to have a multivariate normal distribution. The following somewhat related result has been given by KAC and SIEGERT [6]. These authors consider the continuous parameter case but the proof holds for the discrete case after trivial modifications. The process is supposed to have an absolutely continuous spectrum with a spectral intensity satisfying

$$\int_{-\pi}^{\pi} f^3(\lambda) d\lambda < \infty.$$

Then the stochastic variable  $\sum_1^n x_v^2$  is asymptotically normally distributed as  $n$  tends to infinity.

In connection with empirical spectral analysis the author has needed the following result which is closely related to the results of KAC-SIEGERT and WHITTLE. Let the process be normally distributed and the spectrum absolutely continuous with a spectral intensity  $f(\lambda)$  for which  $0 < f_0 \leq f(\lambda) \leq f_1 < \infty$  holds. Introduce the quantity

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

i.e. the periodogram. When trying to determine empirically the spectrum of a process the following type of estimate has been studied (see [3]).

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

where the weight function  $w(l)$  is supposed to be continuous. It means no restriction to demand that  $w(l)$  shall be an even function in  $l$ . Asymptotic expressions for the mean and variance of  $f^*$  have been given. In this section we shall prove that  $f^*$  is asymptotically normally distributed as  $N$  tends to infinity.

Putting

$$w_\nu = \int_{-\pi}^{\pi} e^{i\nu l} w(l) dl$$



we get easily

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

This is a quadratic form in  $x_{-N}, x_{-N+1}, \dots, x_N$  with the matrix

$$\mathbf{F} = \left\{ \frac{1}{2\pi(2N+1)} w_{\nu-\mu} \right\}, \quad \nu, \mu = -N, -N+1, \dots, N.$$

Using a well-known relation (see [1] p. 120) we get the characteristic function of  $f^*$  in the form of a determinant

$$\varphi(t) = E e^{itf^*} = |I - 2it\mathbf{R}\mathbf{F}|^{-\frac{1}{2}}$$

where  $\mathbf{R}$  denotes the moment matrix of the observed variables. If the eigenvalues of  $\mathbf{R}\mathbf{F}$  are denoted  $\lambda_1^{(N)}, \lambda_2^{(N)}, \dots, \lambda_{2N+1}^{(N)}$ , we have

$$\log \varphi(t) = -\frac{1}{2} \sum_{\nu=1}^{2N+1} \log(1 - 2it\lambda_\nu^{(N)}).$$

Then

$$\left. \begin{aligned} E f^* &= \sum_{\nu=1}^{2N+1} \lambda_\nu \\ D^2 f^* &= 2 \sum_{\nu=1}^{2N+1} \lambda_\nu^2 \end{aligned} \right\}$$

where we have dropped the superscripts for printing convenience. Consider the normed variable

$$z = \frac{f^* - E f^*}{D f^*}$$

and its characteristic function, say  $\psi(t)$ ; we get

$$\log \psi(t) = -\frac{1}{2} \sum_1^{2N+1} \log \left( 1 - 2it \frac{\lambda_\nu}{D f^*} \right) - it \frac{E f^*}{D f^*} = -\frac{t^2}{2} + R_N,$$

where

$$|R_N| \leq K \frac{|t|^3}{D^3 f^*} \sum_1^{2N+1} |\lambda_\nu|^3.$$

We now have to show that  $R_N$  tends to zero as  $N$  tends to infinity. To do this we shall majorate the  $\lambda$ 's.

Let

$$\mathbf{B} = \{b_{\nu\mu}\} = 2\pi(2N+1)\mathbf{R}\mathbf{F}$$

so that

$$b_{\nu\mu} = \sum_{m=-N}^N r_{\nu-m} w_{m-\mu}.$$

$B$  is in general not a Toeplitz matrix. Introduce

$$b_N = \max_{\|x\|=1} \left| \sum_{\nu,\mu=-N}^N b_{\nu\mu} x_\nu x_\mu \right|^2.$$

We have

$$\sum_{\nu,\mu=-N}^N b_{\nu\mu} x_\nu x_\mu = \sum_{m=-N}^N a_m b_m,$$

where

$$\left. \begin{aligned} a_m &= \sum_{\nu=-N}^N r_{\nu-m} x_\nu = \int_{-\pi}^{\pi} \varphi(x) e^{-imx} f(x) dx \\ b_m &= \sum_{\mu=-N}^N w_{m-\mu} x_\mu = \int_{-\pi}^{\pi} e^{imx} \varphi(-x) w(x) dx \end{aligned} \right\}$$

Here we have put

$$\varphi(x) = \sum_{\nu=-N}^N x_\nu e^{i\nu x}$$

so that

$$\int_{-\pi}^{\pi} |\varphi(x)|^2 dx = 2\pi \sum_{\nu=-N}^N x_\nu^2 = 2\pi.$$

Hence

$$\left. \begin{aligned} \sum_{m=-N}^N |a_m|^2 &\leq \sum_{-\infty}^{\infty} |a_m|^2 \leq 4\pi^2 f_1^2 \\ \sum_{m=-N}^N |b_m|^2 &\leq \sum_{-\infty}^{\infty} |b_m|^2 \leq 4\pi^2 w_{\max}^2 \end{aligned} \right\}$$

This gives us

$$b_N \leq 4\pi^2 f_1 w_{\max} = C$$

where  $C$  does not depend upon  $N$ .

Hence we obtain the wanted relation

$$|\lambda_\nu^{(N)}| < \frac{C}{2\pi(2N+1)}$$

and

$$|R_N| \leq C_1 \frac{t^3}{D^3 f^*} \frac{1}{(2N+1)^2}.$$

In [3] we have shown that  $Df^*$  is of the order of magnitude  $N^{-\frac{1}{2}}$  which could be used to complete the above proof. But in [3] we demanded that the spectral intensity should satisfy certain regularity conditions including differentiability. That is not necessary in the present case. Suppose only that  $f(\lambda)$  has a positive lower bound  $f_0$ . Then

$$\begin{aligned} 2\pi^2(2N+1)^2 D^2 f^* &= \sum_{n,m,k,l=-N}^N r_{n+m} r_{k+l} w_{n+k} w_{m+l} = \\ &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left[ \sum_{n,k=-N}^N w_{n+k} e^{in x + ik y} \right]^2 f(x) f(y) dx dy \geq 4\pi^2 f_0^2 \sum_{n,k=-N}^N w_{n+k}^2 \end{aligned}$$

because the quantity in brackets is real. Hence

$$2\pi^2(2N+1) D^2 f^* \geq 4\pi^2 f_0^2 \sum_{v=-2N}^{2N} \left( 1 - \frac{|v|}{2N+1} \right) w_v^2,$$

so that  $Df^*$  is at least of the order  $N^{-\frac{1}{2}}$  from which the wanted result follows.

**Note.** Using SZEGÖ's results we could have proved that  $b_N$  is bounded in a simpler manner. Because  $\mathbf{B}$  is the product of two matrices,  $\mathbf{R}$  and  $2\pi(2N+1)\mathbf{F}$ , and the eigen-values of each of these is dominated by  $2\pi f_1$  resp.  $2\pi w_{\max}$  (see 2.6). According to [5] p. 210 we then get

$$b_N \leq 4\pi^2 f_1 w_{\max}.$$

4.4. In dealing with Toeplitz matrices it may be worth while to observe the following facts although they must be considered as well-known. Consider the *infinite* real matrices

$$\mathbf{A} = \{ a_{\nu-\mu} \}, \quad (-\infty < \nu, \mu < \infty),$$

$$\mathbf{B} = \{ b_{\nu-\mu} \}, \quad (-\infty < \nu, \mu < \infty),$$

where

$$\begin{cases} a_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda} a(\lambda) d\lambda \\ b_\nu = \int_{-\pi}^{\pi} e^{i\nu\lambda} b(\lambda) d\lambda. \end{cases}$$

If  $\alpha$  and  $\beta$  are constants it is evident that to the matrix  $\alpha\mathbf{A} + \beta\mathbf{B}$  corresponds in the same way an "intensity" which is  $\alpha a(\lambda) + \beta b(\lambda)$ .

Suppose further that both  $a(\lambda)$  and  $b(\lambda)$  are quadratically integrable. Then we can form the product  $\mathbf{A}\mathbf{B}$ , and have using Parseval's relation

$$\sum_{n=-\infty}^{\infty} a_{v-n} b_{n-\mu} = 2\pi \int_{-\pi}^{\pi} e^{i(v-\mu)\lambda} a(\lambda) b(\lambda) d\lambda,$$

so that to  $\mathbf{AB}$  corresponds the intensity  $2\pi a(\lambda) b(\lambda)$ . If the inverse of  $a(\lambda)$  is quadratically integrable we see that  $\mathbf{A}$  has an inverse corresponding to

$$\frac{1}{4\pi^2 a(\lambda)}.$$

### 5. Estimation of the mean value

5.1. We now change our assumptions in that the mean value  $m = E x_n$  is an unknown not necessarily vanishing constant. From the sample  $(x_0, x_1, \dots, x_n)$  we want to construct a linear unbiased estimate  $m^*$  of minimum variance. We get, introducing

$$m^* = \sum_0^n c_v x_v, \quad P_n(z) = \sum_0^n c_v z^v,$$

$$\left. \begin{aligned} \int_{-\pi}^{\pi} |P_n(e^{i\lambda})|^2 f(\lambda) d\lambda &= \min \\ \sum_0^n c_v &= P_n(1) = 1 \end{aligned} \right\}$$

This clearly is a special case of SZEGÖ's problem (see 2.2).

In [2] the author has shown that if 1:  $\{x_n\}$  is purely non-deterministic, 2:  $f(0) \neq 0$ , 3:  $D(e^{i\lambda})$  is differentiable in  $\lambda = 0$ , then the equidistributed estimate

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

is asymptotically efficient as  $n$  tends to infinity. (The third condition can be given a more handy form, see [3]). If we try to obtain this result from SZEGÖ's theory we get only the trivial result  $\mu(1, f) = 0$  (see (5) and [10] p. 179). SZEGÖ's formulas hold in general only for  $|\alpha| < 1$  and in some cases for  $|\alpha| > 1$ . In the estimation problem however  $\alpha = 1$  which demands stronger methods in order to enable us to deal with the deeper problems connected with the behaviour of  $\mu_n(\alpha; f)$  at the circumference  $|\alpha| = 1$ .

5.2. A tool, useful for that purpose, is the following theorem also due to SZEGÖ (see [11] p. 291).

Let  $f(\lambda)$  be integrable in Riemann's sense and let it have the form

$$f(\lambda) = \Phi(\lambda) |(z - z_1)^{\sigma_1} (z - z_2)^{\sigma_2} \dots (z - z_l)^{\sigma_l}|, \quad z = e^{i\lambda},$$

where  $0 < A \leq \Phi(\lambda) \leq B$ , and  $z_\nu = e^{i\lambda_\nu}$  are distinct points on the unit circle,  $\sigma_\nu > 0$ ,  $\nu = 1, 2, \dots, l$ . Let  $f(\lambda)$  be differentiable at the fixed point  $\lambda = a$ ;  $\alpha = e^{ia} \neq z_\nu$ ,  $\nu = 1, 2, \dots, l$ , and let the following be bounded near  $\lambda = a$ :

$$\frac{f(\lambda) - f(a) - f'(a)(\lambda - a)}{(\lambda - a)^2}$$

Then

$$\varphi_n(\alpha) = \alpha^n \{ \overline{D(\alpha)} \}^{-1} + \varepsilon_n, \quad \varepsilon_n \rightarrow 0. \tag{17}$$

We remark that from the above conditions on  $f(\lambda)$  follow conditions 1, 2, and 3 in 5.1 (see [11] p. 274).

Putting  $\alpha = 1$  we get combining (5) and (17)

$$\lim_{n \rightarrow \infty} [n \mu_n(1; f)]^{-1} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\nu=0}^n |\varphi_\nu(1)|^2 = \frac{1}{2\pi f(0)}. \tag{18}$$

But under the stated conditions we have

$$\lim_{n \rightarrow \infty} n \cdot D^2 m_E^* = 2\pi f(0). \tag{19}$$

From (18) and (19) the wanted result follows.

**5.3.** The question arises whether it is possible to weaken the conditions 1, 2 and 3 in 5.1. Let us consider 1 in this section.

Suppose that instead of 1 we allow also a discrete and singular part of the spectrum but demand that a) the spectrum of the process is absolutely continuous in the neighbourhood of  $\lambda = 0$ , b)  $\int_{-\pi}^{\pi} \log f(\lambda) d\lambda > -\infty$ . This new condition is denoted by 1'.

1', 2 and 3 imply the asymptotic effectiveness of  $m_E^*$ . To show this we split the process in two orthogonal components in the usual way  $x_n = y_n + z_n$  (see 3.1). For  $y_n$  1, 2 and 3 are valid and hence

$$\frac{D^2 m^*[y]}{D^2 m_E^*[y]} > 1 - \varepsilon \quad \text{if } n > n_\varepsilon,$$

i.e.

$$D^2 m^*[y] > (1 - \varepsilon) \frac{2\pi f(0)}{n} \quad \text{if } n > n'_\varepsilon. \tag{20}$$

But because of 1' we have

$$D^2 m_E^*[x] \sim \frac{2\pi f(0)}{n},$$

which combined with (20) and  $D^2 m^*[z] \geq 0$  proves the result.

5.4. Contrary to 1, the second condition cannot be weakened. This is shown by the following non-pathological counter example used by SZEGÖ for another purpose.

Consider the following moving average which automatically defines a purely non-deterministic process:

$$x_n = \xi_n - \xi_{n-1}$$

where the  $\xi$ 's may be independent variables distributed in a normal distribution  $N(0, 1)$ . The spectrum is absolutely continuous with

$$f(\lambda) = \frac{1}{\pi} (1 - \cos \lambda),$$

and 1 and 3 but not 2 can be seen to be satisfied. SZEGÖ has calculated ([10] p. 189)  $\mu_n(1; f)$  explicitly. It is of the order  $n^{-3}$ . On the other hand  $D^2 m_E^* = \frac{2}{(n+1)^2}$  so that the asymptotic efficiency of the equidistributed estimate  $m_E^*$  is zero.

5.5. Suppose now that  $E x_n = m_{\lambda_0} e^{i\lambda_0 n}$  where  $m_{\lambda_0}$ , i.e. the mean amplitude corresponding to the discrete spectral line of frequency  $\lambda = \lambda_0$ , is unknown. Considering the process

$$y_n = x_n e^{-i\lambda_0 n}$$

with spectral intensity  $f(\lambda - \lambda_0)$  and mean value  $E y_n = m_{\lambda_0}$  we are back at the problem of 5.1. We see that  $m_{\lambda_0}$  is estimated with asymptotic efficiency one by

$$m_{\lambda_0}^* = \frac{1}{n} \sum_{I}^n x_n e^{-i\lambda_0 n},$$

under analogous conditions. The local conditions upon the spectrum shall in this case hold for the neighbourhood of  $\lambda = \lambda_0$ .

5.6. One could proceed and study what happens under a non-stationary hypothesis

$$E x_n = m \alpha^n; \quad \alpha = \rho e^{i\lambda_0}.$$

As we have seen in 3.2 the choice  $\rho = 0$  gives us the problem in prediction. The "stationary case"  $\rho = 1$  has been treated in 5.1-5.5.

For  $\rho \neq 1$  we can use SZEGÖ's results to elucidate the corresponding estimation problem. We will however proceed in another way and only refer to [10].

Let the hypothetical values of the mean value be of the form

$$E x_n = m \mu_n$$

where

$$\mu_n = \int_{-\pi}^{\pi} e^{in\lambda} d\mu(\lambda).$$

The function  $\mu(\lambda)$  shall be of bounded variation in  $(-\pi, \pi)$  but not necessarily real-valued. We want to form a linear unbiased estimate

$$m_n^* = m^*(x_{-n}, \dots, x_n)$$

of  $m$  of minimum variance, and ask under what conditions this minimum variance tends to zero as  $n$  tends to infinity, or, in other words, when there is a *consistent* estimate of  $m$ .

For a fixed value of  $n$  we have to solve

$$\int_{-\pi}^{\pi} |P_n(e^{i\lambda})|^2 dF(\lambda) = \min$$

where the polynomial

$$P_n(e^{i\lambda}) = \sum_{-n}^n c_\nu e^{i\nu\lambda}$$

has to satisfy

$$1 = \sum_{-n}^n c_\nu \mu_\nu = \int_{-\pi}^{\pi} P_n(e^{i\lambda}) d\mu(\lambda).$$

Schwarz' inequality gives

$$\begin{aligned} 1 \leq \left[ \int_{-\pi}^{\pi} |P_n(e^{i\lambda})| |d\mu(\lambda)| \right]^2 &\leq \left[ \sum_{\nu} |P_n(e^{i\lambda_{\nu}^{(m)}})| |\Delta_{\nu}\mu| \right]^2 + \varepsilon \leq \\ &\leq \sum_{\nu} |P_n(e^{i\lambda_{\nu}^{(m)}})|^2 \Delta_{\nu} F \sum_{\nu} \frac{|\Delta_{\nu}\mu|^2}{\Delta_{\nu} F} + \varepsilon \end{aligned}$$

where  $\varepsilon > 0$  can be made arbitrarily small if the division  $-\pi < \lambda_1^{(m)} < \lambda_2^{(m)} < \dots < \lambda_m^{(m)} < \pi$  is made sufficiently fine. But the quantity

$$K_m = \sum_{\nu} \frac{|\Delta_{\nu}\mu|^2}{\Delta_{\nu} F}$$

is known (see e.g. [14]) to be non-decreasing when the division is made finer. The limit is called a *Hellinger* integral and is denoted

$$\lim_{m \rightarrow \infty} K_m = K = \int_{-\pi}^{\pi} \frac{|d\mu(\lambda)|^2}{dF(\lambda)} \leq \infty.$$

Hence

$$D^2 m^* \geq \frac{1}{K}.$$

Actually  $\frac{1}{K}$  is the *infimum* when  $n$  tends to infinity. To show this we consider the stepfunction

$$\varphi^{(m)}(\lambda) = K_m^{-1} \frac{\overline{\Delta_i \mu}}{\Delta_i F} \quad \text{for } \lambda_i^{(m)} < \lambda \leq \lambda_{i+1}^{(m)}.$$

This can be uniformly approximated by the aid of a trigonometric polynomial of sufficiently high order and

$$\left. \begin{aligned} \int_{-\pi}^{\pi} |\varphi^{(m)}(\lambda)|^2 dF(\lambda) &= K_m^{-1} \\ \int_{-\pi}^{\pi} \varphi^{(m)}(\lambda) d\mu(\lambda) &= 1 \end{aligned} \right\}$$

Letting  $m$  tend to infinity we see that

$$D_{\min}^2 m^* = \frac{1}{\int_{-\pi}^{\pi} \frac{|d\mu(\lambda)|^2}{dF(\lambda)}}. \tag{21}$$

Hence a necessary and sufficient condition for the existence of a consistent estimate of  $m$  is the divergence of the Hellinger integral of  $\mu$  with respect to  $F$ . This happens for example if  $\mu(\lambda)$  has a saltus at a continuity point of  $F(\lambda)$ . We see the relation to what has been said above.

**Remark 1.** The problem has been studied in [2] p. 234 (although the results were stated for the continuous case they apply after simple modifications to the discrete case also). For a finite interval of observation we found

$$D_{\min}^2 m_n^* = \frac{1}{\sum_1^{n+1} \frac{|a_\nu|^2}{\lambda_\nu}}. \tag{22}$$

Here  $\lambda_\nu$  are the eigen-values of the covariance matrix and  $a_\nu$  the "Fourier"-coefficients of  $m_n$  developed in terms of the eigen-vectors of the matrix. For  $n = \infty$  the differential solutions will be  $\{e^{i\nu\lambda}\}$  and the coefficients with respect to these will be  $d\mu(\lambda)$ . Previously in 2.6 we have seen that  $\lambda_\nu$  coincide asymptotically with the equidistant ordinates of the spectral intensity. (21) is thus in an intuitive way seen to be a limit of (22).



**Remark 2.** Take  $\mu(\lambda) = \frac{\lambda}{2\pi}$  and  $F(\lambda)$  absolutely continuous with the derivative  $f(\lambda)$ . Then the problem studied reduces to minimizing

$$E \left| x_0 - \sum_{\nu=0}^{\infty} c_{\nu} x_{\nu} \right|^2$$

which is a problem of interpolation which has been solved by KOLMOGOROFF [8] giving as the *infimum* of the error the quantity

$$\frac{4\pi^2}{\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}}$$

which can also be obtained as a special case of (21).

REFERENCES. 1. **Cramér, H.:** Mathematical methods of statistics, Princeton Univ. Press, Princeton 1946. — 2. **Grenander, U.:** Stochastic processes and statistical inference, Arkiv för Matematik, 1950. — 3. —: On empirical spectral analysis of stochastic processes. Arkiv för Matematik, 1951. — 4. **Hanner, O.:** Deterministic and non-deterministic stationary stochastic processes, Arkiv för Matematik, 1949. — 5. **Hardy, G. H., Littlewood, J. E., Pólya, G.:** Inequalities, Cambridge 1934. — 6. **Kac, M., and Siegert, A. J. F.:** An explicit rerepresentation of a stationary Gaussian process, Ann. of Math. Stat., 1947. — 7. **Karhunen, K.:** Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Ann. Ac. Sci. Fennicae, 1947. — 8. **Kolmogoroff, A. H.:** Stationary sequences in Hilbert space (Russian), Bull. of the Univ. of Moscow, 1941. — 9. **Szegő, G.:** Beiträge zur Theorie der Toeplitzschen Formen, Math. Zeitschrift, 1920. — 10. —: Beiträge zur Theorie der Toeplitzschen Formen, (Fortsetzung), Math. Zeitschrift, 1921. — 11. —: Orthogonal polynomials, American Math. Soc. Colloquium publications, Vol. XXIII, 1939. — 12. **Whittle, P.:** Hypothesis testing in time series analysis, Uppsala 1951. — 13. **Wiener, N.:** Extrapolation, interpolation and smoothing of stationary time series, New York 1949. — 14. **Wintner, A.:** Spektraltheorie der unendlichen Matrizen, Leipzig 1929. — 15. **Wold, H.:** A study in the analysis of stationary time series, Uppsala 1938.

Tryckt den 25 september 1951

Uppsala 1951. Almqvist & Wiksells Boktryckeri AB