ARKIV FÖR MATEMATIK Band 4 nr 6

Read 11 March 1959

On the linear prediction problem for certain stochastic processes

By HARALD CRAMÉR

1. Consider an infinite sequence of complex-valued random variables

$$\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots$$

with finite second order mean values. For the sake of simplicity, we shall assume throughout that all first order mean values of the x_n reduce to zero, while there is at least one x_n having a variance different from zero:

$$E\,x_n = 0 \text{ for all } n,$$

$$E\,|x_n|^2 > 0 \text{ for some } n.$$

The *covariance function* of the x_n sequence is

$$R(m, n) = E(x_m \overline{x_n}).$$

We may interpret x_n as a measure of the state of some observed variable system at the time point nt, where t is a given quantity. The sequence of the x_n, with $n = \cdots, -1, 0, 1, \ldots$, will then represent the temporal development of this system, and will constitute a *stochastic process with discrete time*. In the sequel, we shall always take $t = 1$, so that the subscript n may be directly regarded as measuring time.

The *prediction problem* for a process of this kind is the problem of predicting the state of the process at some future time point, when its past development is assumed to be more or less known. In this paper we shall only be concerned with *linear least squares prediction*. Thus we shall want to find the "best possible" prediction of a certain x_n by means of linear operations acting on certain variables belonging to the past of the process, interpreting the "best possible" in the sense of minimizing the mean value of the squared error of prediction.

Consider the expression

$$\operatorname*{Min}_{c_0, \ldots, c_q} E\,|x_n - c_0 x_{n-p} - c_1 x_{n-p-1} - \cdots - c_q x_{n-p-q}|^2 = s^2_{npq} \geqq 0,$$

where n, p and q are fixed integers with $p > 0$, $q > 0$, while the minimum has to be taken for all complex quantities $c_0, \ldots, c_q$. Then s_{npq} will be the least possible error of prediction, when x_n has to be linearly predicted in terms of x_{n-p}, x_{n-p-1},

$\ldots, x_{n-p-q}$. Obviously s_{npq} will never increase when q increases, while n and p remain fixed. The limit

$$\lim_{q\to\infty} s_{npq} = \sigma_{np} \geqq 0$$

will thus always exist, and will be called the *prediction error* for x_n, when prediction is based on all the variables of the process up to and including x_{n-p}. We easily find that

$$0 \leqq \sigma_{n1} \leqq \sigma_{n2} \leqq \cdots.$$

The same thing may be expressed a little differently, if we look at the question from the point of view of Hilbert space geometry. Consider the Hilbert space H_{n-p} spanned by all the variables $x_{n-p}, x_{n-p-1}, \ldots$. The elements of this space are random variables, which are either finite linear combinations of $x_{n-p}, x_{n-p-1}, \ldots$, or limits in the mean of sequences of such combinations. The inner product and the norm are defined by the usual expressions

$$(y, z) = E(y\bar{z}), \qquad \|y\| = (E|y^2|)^{\frac{1}{2}}.$$

By $P_{n-p}(x_n)$ we denote the projection of x_n on H_{n-p}. Then $z = P_{n-p}(x_n)$ is the uniquely determined element of H_{n-p} which minimizes the distance $\|x_n - z\|$. We shall call $P_{n-p}(x_n)$ the *best possible linear prediction* of x_n in terms of $x_{n-p}, x_{n-p-1}, \ldots$. The corresponding error of prediction will be

$$\sigma_{np} = \|x_n - P_{n-p}(x_n)\|.$$

If $\sigma_{np} = 0$ for all n and p, exact linear prediction is always possible. In this case every x_n can be exactly represented in terms of variables belonging to an arbitrarily remote past of the process. A process of this kind will be called a *deterministic* process.

On the other hand, every process such that $\sigma_{np} > 0$ for at least one pair of values of n and p, will be called *non-deterministic*. Every non-deterministic process may be represented as the sum of a deterministic component and a linear combination of certain *innovations*, which represent the "new" random impulses entering into the process at certain moments. In fact, it can be shown (cf. Cramér, 2) that for every non-deterministic process, there exists a uniquely determined, finite or infinite sequence of integers

$$\cdots < r_{-1} < r_0 < r_1 < \cdots$$

such that

$$x_n = \sum_{r_k \leqq n} c_{nr_k} \xi_{r_k} + y_n. \tag{1}$$

Here y_n is the deterministic component of the process, while the ξ_{r_k} are random variables such that

$$E\,\xi_{r_k} = 0, \quad E(\xi_{rj}\bar{\xi}_{rk}) = \delta_{jk},$$

$$E(x_n\,\bar{\xi}_{rk}) = c_{nr_k}, \quad E(y_n\,\bar{\xi}_{rk}) = 0.$$

Further $\sum_{r_k \leqq n} |c_{nr_k}|^2$ is convergent, so that the series in the expression for x_n con-

verges in the mean. If n is a member of the r_k sequence, the quantity c_{nn} is real and positive. Finally, the prediction error σ_{np} is given by the expression

$$\sigma_{np}^2 = \sum_{n-p<r_k\leqq n} |c_{nr_k}|^2.$$

It follows that $\sigma_{n1} > 0$ when and only when n is a member of the r_k sequence. Thus it will be seen that every r_k is a *point of indetermination*, where the process receives an *innovation* proportional to ξ_{r_k}. On the other hand, when n is different from all r_k, we have $\sigma_{n1} = 0$, so that x_n can be exactly predicted in terms of the preceding variables $x_{n-1}, x_{n-2}, \ldots$, and there is no innovation corresponding to the time point n. Thus in the expression (1) of x_n, the non-deterministic component is a linear combination of the innovations received by the process in all its points of indetermination preceding or coinciding with the time point n.

2. The linear prediction problem has been thoroughly studied for the important class of *stationary* processes, which are characterized by the fact that the covariance function $R(m, n) = E(x_m \overline{x_n})$ only depends on the time difference $m - n$. (We are not here concerned with the so called strictly stationary processes, which satisfy more stringent conditions.) For this class of processes, the sequence r_k considered in the preceding section contains every integer n, and the representation (1) reduces to a decomposition theorem due to H. Wold, 1. With respect to the general theory of stationary processes, and in particular the prediction problem, we may refer e.g. to the books by Wiener, 1, and Doob, 1.

One of the most important properties of the class of stationary processes is that they admit a *spectral representation* by means of a stochastic integral of Fourier type. Loève (1) has introduced a more general class of processes, which he calls *harmonizable* processes, and which possess spectral representations of a similar kind. In this paper, we shall consider the prediction problem for a group of harmonizable processes satisfying certain regularity conditions.

Consider a stochastic process such that x_n is given by the stochastic integral

$$x_n = \int_0^{2\pi} e^{inu}\, dz(u), \tag{2}$$

where $z(u)$ denotes, for every u in the interval $0 \leqq u \leqq 2\pi$, a random variable such that

$$E\, z(u) = 0, \quad E(z(u)\, \overline{z(v)}) = F(u, v),$$

where $F(u, v)$—in general complex-valued—is of bounded variation over the square C: $0 \leqq u, v \leqq 2\pi$. Obviously

$$F(v, u) = \overline{F(u, v)},$$

while $F(u, u)$ is real and non-negative. Under these conditions, the stochastic integral (2) can be defined as a limit in the mean of certain Riemann sums, and the variable x_n determines a harmonizable stochastic process. (Cf. also Cramér 1.)

In the sequel, we shall assume that the function $F(u, v)$ satisfies two additional conditions, which we denote by (A) and (B). Thus we shall assume:

$$F(u, v) = \int_0^u \int_0^v f(s, t)\, ds\, dt, \tag{3}$$

where

(*A*) $f(s,t)$ belongs to L^2 over the square C.

(*B*) $f(s,t)$ is bounded in the vicinity of the diagonal $s=t$ of C, so that there exist positive constants h and M such that $|f(s,t)|<M$ for $|s-t|<h$.

The covariance function of the x_n process is then given by

$$R(m,n)=E(x_m\overline{x_n})=\int_0^{2\pi}\int_0^{2\pi}e^{i(mu-nv)}\,dF(u,v)$$
$$=\int_0^{2\pi}\int_0^{2\pi}e^{i(mu-nv)}f(u,v)\,du\,dv. \tag{4}$$

More generally, for any $g(u)$ and $h(u)$ belonging to L^2 over $(0,2\pi)$, the random variables

$$\xi=\int_0^{2\pi}g(u)\,dz(u),\quad \eta=\int_0^{2\pi}h(u)\,dz(u)$$

are well defined, and we have

$$E(\xi\bar{\eta})=\int_0^{2\pi}\int_0^{2\pi}g(u)\,\overline{h(v)}\,f(u,v)\,du\,dv. \tag{5}$$

The function $F(u,v)$ is called the *spectral function* of the x_n process, and is said to define the *spectral distribution* of the process, which is a distribution of complex-valued "mass" over the square C, such that every surface element $du\,dv$ carries the mass $f(u,v)\,du\,dv$. The function $f(u,v)$ is the *spectral density* of the process, while $z(u)$ defines the corresponding *spectral process*.

I have elsewhere (Cramér, 2) given a sufficient condition that a harmonizable process will be deterministic. In the present note, I shall be concerned with an x_n process as defined by (2), and such that the corresponding spectral density $f(u,v)$ satisfies the conditions (*A*) and (*B*). It is proposed to find, for this process, the conditions under which an arbitrarily given sequence of integers r_k will constitute the complete set of points of indetermination of the process.

3. Suppose that we are given an x_n process as defined by (2), with a spectral function $F(u,v)$ satisfying (3), and a spectral density $f(u,v)$ satisfying the conditions (*A*) and (*B*). Let there further be given an increasing sequence of integers:

$$\cdots<r_{-1}<r_0<r_1<\cdots,$$

which may be finite or infinite, in one or both directions.

In order that the given r_k sequence will constitute the complete set of points of indetermination of the given x_n process, the following condition is necessary and sufficient. The spectral function $F(u,v)$ should admit a development of the form:

$$F(u,v)=\sum_{r_k}\varphi_k(u)\,\overline{\varphi_k(v)}+G(u,v), \tag{6}$$

where $G(u,v)$ is the spectral function of a deterministic harmonizable process, while the series in the second member converges absolutely for all u and v, and we have

$$\varphi_k(u) = \sum_{n=r_k}^{\infty} \alpha_{nr_k} e^{-inu} + \beta_k u + \gamma_k, \tag{7}$$

the one-sided trigonometric series occurring here being absolutely convergent for all u, *while the* α, β *and* γ *are constants such that* $\alpha_{r_k r_k} \neq 0$ *for every* $r_k \neq 0$, $\beta_k = 0$ *for* $r_k > 0$, *and the series*

$$\sum_{r_k \leqq n} |\alpha_{nr_k}|^2,$$

extended over all $r_k \leqq n$, *converges for every fixed* n.

We shall first show that the condition is necessary. Since by hypothesis r_k is a point of indetermination of the given x_n process, the innovation

$$c_{r_k r_k} \xi_{r_k} = x_{r_k} - P_{r_k - 1}(x_{r_k})$$

is not identically zero. As in the preceding section, we may take $c_{r_k r_k}$ real and positive, and such that $E|\xi_{r_k}|^2 = 1$. As before we write

$$c_{nr_k} = E(x_n \bar{\xi}_{r_k}),$$

observing that we have

$$c_{nr_k} = 0 \text{ for } n < r_k, \tag{8}$$

as the innovation associated with ξ_{r_k} does not enter into the process until the time point r_k, and is thus uncorrelated with every x_n preceding this time point.

We now define a random variable w by writing

$$w = \int_0^{2\pi} u \, dz(u),$$

where $z(u)$ is the spectral process appearing in (2). The stochastic integral is defined in the way indicated in the preceding section. We further write

$$d_k = E(w \bar{\xi}_{r_k}) \tag{9}$$

and

$$\varphi_k(u) = -\frac{1}{2\pi i} \sum_{n=r_k}^{\infty}{}' \frac{c_{nr_k}}{n} e^{-inu} + \frac{1}{2\pi} c_{0r_k}(u + \pi) - \frac{1}{2\pi} d_k, \tag{10}$$

where the accent on the summation sign indicates that, if the value $n = 0$ falls between the limits of summation, the corresponding term should be omitted. Clearly this is an expression of the form postulated in (7). In order to show that the trigonometric series appearing here is absolutely convergent, we observe that

$$|c_{nr_k}|^2 \leqq E|x_n|^2 E|\xi_{r_k}|^2 = E|x_n|^2 = \int_0^{2\pi}\int_0^{2\pi} e^{in(u-v)} f(u,v)\, du\, dv.$$

Thus by condition (A) the quantities $|c_{nr_k}|^2$ are, for a fixed r_k, the Fourier coefficients of a function in L^2, so that the series $\sum_n |c_{nr_k}|^4$ is convergent. Hence by Hölder's inequality it is easily shown that $\sum_n{}' \frac{c_{nr_k}}{n}$ is absolutely convergent. It has

already been remarked in connection with (1) that the series $\sum_{r_k \leqq n} |c_{nr_k}|^2$ is convergent for every n. We finally observe that, in consequence of (8), the lower limit of summation in (10) may be replaced by $-\infty$, since the new terms thus introduced all reduce to zero.

We now consider the expression

$$G_K(u,v) = F(u,v) - \sum_{k=-K}^{K} \varphi_k(u)\,\overline{\varphi_k(v)}, \tag{11}$$

with appropriate modification in case the r_k sequence is finite in either direction. We shall first show that $G_K(u,v)$ is, for every K, the covariance function of some random variable $z_K(u)$ defined for every u in $(0, 2\pi)$, so that

$$G_K(u,v) = E(z_K(u)\,\overline{z_K(v)})$$

for all u and v in $(0, 2\pi)$. In order to show that $G_K(u,v)$ is a covariance, it is sufficient to show (cf. Loève, 1) that

$$\int_0^{2\pi}\int_0^{2\pi} q(u)\,\overline{q(v)}\,G_K(u,v)\,du\,dv \geqq 0 \tag{12}$$

for any continuous $q(u)$. Obviously it is even sufficient to show that (12) holds for any trigonometric polynomial $q(u)$. Taking

$$q(u) = \sum_{n=A}^{B} \lambda_n e^{inu},$$

$$Q(u) = \sum_{n=A}^{B}{}' \frac{\lambda_n}{in}(e^{inu}-1) + \lambda_0(u-2\pi),$$

we have

$$Q'(u) = q(u), \quad Q(2\pi) = 0.$$

By (3) we have

$$F(u,0) = F(0,v) = 0,$$

so that we obtain by partial integration, using (4), (5) and (9),

$$\int_0^{2\pi}\int_0^{2\pi} q(u)\,\overline{q(v)}\,F(u,v)\,du\,dv = \int_0^{2\pi}\int_0^{2\pi} Q(u)\,\overline{Q(v)}\,f(u,v)\,du\,dv =$$

$$= E\left|\sum_{A}^{B}{}' \frac{\lambda_n}{in}(x_n - x_0) + \lambda_0(w - 2\pi x_0)\right|^2. \tag{13}$$

On the other hand

$$\int_0^{2\pi}\int_0^{2\pi} q(u)\,\overline{q(v)}\,\varphi_k(u)\,\overline{\varphi_k(v)}\,du\,dv = \left|\int_0^{2\pi} q(u)\,\varphi_k(u)\,du\right|^2, \tag{14}$$

and by some simple calculation we find

$$\int_0^{2\pi} q(u)\,\varphi_k(u)\,du = -\sum_{n=A}^{B}{}' \frac{\lambda_n}{in}(c_{nr_k} - c_{0r_k}) - \lambda_0(d_k - 2\pi c_{0r_k}) =$$

$$= -E\left\{\left(\sum_{A}^{B}{}' \frac{\lambda_n}{in}(x_n - x_0) + \lambda_0(w - 2\pi x_0)\right)\overline{\xi_{r_k}}\right\}. \tag{15}$$

Writing

$$X = \sum_{A}^{B}{}' \frac{\lambda_n}{in}(x_n - x_0) + \lambda_0(w - 2\pi x_0),$$

we obtain from (11), (13), (14) and (15)

$$\int_0^{2\pi}\int_0^{2\pi} q(u)\,\overline{q(v)}\,G_K(u,v)\,du\,dv = E|X|^2 - \sum_k |E(X\xi_{rk})|^2.$$

The ξ_{rk} being orthogonal random variables, (12) now follows directly from Bessel's inequality.

Thus $G_K(u,v)$ as defined by (11) is always a covariance function. It follows that $G_K(u,u) \geqq 0$, so that, allowing K to tend to infinity, the sum over all k

$$\sum_k |\varphi_k(u)|^2$$

is convergent, and consequently by the Schwarz inequality

$$\sum_k \varphi_k(u)\,\overline{\varphi_k(v)}$$

is absolutely convergent. Thus $G_K(u,v)$ tends to a limit $G(u,v)$ as $K \to \infty$.

The limit of a sequence of covariance functions being itself a covariance function (cf. Loève 1), we have now proved that $F(u,v)$ admits a development of the form (6), where $G(u,v)$ is a covariance function, while $\varphi_k(u)$ has the form (7), the stated convergence conditions being satisfied. It remains to show that $G(u,v)$ is the spectral function of a deterministic harmonizable process.

In order that a covariance function $G(u,v)$ defined in the square C: $0 \leqq u, v \leqq 2\pi$, should be the spectral function of some harmonizable process, it is necessary and sufficient that $G(u,v)$ should be of bounded variation over C. We shall first show that this property holds here.

Since $G(u,v)$ is a covariance function, there exists for every u in $(0, 2\pi)$ a random variable $Z(u)$ such that

$$G(u,v) = F(u,v) - \sum_k \varphi_k(u)\,\overline{\varphi_k(v)} = E(Z(u)\,\overline{Z(v)}). \tag{16}$$

For an arbitrary sub-interval $(u, u+h)$ of $(0, 2\pi)$ we then have, taking differences in the obvious way,

$$\Delta_2 G(u,u) = \Delta_2 F(u,u) - \sum_k |\Delta\varphi_k(u)|^2 = E|\Delta Z(u)|^2,$$

and thus

$$E|\Delta Z(u)|^2 \leqq \Delta_2 F(u,u). \tag{17}$$

Let now

$$0 = u_0 < u_1 < \cdots < u_m = 2\pi,$$
$$0 = v_0 < v_1 < \cdots < v_n = 2\pi,$$

be two arbitrary sub-divisions of the interval $(0, 2\pi)$. All differences occurring in the sequel will be understood to be related in the obvious way to the sub-intervals in these divisions.

From (16) we obtain

$$\Delta_2 G(u_r, v_s) = E(\Delta Z(u_r)\,\overline{\Delta Z(v_s)}),$$

and further by the Schwarz inequality, using (17),

$$\begin{aligned}|\Delta_2 G(u_r, v_s)| &\leqq E\{|\Delta Z(u_r)| \cdot |\Delta Z(v_s)|\} \\ &\leqq \sqrt{E|\Delta Z(u_r)|^2 \cdot E|\Delta Z(v_s)|^2} \\ &\leqq \sqrt{\Delta_2 F(u_r, u_r) \cdot \Delta_2 F(v_s, v_s)}.\end{aligned}$$

Summing over all the sub-intervals in both variables, we thus obtain by means of the condition (B), assuming that all sub-intervals are sufficiently small,

$$\sum_{r,s} |\Delta_2 G(u_r, v_s,)| \leqq \sum_r \sqrt{\Delta_2 F(u_r, u_r)} \cdot \sum_s \sqrt{\Delta_2 F(v_s, v_s)} \leqq 4\pi^2 M,$$

which shows that $G(u, v)$ is of bounded variation over C.

Comparing now the general representation (1) of a non-deterministic x_n process, and the development (6) of the spectral function $F(u, v)$ in the present case, it is easily verified that the term $\sum \varphi_k(u)\overline{\varphi_k(v)}$ in (6), where $\varphi_k(u)$ is determined by (10), is the spectral function of the sum $\sum c_{nr_k}\xi_{r_k}$ in (1), while the remaining term $G(u, v)$ is the spectral function of the deterministic component y_n of the x_n process. We have thus completed the proof that the given condition is necessary.

The proof that the condition is also sufficient is now very simple. We first observe that the predictionary properties (prediction being always understood in the sense of linear least squares prediction) of a stochastic process are entirely determined by the covariance function of the process. Thus if we are given an x_n process with the covariance function $R(m, n)$, and if we can show that $R(m, n)$ is the covariance function of *some* stochastic process having the given sequence of the r_k for its points of indetermination, it follows that the given x_n process will have precisely the same points of indetermination. For a harmonizable process, the covariance function is uniquely determined by the spectral function $F(u, v)$, so that the same remark applies here to $F(u, v)$.

Suppose now that we are given an r_k sequence, and an x_n process with a spectral function $F(u, v)$ satisfying all our conditions. In particular, $F(u, v)$ will then be given by the development (6), where $\varphi_k(u)$ is given by (7). According to the remark just made, we shall then only have to show that $F(u, v)$ is the spectral function of some stochastic process having the given r_k for its points of indetermination.

Consider a stochastic process x_n^* represented in the form (1), where we take

$$\begin{aligned}c_{nr_k} &= -2n\pi i\alpha_{nr_k} \text{ for } r_k \leqq n,\ n \neq 0, \\ c_{0r_k} &= 2\pi\beta_k \qquad\qquad \text{for } r_k \leqq 0,\end{aligned}$$

while the ξ_{r_k} are orthogonal random variables, and also orthogonal to the y_n, which are the variables of a deterministic process with the spectral function $G(u, v)$ appearing in (6). Then it will be immediately seen that the x_n^* process has the given r_k sequence for its points of indetermination, so that the proof is hereby completed.

We finally remark that the development of the difference $F(u,v)-G(u,v)$, which follows from (6), is formally analogous to the well-known development of the kernel $F(u,v)-G(u,v)$ in terms of its characteristic functions. However, it is easily shown by means of examples that the $\varphi_k(u)$ appearing in (6) are not necessarily identical with the characteristic functions of the corresponding kernel.

REFERENCES

CRAMÉR, H., 1. A contribution to the theory of stochastic processes. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley 1951.

2. Remarques sur le problème de prédiction pour certaines classes de processus stochastiques. To appear in the Publications de l'Institut de Statistique de l'Université de Paris.

DOOB, J. L., Stochastic processes. New York (Wiley) 1953.

LOÈVE, M., Fonctions aléatoires du second ordre. Supplément to P. LÉVY, Processus stochastiques et mouvement brownien. Paris. (Gauthier-Villars) 1948.

WIENER, N., Extrapolation, interpolation and smoothing of stationary time series. New York (Wiley) 1949.

WOLD, H., A study in the analysis of stationary time series. Thesis, Stockholm. Uppsala 1938.

Tryckt den 28 augusti 1959

Uppsala 1959. Almqvist & Wiksells Boktryckeri AB