On the Prediction of Non-Stationary Processes

by N. A. Abdrabbo[a], M. B. Priestley[a]

a. University of Manchester
[Received August 1966. Revised January 1967]

Abstract

We consider the problem of linear least-squares prediction for a class of nonstationary processes which possess *evolutionary spectral representations.* It is shown that, under certain conditions, such processes admit moving-average representations in terms of time-dependent coefficients. This feature enables us to develop a close analogue of the Wiener-Kolmogorov approach to the corresponding problem for stationary processes.

Table of contents

1 Introduction

The Wiener-Kolmogorov "linear least-squares" approach to the prediction and filtering of stationary processes is now well established, and may, in principle, be applied to any practical problem involving stationary processes whose second-order properties are known. Accounts of this theory have been written at various levels of mathematical rigour (see, for example, [Doob1953, Bartlett1955, GrenanderRosenblatt1957, Yaglom1962, Whittle1965]), but so far the corresponding problem for non-stationary processes has received little attention.

There have been a few isolated attempts to deal with this topic, but in the main the approaches have been either too general or too restricted to be useful in practical applications. For example, Parzen [Parzen1961] has solved the non-stationary prediction problem in principle, but his approach is a somewhat abstract one, and his "solution" for the optimum predictor is expressed as a certain inner-product in a Hilbert space. Cramér [Cramer1961a, Cramer1961b] considered the same problem, and obtained some interesting results in the form of "existence theorems", but did not present a method for determining the explicit form of a predictor in terms of the observed variables. Similar remarks apply to the work of Davis [Davis1952]. On the other hand, there have appeared several papers written from an engineering standpoint (see, for example, [Booton1952, Zadeh1953, Bendat1956]), but in most cases the "general solution" stops with the construction of the well-known integral equation (cf. (?)) involving the covariance function of the process and the unknown "coefficients" of the optimal predictor, the solution of which is obtainable only when the process obeys some very simple model.

On a different level, Kalman [Kalman1960] has studied multivariate non-stationary processes corresponding to dynamical systems which are governed by known linear differential equations, and has established some basic results regarding the analytic structure of optimal predictors for such models. However, from the point of view of practical application, it would appear that the most useful results so far obtained are due to Whittle [Whittle1965], who considered non-stationary processes generated by autoregressive models with time-dependent coefficients, and obtained explicit recursive relations for the optimal predictors. In fact, some of our results for these particular models correspond very closely to those obtained by Whittle.

The success of classical prediction theory for stationary processes is due essentially to the fact that such processes admit a spectral representation in terms of an orthogonal process. This feature not only simplifies the solution of the prediction problem, but also enables one to treat a general class of stationary processes by means of a "canonical" representation, so that the discussion need not be restricted to particular models, such as the autoregressive, moving-average, etc. Hitherto, the lack of a similar spectral representation for a general class of non-stationary processes has no doubt been one of the major stumbling blocks in attempts to generalize the classical theory. However, it turns out that the recently developed theory of evolutionary spectral representations [Priestley1965] provides an ideal framework for the formulation and solution of non-stationary prediction problems. In fact, by using evolutionary spectral representations one obtains a prediction theory which is almost an exact parallel of the Wiener-Kolmogorov theory.

The basic idea underlying this approach is the introduction of the evolutionary (i.e., time-dependent) spectrum of a nonstationary process, whose form completely determines the values of the "coefficients" of the optimal linear predictor. This means that even if we are presented with observations from a process whose structure is completely unknown we may still usefully apply this prediction theory by first estimating the evolutionary spectrum (methods for estimating evolutionary spectra have been discussed by [Priestley1965, Priestley1966]). Of course, this estimation procedure introduces further complications

which will not be discussed here (see also Section 6). In this paper we assume throughout that the second-order properties of the process are known a priori. We should point out, further, that in this account we are not aiming at complete mathematical rigour: rather we aim to show how our approach may be used to obtain explicit results for non-stationary processes.

Before discussing the prediction problem, we summarize briefly relevant parts of the theory of evolutionary spectra.

2 Evolutionary Spectral Representations

We consider a class of continuous parameter processes, $\{X(t)\}$, with $E\{X(t)\} = 0$, $E\{X^2(t)\} < \infty$ for all t, for which there exists a family $\mathscr{F}$ of functions $\{\phi_t(\omega)\}$ (defined on the real line and indexed by t) and a measure μ on the real line such that $\{X(t)\}$ admits a spectral representation of the form

$$X(t) = \int_{-\infty}^{\infty} \phi_t(\omega)\, d\, Z(\omega) \tag{1}$$

where $\{Z(\omega)\}$ is an orthogonal process with $E\, |d\, Z(\omega)|^2 = \mu\, (d\, \omega)$ (see also [GrangerHatanaka1964]).

If there exists a family $\mathscr{F}$ for which each ϕ_t may be written in the form

$$\phi_t(\omega) = e^{i\omega t}\, A_t(\omega) \tag{2}$$

where, for each fixed ω, the generalized Fourier transform of $A_t(\omega)$ (considered as a function of t) has an absolute maximum at the origin, $X(t)$ is termed an *oscillatory process* (cf. [Priestley1965]), and the evolutionary spectrum at time t with respect to the family $\mathscr{F}$ is defined by

$$d\, F_t(\omega) = |A_t(\omega)|^2\, \mu\, (d\, \omega), \quad -\infty < \omega < \infty \tag{3}$$

For discrete parameter oscillatory processes (i.e., processes defined only for $t = 0, \pm 1, \pm 2, \dots$), we have a corresponding evolutionary spectral representation,

$$X(t) = \int_{-\pi}^{\pi} e^{i\omega t}\, A_t(\omega)\, d\, Z(\omega) \tag{4}$$

and the evolutionary spectrum at time t with respect to $\mathscr{F}$ is defined by

$$d\, F_t(\omega) = |A_t(\omega)|^2\, \mu\, (d\, \omega), \quad -\pi \le \omega \le \pi \tag{5}$$

3

When the measure μ is absolutely continuous (with respect to Lebesgue measure), the derivative $f_t(\omega) = F_t'(\omega)$ exists for all ω, and is termed the *evolutionary spectral density function* at time t.

In both the continuous and discrete cases we may normalize the functions $A_t(\omega)$ so that $A_0(\omega) = 1$ for all ω. With this convention, $\mu(d\omega) \equiv d F_0(\omega)$, the evolutionary spectrum at time zero.

3 Moving-average Representations for Oscillatory Processes

As in the prediction theory of stationary processes, we begin by constructing a one-sided moving-average representation for a class of oscillatory processes, the distinction with the stationary case being that the coefficients in the moving-average scheme are now time-dependent.

3.1 Continuous-parameter Processes

As in the case of stationary processes (see [Whittle1963]), the results for discrete-parameter processes can readily be adapted to the continuous case. The measure μ is again assumed to be absolutely continuous (with respect to Lebesgue measure) and, in place of C1, we assume that

$$\int_{-\infty}^{\infty} \frac{\log f(\omega)}{1+\omega^2} \, d\omega > -\infty \tag{6}$$

(Condition C1*)

Then there exists a function $\psi(\omega)$ such that $|\psi(\omega)|^2 = f(\omega)$, $\psi(z)$ having no poles or zeros in the lower half-plane. The function $\psi(\omega)$ may now be written as a one-sided Fourier integral. Corresponding to (?), we assume now that, for each t,

$$A_t(\omega) = \int_0^{\infty} e^{-i\omega u} \, g_t(u) \, d u \tag{7}$$

A necessary condition being

$$\int_{-\infty}^{\infty} \frac{\log |A_t(\omega)|^2}{1+\omega^2} \, d\omega > -\infty \tag{8}$$

for all t (Condition C3*).

It then follows that we may write $X(t)$ in the form

$$X(t) = \int_{-\infty}^{\infty} e^{it\omega} \alpha_t(\omega) \, d z(\omega) \tag{9}$$

where $z(\omega)$ is an orthogonal process on $(-\infty, \infty)$, with $E |d z(\omega)|^2 = d\omega$, and $\alpha_t(\omega)$ has a one-sided Fourier integral representation of the form

$$\alpha_t(\omega) = \int_0^\infty e^{-i\omega u}\, h_t(u)\, d u \tag{10}$$

and $|\alpha_t(\omega)|^2 = f_t(\omega)$.

A necessary condition for the validity of (10) is

$$\int_{-\infty}^\infty \frac{\log f_t(\omega)}{1 + \omega^2}\, d\omega > -\infty \tag{11}$$

for all t (Condition C4*).

Then, formally, we may write

$$X(t) = \int_0^\infty h_t(u)\, \xi\, (t - u)\, d u \tag{12}$$

where $\xi(t) = \int_{-\infty}^\infty e^{i\omega t}\, d z(\omega)$.

4   The Time-domain Approach

The basic problem of linear least-squares prediction may be stated as follows: we are given the observed values of the process over the semi-infinite interval $(-\infty, t)$ and wish to predict the value of $X\,(t + m)$ $(m > 0)$. The predictor, $\tilde{X}\,(t + m)$, is to be chosen as that linear combination of $\{X(s), s \leq t\}$ which is such that

$$M(m) = E\, \{\tilde{X}\,(t + m) - X\,(t + m)\}^2 \tag{13}$$

is minimized.

Assuming that the process has a one-sided moving average form, we may easily obtain an explicit expression for $\tilde{X}\,(t + m)$ in terms of the process $\{\xi(t)\}$.

4.1   Continuous-parameter Processes

The results for continuous-parameter processes are very similar. We are given $\{X(s), -\infty < s \leq t\}$ and wish to predict $X\,(t + m)$ by a linear combination of past values:

$$\tilde{X}\,(t + m) = \int_{-\infty}^t X(s)\, b(s)\, d s \tag{14}$$

5

or, equivalently,

$$\tilde{X}(t+m) = \int_{-\infty}^{t} \xi(u)\, a(u)\, d\,u \tag{15}$$

Assuming $X(t)$ has a moving-average representation of the form (12), it follows that

$$\tilde{X}(t+m) = \int_{-\infty}^{t} h_{t+m}(t+m-u)\, \xi(u)\, d\,u = \int_{m}^{\infty} h_{t+m}(u)\, \xi(t+m-u)\, d\,u \tag{16}$$

To express $\tilde{X}(t+m)$ in terms of $\{X(s)\}$, write

$$\xi(t) = \int_{0}^{\infty} k_t(v)\, X(t-v)\, d\,v \tag{17}$$

and substitute in (12), leading to the integral equations

$$\int_{0}^{v} h_t(u)\, k_{t-u}(v-u)\, d\,u = \delta(v) \tag{18}$$

and

$$\int_{0}^{u} k_t(v)\, h_{t-v}(u-v)\, d\,v = \delta(u) \tag{19}$$

where $\delta(\cdot)$ is the Dirac delta function.

5   The Frequency-domain Approach

We now consider an alternative approach based on a generalization of the Wiener-Hopf technique. We treat first the continuous-parameter case, and write the predictor as

$$\tilde{X}(t+m) = \int_{0}^{\infty} b_t(u)\, X(t-u)\, d\,u \tag{20}$$

or, formally,

$$\tilde{X}(t+m) = \int_{0}^{\infty} X(t-u)\, d\,\beta_t(u) \tag{21}$$

where $b_t(u)$ may depend on both t and m.

Using the evolutionary spectral representation, the predictor may be written

$$\tilde{X}(t+m) = \int_{-\infty}^{\infty} e^{i\omega t}\, B_t(\omega)\, d\,z(\omega) \tag{22}$$

where

$$B_t(\omega) = \int_0^\infty b_t(u)\, \alpha_{t-u}(\omega)\, e^{-iu\omega}\, du \tag{23}$$

and

$$X(t+m) = \int_{-\infty}^\infty e^{i\omega(t+m)}\, \alpha_{t+m}(\omega)\, dz(\omega) \tag{24}$$

The mean squared error is

$$M(m) = E\,|\tilde{X}(t+m) - X(t+m)|^2 = \int_{-\infty}^\infty |e^{i\omega m}\,\alpha_{t+m}(\omega) - B_t(\omega)|^2\, d\omega \tag{25}$$

Let

$$e^{i\omega m}\,\alpha_{t+m}(\omega) = C_t^{(1)}(\omega) + C_t^{(2)}(\omega) \tag{26}$$

where $C_t^{(1)}(\omega)$ is a "backward transform" and $C_t^{(2)}(\omega)$ is a "forward transform" (see [Bartlett1955], p. 201). Then

$$M(m) = \int_{-\infty}^\infty |C_t^{(1)}(\omega) - B_t(\omega)|^2 d\omega + \int_{-\infty}^\infty |C_t^{(2)}(\omega)|^2\, d\omega \tag{27}$$

and the minimum is attained when

$$B_t(\omega) = C_t^{(1)}(\omega) \tag{28}$$

with

$$M(m) = \int_{-\infty}^\infty |C_t^{(2)}(\omega)|^2\, d\omega \tag{29}$$

Taking Fourier transforms, for $v \geq 0$,

$$K_t(v) = \int_{-\infty}^\infty e^{i\theta(v+m)}\,\alpha_{t+m}(\theta)\, d\theta = h_{t+m}(v+m) \tag{30}$$

and

$$\int_0^v b_t(u)\, h_{t-u}(v-u)\, du = h_{t+m}(v+m) \tag{31}$$

for all t and $v \geq 0$.

If $h_{t-v}(0) \neq 0$ for all $v \in (0, \infty)$ and $\partial h_{t-u}(v-u)/\partial v = h'_{t-u,v}(v-u)$ exists and is continuous, then by differentiating both sides with respect to v,

$$h_{t-v}(0)\, b_t(v) + \int_0^v h'_{t-u,v}(v-u)\, b_t(u)\, d\,u = h'_{t+m,v}(v+m) \tag{32}$$

or

$$b_t(v) + \int_0^v \frac{h'_{t-u,v}(v-u)}{h_{t-v}(0)}\, b_t(u)\, d\,u = \frac{h'_{t+m,v}(v+m)}{h_{t-v}(0)} \tag{33}$$

The solution is (see [Tricomi1957], p. 10)

$$b_t(v) = \frac{h'_{t+m,v}(v+m)}{h_{t-v}(0)} + \int_0^v H_t^*(v,u,-1)\, \frac{h'_{t+m,v}(u+m)}{h_{t-u}(0)}\, d\,u \tag{34}$$

where the "resolvent kernel" $H_t^*(v,u,-1)$ is

$$H_t^*(v,u,-1) = -\sum_{p=0}^{\infty} (-1)^p\, K_{t,p+1}^*(u,v) \tag{35}$$

with the "iterated kernels"

$$K_{t,p+1}^*(u,v) = \int_u^v K_{t,1}^*(u,z)\, K_{t,p}^*(z,v)\, d\,z \quad (p=1,2,\dots) \tag{36}$$

and

$$K_{t,1}^*(u,v) = \frac{h'_{t-u,v}(v-u)}{h_{t-v}(0)} \tag{37}$$

6 Discussion

As mentioned in Section 1, in this paper we are concerned primarily with the problem of determining the optimum predictor on the assumption that the second-order properties of the process are fully known. However, we have shown that the predictor $\tilde{X}(t+m)$ is determined uniquely by the form of the evolutionary spectra $\{f_t(\omega)\}$, so that in cases where we have no precise knowledge of the underlying model we may attack the problem by first estimating the functions $\{f_t(\omega)\}$.

It will be seen from Section 5 that $\tilde{X}(t+m)$ depends on the form of $h_{t+m}(u)$, i.e., on the form of the evolutionary spectrum at the future time point $(t+m)$. In a purely empirical analysis we would not, of course, be able to estimate $f_{t+m}(\omega)$ directly from observations extending only up to time t, and we would have to assume that the spectra were changing with sufficient "smoothness" over time to enable the form of $f_{t+m}(\omega)$ to be inferred from the spectra which have been estimated up to time t.

If the prediction step m is much smaller than the characteristic width B_X of the process (see [Priestley1965]), it would be a reasonable approximation to replace $f_{t+m}(\omega)$ by $f_t(\omega)$—the most recently available spectrum. In fact, it seems clear intuitively that the very nature of a non-stationary process precludes long-range prediction, unless we are prepared to make assumptions about the character of the non-stationarity. Consequently, if $m \gg B_X$, we would have to extrapolate the available spectra to the time point $(t + m)$. This extrapolation could be performed on the spectral ordinates themselves (by fitting suitable functions of t to a range of values of ω), but in general we would presumably start from a model of the process which determines the form of the evolutionary spectra in terms of a set of time-varying parameters $\{\alpha_i(t)\}$, and then extrapolate the values of these parameters. If our model specifies further the functional time-dependence of the $\{\alpha_i(t)\}$ in terms of another set of constant parameters $\{\beta_j\}$, then the extrapolation is a straightforward problem and involves merely the estimation of the $\{\beta_j\}$. If, on the other hand, the functional form of the $\{\alpha_i(t)\}$ is not specified by the model, the extrapolation would have to be performed by regression techniques.

References

Bibliography

[Bartlett1955] Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge: University Press.

[Bendat1956] Bendat, J. S. (1956). A general theory of linear prediction and filtering. *J. Soc. Industr. Appl. Math.*, 4, 131–151.

[Bendat1957] Bendat, J. S. (1957). Exact integral equation solutions and synthesis for a large class of optimum time variable linear filters. *Trans. Inst. Radio Engrs, IT-3*, 71–80.

[Bendat1959] Bendat, J. S. (1959). *Principles and Applications of Random Noise Theory*. New York: Wiley.

[Booton1952] Booton, R. C. (1952). An optimization theory for time-varying linear systems with non-stationary statistical inputs. *Proc. Inst. Radio Engrs*, 40, 977–981.

[Cramer1961a] Cramér, H. (1961a). On some classes of non-stationary processes. *Proc. 4th Berkeley Symp. Math. Statist. & Prob.*, 2, 57–78.

[Cramer1961b] Cramér, H. (1961b). On the structure of purely non-deterministic stochastic processes. *Ark. Mat.*, 19, 249–266.

[Davis1952] Davis, R. C. (1952). On the theory of prediction of non-stationary stochastic processes. *J. Appl. Phys.*, 23, 1047–1053.

[Doob1953] Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.

[GrangerHatanaka1964] Granger, C. W. J. and Hatanaka, M. (1964). *Spectral Analysis of Economic Time Series*. Princeton: University Press.

[GrenanderRosenblatt1957] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.

[Kalman1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng. (Trans. A.S.M.E., Series D)*, 82, 35–45.

[Parzen1961] Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.*, 32, 951–989.

[Priestley1965] Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc. B*, 27, 204–237.

[Priestley1966] Priestley, M. B. (1966). Design relations for non-stationary processes. *J. R. Statist. Soc. B*, 28, 228–240.

[Tricomi1957] Tricomi, F. G. (1957). *Integral Equations*. New York: Wiley.

[Whittle1963] Whittle, P. (1963). *Prediction and Regulation*. London: English Universities Press.

[Whittle1965] Whittle, P. (1965). Recursive relations for predictors of non-stationary processes. *J. R. Statist. Soc. B*, 27, 523–532.

[Yaglom1962] Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions*. Englewood Cliffs, N.J.: Prentice-Hall.

[Zadeh1953] Zadeh, L. A. (1953). Optimum non-linear filters. *J. Appl. Phys.*, 24, 396–404.