

LINEAR PROCESSES AND BISPECTRA

M. ROSENBLATT,* *University of California, San Diego*

Abstract

A linear process is generated by applying a linear filter to independent, identically distributed random variables. Only the modulus of the frequency response function can be estimated if only the linear process is observed and if the independent identically distributed random variables are Gaussian. In this case a number of distinct but related problems coalesce and the usual discussion of these problems assumes, for example, in the case of a moving average that the zeros of the polynomial given by the filter have modulus greater than one. However, if the independent identically distributed random variables are non-Gaussian, these problems become distinct and one can estimate the transfer function under appropriate conditions except for a possible linear phase shift by using higher-order spectral estimates.

LINEAR PROCESS; BISPECTRUM; NON-GAUSSIAN; FREQUENCY RESPONSE FUNCTION; MODULUS; PHASE; AUTOREGRESSIVE MOVING AVERAGE MODEL; LINEAR AND NON-LINEAR PREDICTION; ESTIMATION

Let v_t , $t = \dots, -1, 0, 1, \dots$, be independent identically distributed random variables with mean zero, $Ev_t \equiv 0$, and variance one, $Ev_t^2 \equiv 1$. Consider a sequence of real constants $\{\alpha_j\}$ with

$$\sum_{j=-\infty}^{\infty} \alpha_j^2 < \infty.$$

The process

$$(1) \quad x_t = \sum_{j=-\infty}^{\infty} \alpha_j v_{t-j}$$

is the linear process generated by $\{\alpha_j\}$ and $\{v_t\}$. The frequency response function is

$$(2) \quad \alpha(e^{-i\lambda}) = \sum_j \alpha_j e^{-ij\lambda}$$

Received 5 December 1978; revision received 23 February 1979.

* Postal address: Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, U.S.A.

Research partially supported by the Office of Naval Research Contract N00014–75–C–0428.

and the spectral density of $\{x_t\}$ is

$$(3) \quad f(\lambda) = \frac{1}{2\pi} |\alpha(e^{-i\lambda})|^2.$$

We shall actually assume that $\sum |\alpha_j| < \infty$ (and often more) so that $\alpha(e^{-i\lambda})$ and $f(\lambda)$ are continuous. Of course, if both x_t and v_t , $t = 1, \dots, T$, are observed $\alpha(e^{-i\lambda})$ can be estimated by estimating the cross-spectral density of $\{x_t\}$ and $\{v_t\}$.

However, if only $\{x_t\}$ is observed a number of distinct but related problems arise. If $\{v_t\}$ (and hence $\{x_t\}$) is Gaussian, then only $f(\lambda)$ or equivalently the modulus of $|\alpha(e^{-i\lambda})|$ can be estimated. It is then usual to assume that if $\{x_t\}$ is an autoregressive moving average process

$$(4) \quad \sum_{j=0}^p b_j x_{t-j} = \sum_{k=0}^q a_k v_{t-k}, \quad (b_0, q_0 \neq 0)$$

that the roots of the polynomials

$$(5) \quad \begin{aligned} A(z) &= \sum_{k=0}^q a_k z^k \\ B(z) &= \sum_{j=0}^p b_j z^j \end{aligned}$$

are distinct and all have modulus greater than one (see [1] and [4]). It is clear that

$$(6) \quad \alpha(e^{-i\lambda}) = \frac{A(e^{-i\lambda})}{B(e^{-i\lambda})}.$$

In the Gaussian case any zero z_j can be replaced by \bar{z}_j^{-1} without changing $f(\lambda)$ so that there are generally 2^{p+q} possible specifications of the zeros without changing the structure of $\{x_t\}$. However, the specification in which the roots all have modulus greater than one corresponds to the prediction problem in the following sense. Then $\{x_t\}$ can be represented as a one-sided moving average in terms of $\{v_t\}$:

$$(7) \quad x_t = \sum_{j=0}^{\infty} \beta_j v_{t-j}, \quad \sum |\beta_j| < \infty.$$

v_t can be represented as a one-sided moving average in terms of x_t

$$(8) \quad v_t = \sum_{j=0}^{\infty} \gamma_j x_{t-j}, \quad \sum |\gamma_j| < \infty.$$

Thus v_t is independent of the past of $\{x_t\}$, that is, x_{t-1}, x_{t-2}, \dots , and the best one-step predictor of x_t in terms of x_{t-1}, x_{t-2}, \dots is

$$(9) \quad x_t^* = - \sum_{j=1}^p \frac{b_j}{b_0} x_{t-j} + \sum_{k=1}^q \frac{a_k}{b_0} v_{t-k}$$

with prediction error

$$(10) \quad x_t - x_t^* = \frac{a_0}{b_0} v_t, \quad E |x_t - x_t^*|^2 = \left| \frac{a_0}{b_0} \right|^2.$$

It is also true that if one regards an autoregressive scheme

$$\sum_{j=0}^p b_j x_{t-j} = v_t$$

as a system characterized by x_t being driven by v_t , then the condition on roots of $B(z)$ having modulus greater than one is equivalent to ‘stability’ of the system (see [2], [5]). Of course, if one is just interested in estimating the spectral density

$$(11) \quad f(\lambda) = \frac{1}{2\pi} \left| \frac{A(e^{-i\lambda})}{B(e^{-i\lambda})} \right|^2,$$

of an autoregressive moving average scheme, the roots taken with modulus greater than one may be as convenient a set of parameters for such a finite parameter scheme as any others.

The prediction problem for a Gaussian sequence has the same character for forward prediction as for backward prediction. This is decidedly not the case for linear processes with non-Gaussian v_t . The simplest case is that of the stationary autoregressive sequence

$$(12) \quad x_t = \frac{1}{2} x_{t-1} + v_t$$

with

$$v_t = \begin{cases} \frac{1}{2} & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases}$$

and v_t independent of the past of the x_t process. Here the best predictor of x_t in terms of the past (in the sense of minimizing mean square error of prediction) is

$$x_t^* = \frac{1}{2} x_{t-1} + \frac{1}{4}.$$

The fact that v_t, x_t do not have mean zero here should not trouble one. It just accounts for the constant $\frac{1}{4}$ in the expression for the best predictor. Notice that the marginal distribution of x_t is uniform $[0, 1]$. The best predictor of x_t given the future is

$$\bar{x}_t^* = 2x_{t+1} \text{ modulo one} = x_t.$$

Thus the backward predictor is non-linear and has variance of prediction error zero (perfect prediction) while the forward predictor is linear and has positive variance of prediction error. The process (12) is purely non-deterministic going forwards in time and purely deterministic going backwards in time. Of course, this simple example is a bit extreme. If one has an autoregressive moving

average process $\{x_t\}$ with the v_t 's non-Gaussian and by chance the zeros of $A(z)$, $B(z)$ of modulus greater than one, then it is still true that the best predictor of x_t in terms of the past is given by (9) with prediction error (10). But if the zeros of $A(z)$, $B(z)$ do not all have modulus greater than one, then the best predictor of x_t given the past will no longer have the form (9) and will generally be non-linear. Also, as we shall see, the function (6) and hence the zeros (as contrasted with the Gaussian case) will generally be identifiable.

A simple example is now given of a situation in which there is interest in estimating $\alpha(e^{-i\lambda})$ or as much of it as one can manage. Suppose $\{v_t\}$ is independent, identically distributed non-Gaussian. It is passed through the linear filter with coefficients $\{\alpha_j\}$ by (1) to get $\{x_t\}$. The experimenter observes only $\{x_t\}$ (not $\{v_t\}$) and wishes to estimate $\alpha(e^{-i\lambda})$. A particular case is that of the moving average

$$(13) \quad x_t = 6v_t - 5v_{t-1} + v_{t-2}$$

where the roots of $A(z)$ are 2 and 3. The moving average

$$(14) \quad x_t = 3v_t - 7v_{t-1} + 2v_{t-2}$$

has a polynomial $A(z)$ with roots at $\frac{1}{2}$ and 3. Both schemes (13) and (14) have the same spectral density but if the independent, identically distributed v_t 's are exponential, the marginal distributions of x_t for these schemes are different.

Assume that one has a linear process where the independent, identically distributed non-Gaussian variables v_t have finite moment $m_k = Ev_t^k (k > 2)$ and corresponding cumulant γ_k . The k th-order cumulant spectral density is then

$$(15) \quad \frac{1}{(2\pi)^{k-1}} \sum_{j_1, \dots, j_{k-1}} \text{cum}(x_t, x_{t+j_1}, \dots, x_{t+j_{k-1}}) \exp\left(-\sum_{s=1}^{k-1} ij_s \lambda_s\right) \\ = \frac{1}{(2\pi)^{k-1}} \alpha(e^{-i\lambda_1}) \alpha(e^{-i\lambda_2}) \dots \alpha(e^{-i\lambda_{k-1}}) \alpha(e^{i(\lambda_1 + \lambda_2 + \dots + \lambda_{k-1})}) \gamma_k.$$

We shall now indicate circumstances under which $\alpha(e^{-i\lambda})$ can be identified except for a linear phase shift.

Theorem. Let $\{x_t\}$ be a linear process generated by independent, identically distributed random variables $\{v_t\}$ with finite third cumulant $\gamma = \gamma(v) \neq 0$. Further let the coefficients α_j be such that

$$(16) \quad \sum_j |j| |\alpha_j| < \infty$$

and $\alpha(e^{-i\lambda}) \neq 0$ for any λ . Then $\alpha(e^{-i\lambda})$ is identifiable except for a linear phase shift if one observes the process $\{x_t\}$. The conclusion of the theorem is still valid if the $\{v_t\}$ have finite k th ($k > 2$) cumulant $\gamma_k = \gamma_k(v) \neq 0$.

Obviously $|\alpha(e^{-i\lambda})|$ is identifiable since $|\alpha(e^{-i\lambda})| = (2\pi)^{\frac{1}{2}}[f(\lambda)]^{\frac{1}{2}}$. We thus have only to deal with $\arg\{\alpha(e^{-i\lambda})\}$. Let

$$h(\lambda) = \arg\{\alpha(e^{-i\lambda})\} - \arg\{\alpha(1)\}.$$

Since the α_j 's are real, $h(-\lambda) = -h(\lambda)$. The bispectrum of the x process is

$$b(\lambda, \mu) = \frac{1}{(2\pi)^2} \alpha(e^{-i\lambda})\alpha(e^{-i\mu})\alpha(e^{i(\lambda+\mu)})\gamma.$$

$(\alpha(1)/|\alpha(1)|)\gamma$ can be determined since

$$\frac{\alpha(1)}{|\alpha(1)|} \gamma = (2\pi)^{\frac{1}{2}}\{b(0, 0)\}/\{f(0)\}^{\frac{3}{2}}.$$

Also

$$h(\lambda) + h(\mu) - h(\lambda + \mu)$$

can be determined since

$$\begin{aligned} & h(\lambda) + h(\mu) - h(\lambda + \mu) \\ &= \frac{1}{i} \log[(2\pi)^{\frac{1}{2}}\gamma^{-1}b(\lambda, \mu)\{f(\lambda)f(\mu)f(\lambda + \mu)\}^{-\frac{1}{2}}] - \arg\{\alpha(1)\}. \end{aligned}$$

Notice that

$$h'(0) - h'(\lambda) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \{h(\lambda) + h(\Delta) - h(\lambda + \Delta)\}.$$

Set $c = h'(0)$. Then

$$h(\lambda) = \int_0^\lambda \{h'(u) - h'(0)\}du + c\lambda.$$

Thus

$$\alpha(e^{-i\lambda}) = u(\lambda)e^{ic\lambda}$$

and by the simple argument given above $\alpha(e^{-i\lambda})$ is identifiable except for sign and for the linear phase shift $c\lambda$. The shift $c\lambda$ is clearly not identifiable under our assumptions unless one makes additional specifications. Notice that if $c = n$ (integer), this just corresponds to a time shift of n units.

The argument given can be immediately elaborated to take care of finite k th cumulant $\gamma_k \neq 0$ ($k > 3$) if $\gamma_3 = 0$.

The k th cumulant spectrum of the process is given by (15) and a simple adaptation of the proof of the theorem yields the result.

Given that x_1, x_2, \dots, x_N is observed one can easily suggest a consistent estimate of $\alpha(e^{-i\lambda})$ up to the unidentifiable phase shift $c\lambda$. The estimate is suggested by the approximation

$$\int_0^\lambda \{h'(u) - h'(0)\} du \cong \sum_{j=1}^{[m\lambda]} \left\{ h' \left(\frac{j}{m} \right) - h'(0) \right\} \frac{1}{m}$$

$$\cong - \sum_{j=1}^{[m\lambda]} \left\{ h \left(\frac{j}{m} \right) + h \left(\frac{1}{m} \right) - h \left(\frac{j+1}{m} \right) \right\}$$

as $m \rightarrow \infty$. Let $f_N(\lambda)$, $b_N(\lambda, \mu)$ be estimates of the spectral density and bispectral density with symmetric weight functions (see [3] for a discussion of such estimates). Let β_N be the linear bandwidth of the estimates so that $\beta_N \downarrow 0$, $N\beta_N^2 \rightarrow \infty$. Now $f_N(\lambda)^{\frac{1}{2}}(2\pi)^{\frac{1}{2}}$ is a consistent estimate of $|\alpha(e^{-i\lambda})|$. Further

$$(2\pi)^{\frac{1}{2}}\{b_N(0, 0)\}\{f_N(0)\}^{-3/2}$$

is a consistent estimate of $(\alpha(1)/|\alpha(1)|)\gamma$. This implies that

$$(2\pi)^{\frac{1}{2}}f_N(\lambda)^{\frac{1}{2}} \prod_{j=1}^{[m\lambda]} \left(\left[b_N \left(\frac{1}{m}, \frac{j}{m} \right) \left\{ f_N \left(\frac{1}{m} \right) f_N \left(\frac{j}{m} \right) f_N \left(\frac{j+1}{m} \right) \right\}^{-\frac{1}{2}} \right]^{-1} b_N(0, 0) f_N(0)^{-\frac{3}{2}} \right)$$

will be a consistent estimate of $\alpha(e^{-i\lambda})$ up to sign and the linear phase shift $c\lambda$ if $m\beta_N^2 \rightarrow 0$, $m(N\beta_N^2)^{-1} \rightarrow 0$ as $m, N \rightarrow \infty$, $\beta_N \rightarrow 0$. Of course, this is a complicated and unwieldy estimate.

Notice that when dealing with the autoregressive moving average scheme (4), the specification $a_0, b_0 \neq 0$ implies that the phase shift $c\lambda \equiv 0$.

References

[1] ANDERSON, T. W. (1977) Estimation for autoregressive moving average models in the time and frequency domain, *Ann. Statist.* 5, 842–865.
 [2] ÅSTRÖM, K. (1970) *Introduction to Stochastic Control Theory*. Academic Press, New York.
 [3] BRILLINGER, D. R. AND ROSENBLATT, M. (1967) Asymptotic theory of estimates of k th order spectra. In *Spectral Analysis of Time Series*, ed. B. Harris, Wiley, New York, 153–188.
 [4] HANNAN, E. J. (1970) *Multiple Time Series*. Wiley, New York.
 [5] WHITTLE, P. (1963) *Prediction and Regulation*. English Universities Press, London.