PREDICTION AND NONGAUSSIAN AUTOREGRESSIVE STATIONARY SEQUENCES[1]

Murray Rosenblatt

University of California, San Diego

Abstract. The object of this paper is to show that under certain auxiliary assumptions a stationary autoregressive sequence has a best predictor in mean square that is linear if and only if the sequence is minimum phase or is Gaussian when all moments are finite.

1. Introduction. We consider a stationary autoregressive sequence, that is, a stationary sequence x_t satisfying the system of equations

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = \xi_t, \quad t = \cdots, -1, 0, 1, \cdots \tag{1.1}$$

with the ξ_t's independent identically distributed, the ϕ_i's real and $E\xi_t \equiv 0, E\xi_t^2 = \sigma^2 > 0$. Let

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p. \tag{1.2}$$

The system of equations is satisfied by a strictly stationary sequence (which is uniquely determined) if and only if $\phi(z)$ has no roots of absolute value 1. In [4] a simple result of the type considered in this paper was established for a first order autoregressive scheme x_t satisfying

$$x_t - \beta x_{t-1} = \xi_t, \quad t = \cdots, -1, 0, 1, \cdots, \quad 0 < |\beta| < 1. \tag{1.3}$$

Clearly the best one-step predictor (predicting ahead) of x_{t+1} is the linear predictor βx_t. However, the best one-step predictor with time reversed for the process (1.3),

$$E(x_t \mid x_{t+1})$$

is linear if and only if the distribution of x_t is Gaussian. Let G be the distribution function of ξ_t and let F the distribution function of x_t. It is clear that F satisfies the equation

$$F(\cdot) = G(\cdot) * F(\beta^{-1}\cdot),$$

where the asterisk $(*)$ denotes the convolution operation. If φ is the characteristic function of ξ_t, η the characteristic function of x_t

$$\eta(\tau) = \prod_{j=0}^{\infty} \varphi(\beta^j \tau),$$

1991 *Mathematics Subject Classification.* Primary 60G25, 62M20; secondary 60G10, 10J10.

Key words and phrases. Autoregressive sequence, nonGaussian, nonminimum phase, nonlinear prediction.

Research supported by ONR Grant N00014-90-J1371

and $\phi(\tau_1,\tau_2) = E\exp(i\tau_1 x_{-1} + i\tau_2 x_0)$ the joint characteristic function of x_{-1} and x_0, the following relation is satisfied:

(1.4)
$$\phi_{\tau_1}(0,\tau_2) = \frac{1}{\beta}\,\eta'(\tau_2) - \frac{1}{\beta}\,\varphi'(\tau_2)\eta(\beta\tau_2) = \int \exp(i\tau_2 x) iE(x_{-1} \mid x_0 = x)\,dF(x).$$

Let
$$\overline{G}(x) = \int_{-\infty}^{x} u dG(u).$$

The relation (1.4) implies that $E(x_{-1} \mid x_0 = x)$ is given by

$$\frac{1}{\beta}x - \frac{1}{\beta}\left\{\overline{G} * F(\beta^{-1}\cdot)\right\}(dx)/F(dx).$$

A related problem for heavy-tailed distributions is considered in [2].

Factor the polynomial (1.2),

$$\phi(z) = \phi^{+}(z)\phi^{*}(z)$$

where
$$\phi^{+}(z) = 1 - \theta_1 z - \cdots - \theta_r z^r \neq 0 \quad \text{for } |z| \leq 1,$$
$$\phi^{*}(z) = 1 - \theta_{r+1} z - \cdots - \theta_p z^s \neq 0 \quad \text{for } |z| \geq 1.$$

and $r, s \geq 0, r+s = p$. Given that $m_1, \ldots, m_r, m_{r+1}, \ldots, m_p$ are the p zeros of $\phi(z)$ let $|m_i| > 1, i = 1, \ldots, r$ and $|m_i| < 1, i = r+1, \ldots, p$. Then

$$\phi^{+}(z) = \prod_{i=1}^{r}\left(1 - m_i^{-1}z\right), \quad \phi^{*}(z) = \prod_{i=r+1}^{p}\left(1 - m_i^{-1}z\right).$$

The autoregressive sequence (1.1) is called minimum phase if $r = p$ and nonminimum phase otherwise. If the sequence is minimum phase, clearly one can write

$$x_t = \sum_{j=0}^{\infty} \alpha_j \xi_{t-j} \tag{1.5}$$

where

$$\phi(z)^{-1} = \sum_{j=0}^{\infty} \alpha_j z^j \tag{1.6}$$

and the α_j's decay to zero exponentially fast as $j \to \infty$. The relations (1.5) and (1.6) imply that the σ-algebras generated by $\{\xi_j, j \leq t\}$ and $\{x_j, j \leq t\}$ are the same. This implies that in the minimum phase case the best predictor in mean square of x_{t+1} given $x_j, j \leq t$, is linear and given by

$$x_t^{*} = \sum_{j=1}^{\infty} \alpha_j \xi_{t+1-j}.$$

Our object is to show that in the nonminimum phase nonGaussian case, the best predictor in mean square of x_{t+1} given $x_j, j \leq t$, is nonlinear if all moments of ξ_t are finite and the roots $m_i, i = r+1, \ldots, p$, are distinct. In Section 2 we show that the stationary solution of (1.1) is pth order Markovian. This implies that the best one-step predictor in mean square in terms of the past is a function of the p preceding variables. That the solution of (1.1) is pth order Markovian is obvious in the minimum phase case since the solution is causal, implying that ξ_t is independent of the past of the x process, that is $x_{t-1}, x_{t-2}, \ldots$. In the nonminimum phase case the x_t process is noncausal and so ξ_t is not independent of the past of the x process. The Markovian property of the x process is used in Section 3 where the principal result on the nonlinearity of the best predictor in mean square of x_{t+1} given $x_j, j \leq t$, is derived in the nonminimum phase nonGaussian case when all the moments of the ξ_t are finite and the roots $m_i, i = r+1, \ldots, p$ are distinct. One should note that the first order autoregressive scheme with time reversed discussed in this section is not minimum phase.

2. The Markov Property. Our object in this section is to show that the stationary autoregressive sequence is pth order Markovian, whether it is minimum phase or not. Part of the argument parallels one given in [1]. The argument is carried out in the case $r, s > 0$, since it is obvious otherwise.

Introduce the causal and purely noncausal sequences

$$U_t = \phi^*(B)x_t \ , \quad V_t = \phi^+(B)x_t,$$

with B the one-step backshift symbol, that is, $Bx_t = x_{t-1}$. We then have

$$U_t = \sum_{j=0}^{\infty} \alpha_j \xi_{t-j}, \quad V_t = \sum_{j=s}^{\infty} \beta_j \xi_{t+j}$$

where

$$\phi^+(z)^{-1} = \sum_{j=0}^{\infty} \alpha_j z^j \ , \quad \phi^*(z)^{-1} = \sum_{j=s}^{\infty} \beta_j z^{-j}.$$

Let us also note that we have

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j \xi_{t-j},$$

where

$$\phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j z^j. \tag{2.1}$$

We shall carry through the argument assuming the existence of positive density functions. However, essentially the same argument can be carried through without this assumption using a more elaborate notation. Let the density function of the ξ random variables be g. The random variables $U_\ell, \ell \leq t$, are independent of $V_\ell, \ell \geq t-s+1$, and so the joint probability density function of $(U_1, \ldots, U_n, V_{n-s+1}, \ldots, V_n)$ is

$$h_U(U_{1,\ldots,}U_r) \left\{ \prod_{t=r+1}^{n} g\left(U_t - \theta_1 U_{t-1} - \cdots - \theta_r U_{t-r}\right) \right\} h_V(V_{n-s+1}, \ldots, V_n)$$

where h_U and h_V are the joint probability density functions of $(U_1, \ldots, U_r)$ and $(V_{n-s+1}, \ldots, V_n)$ respectively. Consider the linear transformation T_n given by

$$\begin{bmatrix} U_1 \\ \vdots \\ U_s \\ U_{s+1} \\ \vdots \\ U_n \\ V_{n-s+1} \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} U_1 \\ \vdots \\ U_s \\ x_{s+1} - \theta_{r+1}x_s - \cdots - \theta_p x_1 \\ \vdots \\ x_n - \theta_{r+1}x_{n-1} - \cdots - \theta_p x_{n-s} \\ x_{n-s+1} - \theta_1 x_{n-s} - \cdots - \theta_r x_{n-s+1-r} \\ \vdots \\ x_n - \theta_1 x_{n-1} - \cdots - \theta_r x_{n-r} \end{bmatrix} = T_n \begin{bmatrix} U_1 \\ \vdots \\ U_s \\ x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Using this transformation one can see that the joint density of $(U_1, \ldots, U_s, x_1, \ldots, x_n)$ is

$$\begin{aligned} h_U\left(\widetilde{U}_1, \ldots, \widetilde{U}_r\right) &\left\{ \prod_{t=r+1}^{p} g\left(\widetilde{U}_t - \theta_1 \widetilde{U}_{t-1} - \cdots - \theta_r \widetilde{U}_{t-r}\right)\right\} \\ &\times \left\{ \prod_{t=p+1}^{n} g\left(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p}\right)\right\} \\ &\times h_V(\phi^+(B)x_{n-s+1}, \ldots, \phi^+(B)x_n) |\det(T_n)| \end{aligned}$$

where

$$\widetilde{U}_\ell = \begin{cases} U_\ell & \text{if } \ell \le s \\ x_\ell - \theta_{r+1}x_{\ell-1} - \cdots - \theta_p x_{\ell-s} & \text{if } \ell > s. \end{cases}$$

If $s > 0$, $\ln|\det(T_n)| \sim \ln|\theta_p|^{n-p}$. Let us compute the conditional density of $x_n, x_{n-1}, \ldots, x_{n-p}$ given $x_{n-d}, x_{n-1-d}, \ldots, x_1, U_s, \ldots, U_1$. The one-step ($d = 1$) conditional density is given by

$$g(x_n - \phi_1 x_{n-1} - \cdots - \phi_p x_{n-p}) \frac{h_V(\phi^+(B)x_{n-s+1}, \cdots, \phi^+(B)x_n)}{h_V(\phi^+(B)x_{n-s}, \ldots, \phi^+(B)x_{n-1})} \; \frac{|\det(T_n)|}{|\det(T_{n-1})|},$$

whereas if $1 < d \le p+1$, one obtains

$$\begin{aligned} &\left\{ \prod_{u=0}^{d-1} g(x_{n-u} - \phi_1 x_{n-1-u} - \cdots - \phi_p x_{n-u-p}) \right\} \\ &\qquad \times \frac{h_V(\phi^+(B)x_{n-s+1}, \ldots, \phi^+(B)x_n)}{h_V(\phi^+(B)x_{n-d-s+1}, \ldots, \phi^+(B)x_{n-d})} \; \frac{\det(T_n)}{\det(T_{n-d})}. \end{aligned}$$

If $d > p+1$ the conditional probability density is

$$\begin{aligned} \int \cdots \int &\left\{ \prod_{t=n-d+1}^{n} g(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p}) \right\} dx_{n-d+1} \cdots dx_{n-p-1} \\ &\times \frac{h_V(\phi^+(B)x_{n-s+1}, \ldots, \phi^+(B)x_n)}{h_V(\phi^+(B)x_{n-d-s+1}, \ldots, \phi^+(B)x_{n-d})} \; \frac{\det(T_n)}{\det(T_{n-d})}. \end{aligned}$$

Notice that in all these cases the conditional probability density depends on $x_{n-d}, x_{n-d-1}, \cdots, x_1, U_s, \cdots, U_1$ only through $x_{n-d}, \ldots, x_{n-d-p+1}$. However, this implies that the conditional probability density of $x_n, x_{n-1}, \ldots, x_{n-p}$ given $x_{n-d}, x_{n-1-d}, \ldots, x_1$ is the same by a standard argument using conditional expectations. The argument is that if f is integrable, $\mathcal{B}$ and $\mathcal{A}$ are σ-algebras, then if $E(f \mid \mathcal{B}, \mathcal{A}) = h$ is $\mathcal{B}$ measurable, it follows that $E(f \mid \mathcal{B}) = E(E(f \mid \mathcal{B}, \mathcal{A}) \mid \mathcal{B}) = E(h \mid \mathcal{B}) = h$. Thus $\{X_n\}$ is a sequence that is pth order Markovian.

3. A functional equation for the characteristic function. The characteristic function of x_k is clearly

$$\eta(t) = \prod_{k=-\infty}^{\infty} \varphi(\psi_k t).$$

The joint characteristic function of the random variables $x_{-s}, x_{-s+1}, \ldots, x_0$ is

$$\eta(\tau_s, \tau_{s-1}, \ldots, \tau_0) = E\left\{\exp\left[i \sum_{\ell=0}^{s} \tau_\ell x_{-\ell}\right]\right\} = \prod_{k=-\infty}^{\infty} \varphi\left(\sum_{\ell=0}^{s} \tau_\ell \psi_{k-\ell}\right),$$

whereas the joint characteristic function of $x_{-s}, x_{-s+1}, \ldots, x_{-1}$ is

$$\widetilde{\eta}(\tau_s, \tau_{s-1}, \ldots, \tau_1) = \prod_{k=-\infty}^{\infty} \varphi\left(\sum_{\ell=1}^{s} \tau_\ell \psi_{k-\ell}\right).$$

It is clear that

$$\begin{aligned}
&\frac{\partial}{\partial \tau_0} \eta(\tau_s, \ldots, \tau_1, \tau_0)_{|\tau_0=0} = \eta_{\tau_0}(\tau_2, \ldots, \tau_1, 0) \\
&\quad = \int i x_0 \exp\left(i \sum_{\ell=1}^{s} \tau_\ell x_{-\ell}\right) dF(x_{-s}, \ldots, x_{-1}, x_0) \\
&\quad = i \int E(x_0 \mid x_{-1}, \ldots, x_{-s}) \exp\left(i \sum_{\ell=1}^{s} \tau_\ell x_{-\ell}\right) dF(x_{-s}, \ldots, x_{-1}),
\end{aligned}$$

where $F(x_{-s}, \ldots, x_{-1})$ is the joint distribution function of $x_{-s}, \ldots, x_{-1}$. In the case of the pth order autoregressive sequence x_t, since the sequence is a Markov process of order p, it is sufficient in considering the best one-step predictor (in mean square) to consider $s = p$, since the one-step predictor of x_0 given the whole past will depend only on the p immediately preceding random variables. Now

$$\begin{aligned}
\frac{\partial}{\partial \tau_0} \log \eta(\tau_p, \ldots, \tau_1, \tau_0)_{|\tau_0=0} &= \frac{\eta_{\tau_0}(\tau_p, \ldots, \tau_1, 0)}{\widetilde{\eta}(\tau_p, \ldots, \tau_1)} \\
&= \sum_{k=-\infty}^{\infty} \psi_k \varphi'\left(\sum_{\ell=1}^{p} \tau_\ell \psi_{k-\ell}\right) \Big/ \varphi\left(\sum_{\ell=1}^{p} \tau_\ell \psi_{k-\ell}\right)
\end{aligned}$$

while

$$\frac{\partial}{\partial \tau_j} \log \widetilde{\eta}(\tau_p, \ldots, \tau_1) = \sum_{k=-\infty}^{\infty} \psi_{k-j} \varphi'\left(\sum_{\ell=1}^{p} \tau_\ell \psi_{k-\ell}\right) \Big/ \varphi\left(\sum_{\ell=1}^{p} \tau_\ell \psi_{k-\ell}\right)$$

for some neighborhood of the origin $|\tau_p|, \ldots, |\tau_1| \leq \varepsilon, \varepsilon > 0$. If the best predictor is linear we must have

$$\eta_{\tau_0}(\tau_p, \ldots, \tau_1, 0) = \sum_{j=1}^{p} b_j \widetilde{\eta}_{\tau_j}(\tau_p, \ldots, , \tau_1), \tag{3.1}$$

where the b_j's are the coefficients of the best linear predictor of x_0 in mean square

$$x_0^* = \sum_{j=1}^{p} b_j x_{-j}.$$

This is in turn equivalent to

$$\sum_{k=-\infty}^{\infty} \left(\psi_k - \sum_{\ell=1}^{p} b_\ell \psi_{k-\ell} \right) h \left(\sum_{j=1}^{p} \tau_j \psi_{k-j} \right) = 0 \tag{3.2}$$

where $h(\tau) = \varphi'(\tau) \big/ \varphi(\tau)$ for $(\tau_1, \ldots, \tau_p)$ such that $\widetilde{\eta}(\tau_p, \ldots, \tau_1) \neq 0$. That (3.1) is equivalent to (3.2) follows from the fact that

$$\left| \eta_{\tau_j}(\tau_p, \ldots, \tau_0) \right| \leq \left\{ E\left(x_j^2 \right) \right\}^{1/2}$$

and that the ψ_k tend to zero exponentially as $|k| \to \infty$. The equation (3.2) is similar to the type of functional equation taken up in [3].

The ψ_k are the coefficients in the Laurent expansion of $\phi(z)^{-1}$. The b_ℓ can be read off from the polynomial with constant coefficient positive, having the same absolute value as $\phi(z)$ when $z = e^{-i\lambda}$ and with all its zeros outside the unit disc. Let

$$\phi^{**}(z) = (-1)^s z^s \phi^* \left(\frac{1}{z} \right).$$

Notice that the roots of $\phi^{**}(z)$ are the inverses of the roots of $\phi^*(z)$. Thus the polynomial

$$\zeta(z) = \phi^+(z) \phi^{**}(z)$$

has all its roots outside the unit disc and has the same absolute value as $\phi(z)$. Notice that the coefficients $\psi_k - \sum_{\ell=1}^{p} b_\ell \psi_{k-\ell}$ are those in the Laurent expansion of

$$\zeta(z) \phi(z)^{-1} = \phi^{**}(z) \phi^*(z)^{-1}.$$

If the sequence is minimum phase this function is 1 and so the coefficients are 0 for $k \neq 0$ and 1 when $k = 0$. Further, in the minimum phase case, $\psi_k = 0$ for $k < 0$. The relation (3.2) is automatically satisfied in the minimum phase case whatever the distribution G (as long as $E\xi_t = 0$ and $E\xi_t^2 < \infty$). However, we had already seen this before using another argument.

Proposition 1. *Consider the stationary solution x_t of the system of equations (1.1), where the ξ_t are independent, identically distributed with $E\xi_t \equiv 0, E\xi_t = \sigma^2 < \infty$ and characteristic polynomial $\phi(\cdot)$ [clearly $\phi(z) \neq 0$ if $|z| = 1$]. Then the best one-step predictor for x_t is linear if and only if (3.2) holds where the ψ_k's are given by (2.1) and the b_ℓ's the coefficients of the best linear predictor.*

It is of some interest to essentially characterize the sequence

$$\gamma_k = \psi_k - \sum_{\ell=1}^{p} b_\ell \psi_{k-\ell}, \quad k = \ldots, -1, 0, 1, \ldots$$

The generating function of the ψ_k is $\phi(z)^{-1}$ [see(2.1)]. Since the coefficients b_ℓ correspond to best linear one-step predictor in mean square

$$1-\sum_{\ell=1}^{p} b_\ell z^\ell = c\phi(z)\prod_{i=r+1}^{p}\left\{\frac{(1-m_i z)m_i^{-1}}{(1-m_i^{-1}z)}\right\}$$

with c a nonzero constant. It then follows that

$$\gamma(z)=\sum_{k=-\infty}^{\infty}\gamma_k z^k = c\prod_{i=r+1}^{p}\left\{\frac{(1-m_i z)m_i^{-1}}{(1-m_i^{-1}z)}\right\}. \tag{3.3}$$

Since

$$m^{-1}(1-zm)(1-m^{-1}z)^{-1} = m+(m^2-1)\sum_{j=1}^{\infty} m^{j-1}z^{-j}$$

when $|m|<1$, if follows that

$$\gamma_k = 0 \quad \text{for } k>0.$$

Theorem 1. *Consider the stationary autoregressive sequence $\{x_t\}$ satisfying (1.1) and assume that the random variables ξ_t have all moments finite. Further let the sequence be nonminimum phase with all the zeros $m_i, i=r+1,\dots,p$ simple. Then if the best one-step predictor is linear, the ξ distribution is Gaussian.*

Suppose that the ξ distribution is nonGaussian. There are then an infinite number of nonzero cumulants $\mu_a \neq 0, a>2$, of the ξ distribution. Also

$$\psi_{-k}=\sum_{j=r+1}^{p}\alpha_j m_j^k, \quad k>0$$

for some coefficients $\alpha_j \neq 0, j=r+1,\dots,p$. If $\mu_{a+1}\neq 0$ for some $a\geq 2$, the relation (3.2) implies that

$$\sum_{k=0}^{\infty}\gamma_{-k}\psi_{-k-\ell_1}\cdots\psi_{-k-\ell_a}=0, \quad \ell_1,\dots,\ell_a=1,\dots,p. \tag{3.4}$$

For the ath order partial derivative of the expression in (3.2) with respect to $\tau_{\ell_1},\dots,\tau_{\ell_a}$ at $\tau_{\ell_1}=\cdots=\tau_{\ell_a}=0$, $i^{a+1}\mu_{a+1}a!$ multiplied by the expression on the left of (3.4). The equations (3.4) can be rewritten

$$\sum_{j_1,\dots,j_a=r+1}^{p}\alpha_{j_1}\cdots\alpha_{j_a}m_{j_1}^{\ell_1}\cdots m_{j_a}^{\ell_a}\sum_{k=0}^{\infty}\gamma_{-k}(m_{j_1}\dots m_{j_a})^k=0,$$

$\ell_1,\dots,\ell_a=1,\dots,p$. Consider the set of equations obtained by letting $\ell_1,\dots,\ell_a=1,\dots,s$. The matrix of this set of equations is

$$M=(M_{j,\ell})=\left\{\alpha_{j_1}\cdots\alpha_{j_a}m_{j_1}^{\ell_1}\cdots m_{j_a}^{\ell_a}\right\},$$

$j=(j_1,\dots,j_a), \ell=(\ell_1,,\dots,\ell_a), j_1,\dots,j_a=r+1,\dots,p,\ell_1,\dots,\ell_a=1,\dots,s$. The determinant of this matrix is $\left(\prod_{u=r+1}^{p}\alpha_u\right)^a$ times the a^{th} power of the Vandermonde determinant

$$\left|m_j^\ell;\ j=r+1,\dots,p,\ \ell=1,\dots,s\right|.$$

Since the determinant is nonzero we must have

$$\gamma\left((m_{j_1}\ldots m_{j_a})^{-1}\right)=\sum_{k=0}^{\infty}\gamma_{-k}(m_{j_1}\ldots m_{j_a})^{k}=0,$$

$j_1,\ldots,j_a=r+1,\ldots,p$. These are too many zeros for the function $\gamma(z)$ and so we must have $\mu_{a+1}=0$. Since this holds for any $a\geq 2$, the ξ distribution must be Gaussian

NonGaussian nonminimum phase autoregressive sequences arise naturally when considering transects of certain classes of random fields.

References

[1] F. Breidt, R. Davis, K. S. Lii and M. Rosenblatt, *Maximum likelihood estimation for noncausal autoregressive processes*, J. Mult. Anal. **36,** (1991), 175–198.

[2] S. Cambanis and I. Fakhre-Zaker, *On prediction of heavy-tailed autoregressive sequences: regression versus best linear prediction*, Technical report No. 383, Center for Stochastic Processes, University of North Carolina.

[3] B. Ramchandran and Ka-Sing Lau, *Functional Equations in Probability Theory*, Academic Press, 1991.

[4] M. Rosenblatt, *A note on prediction and an autoregressive sequence*, Stochastic Processes (S. Cambanis, J. Ghosh, R. Karandikar and P. Sen, eds.), 1993, pp. 291-295.

Department of Mathematics
University of California, San Diego
La Jolla, California 92093