Stochastic Processes and their Applications 98 (2002) 67–76

stochastic processes and their applications

www.elsevier.com/locate/spa

Prediction with incomplete past of a stationary process

Pascal Bondon *

Laboratoire des Signaux et Systèmes, CNRS UMR 8506, 3 Rue Joliot-Curie, F-91192 Gif-sur-Yvette Cedex, France

Received 22 November 2000; received in revised form 22 May 2001; accepted 15 June 2001

Abstract

An explicit formula is obtained for the prediction error of a future value of a stationary process when the infinite past is altered by some missing observations with an arbitrary pattern. Then the autoregressive representation of the predictor is derived and the processes for which the missing observations in the past do not affect the prediction of a future value are characterized. Some properties for autoregressive processes and for moving average processes with finite orders are established. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Prediction theory; Stationary process; Missing value problems; Autoregressive parameters; Autoregressive representation

1. Introduction

Let $(X_k)_{k\in\mathbb{Z}}$ be a zero mean weakly stationary stochastic process. The problem of finding the best linear mean square predictor $\hat{X}_0$ of X_0 based on the infinite past $\{X_k; k \leqslant -1\}$ is of fundamental interest. This problem was studied by Kolmogorov and Wiener in the forties, and the spectral representation and the Wold decomposition theorems are the two most powerful tools for solving it, see Kolmogorov (1941).

Assume that the observations $X_{-n_1},\dots,X_{-n_N}$ are missing in the past, and let $\hat{X}_0'$ be the best linear mean square predictor of X_0 based on the incomplete past $\{X_k;\ k \leqslant -1,\ k \neq -n_1,\dots,-n_N\}$. In the case where $n_1 = 1,\dots,n_N = N$, $\hat{X}_0'$ is the $(N+1)$-step predictor of X_0 which is easily calculated from the Wold decomposition of (X_k), see for instance (Brockwell and Davis, 1991, p. 189). When $n_1,\dots,n_N$ are arbitrary integers, Cheng and Pourahmadi (1997) proposed an algorithm to compute $\hat{X}_0'$ whose idea is the following. Let $X_{-k_1},\dots,X_{-k_K}$ be the observed values of the process between the moments $k = -n_N$ and $k = 0$. We have $K = n_N - N$. The observation space is decomposed as the orthogonal sum

$$\overline{\text{sp}}\{X_k;\ k \leqslant -1,\ k \neq -n_1,\dots,-n_N\} = \overline{\text{sp}}\{X_k;\ k < -n_N\} + \text{sp}\{Y_1,\dots,Y_K\},$$

* Corresponding author. Fax: +33-169851765.

E-mail address: bondon@lss.supelec.fr (P. Bondon).

PII: S0304-4149(01)00116-8

where $Y_i = X_{-k_i} - \tilde{X}_{-k_i}$ and $\tilde{X}_{-k_i}$ is the orthogonal projection of X_{-k_i} onto $\overline{\mathrm{sp}}\{X_k; k < -n_N\}$. Next, the space $\mathrm{sp}\{Y_1, \ldots, Y_K\}$ is decomposed into the sum of K orthogonal one-dimensional subspaces by means of the Gram–Schmidt procedure (innovation algorithm). Hence, the observation space is decomposed as the sum of $K+1$ orthogonal subspaces, and $\hat{X}'_0$ is the sum of the projections of X_0 onto each subspace. The projection of X_0 onto the infinite-dimensional subspace $\overline{\mathrm{sp}}\{X_k;\ k < -n_N\}$ is the (n_N+1)-step predictor of X_0, and computing a projection onto a one-dimensional subspace is an easy task. This method generalizes the innovation algorithm presented in Brockwell and Davis (1991, Proposition 5.2.2) and it is efficient for computing recursively $\hat{X}'_0$ and the prediction error variance $\mathrm{var}(X_0 - \hat{X}'_0)$. The resulting expressions depend only on the moving average (MA) parameters in the Wold decomposition of (X_k) and on the innovation variance. The numerical complexity of this algorithm is related to the integer K which depends on the number of missing values and of their pattern. Thus, the calculation of $\hat{X}'_0$ may be complicated even when only one observation in the past is missing.

Using a result of Grenander and Rosenblatt (1954), a closed form expression for $\mathrm{var}(X_0 - \hat{X}'_0)$ that involves only the autoregressive (AR) parameters of (X_k) and the innovation variance was derived in Pourahmadi (1994). In the present paper, we establish an explicit formula for $X_0 - \hat{X}'_0$, see Theorem 3.1. This formula leads to the above-mentioned expression of $\mathrm{var}(X_0 - \hat{X}'_0)$ and allows us to derive the AR representation of $\hat{X}'_0$ in Theorem 3.2. The calculation of $\hat{X}'_0$ requires to invert a matrix whose dimension depends on the number N of missing values but is independent of their pattern, and whose elements depend only on the AR parameters of (X_k). Theorem 3.2 generalizes to arbitrary integers $n_1, \ldots, n_N$, Theorem 1 of Bloomfield (1985) wherein $n_1 = 1, \ldots, n_N = N$, and it generalizes to nonminimal processes Corollary 1 of Bondon (2000). In Theorem 3.3, we characterize the processes for which the loss of observations in the past does not affect the prediction of X_0, i.e., $\hat{X}'_0 = \hat{X}_0$. This reveals the important role played by the AR parameters in forecasting. Then, some properties are established in Examples 4.2 and 4.3 when (X_k) is either an AR process or an MA process with finite orders.

2. Preliminaries

Let $(\Omega, \mathscr{F}, \mathsf{P})$ be the probability space on which (X_k) is defined and L_2 be the space of the equivalence classes of real valued random variables defined on $(\Omega, \mathscr{F}, \mathsf{P})$ which are square integrable. L_2 endowed with the inner product $\langle X, Y \rangle = \mathsf{E}XY$, where E stands for the expectation operator, is a Hilbert space. We say that (X_k) is (weakly) stationary if $(X_k) \subset L_2$ and $\langle X_k, X_n \rangle$ depends only on $k - n$ for any $(k, n) \in \mathbb{Z}^2$. It is well known that every stationary process (X_k) has a nonnegative measure F on $(-\pi, \pi]$ called its spectral measure such that $\langle X_k, X_0 \rangle = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}k\lambda}\, \mathrm{d}F(\lambda)$. Let $\mathrm{d}\lambda$ be the normalized Lebesgue measure on $(-\pi, \pi]$. The derivative f of the absolutely continuous part of F with respect to $\mathrm{d}\lambda$ is called the spectral density of (X_k).

If S is an arbitrary nonempty family of random variables in L_2, the subspace of all (finite) linear combinations of elements of S is denoted by $\mathrm{sp}\, S$ and its closure in L_2

by $\overline{\mathrm{sp}}\, S$. For any $k \in \mathbb{Z}$, we set

$$H_k = \overline{\mathrm{sp}}\{X_i;\ i \leqslant k\},$$

$$H_{-\infty} = \bigcap_{k \in \mathbb{Z}} H_k,$$

$\hat{X}_k$ denotes the orthogonal projection of X_k onto H_{k-1}, and $\varepsilon_k = X_k - \hat{X}_k$. The uncorrelated sequence (ε_k) is called the innovation process of (X_k).

The process (X_k) is said to be nondeterministic if $X_k \notin H_{k-1}$ for some k. According to Kolmogorov (1941, Theorem 23) a stationary process (X_k) is nondeterministic if and only if (iff) f is positive almost everywhere and $\log f$ is integrable with respect to $\mathrm{d}\lambda$. In this case,

$$\mathrm{var}(\varepsilon_k) = \exp\left[\int_{-\pi}^{\pi} \log f(\lambda)\,\mathrm{d}\lambda\right] = \sigma^2$$

and (X_k) has the Wold decomposition

$$X_k = \sum_{i=0}^{\infty} c_i \varepsilon_{k-i} + V_k, \tag{2.1}$$

where the coefficients (c_k) are unique, $c_0 = 1$ and $\sum_{k=0}^{\infty} c_k^2 < \infty$, (V_k) is deterministic and is uncorrelated with (ε_k), and $(V_k) \subset H_{-\infty}$. The coefficients (c_k) are called the MA parameters of (X_k).

A nondeterministic process (X_k) is said to be purely nondeterministic if $H_{-\infty} = \{0\}$ and in this case (2.1) reduces to

$$X_k = \sum_{i=0}^{\infty} c_i \varepsilon_{k-i}. \tag{2.2}$$

Suppose that (2.2) may be inverted to give

$$\varepsilon_k = -\sum_{i=0}^{\infty} a_i X_{k-i}, \tag{2.3}$$

where the series converges in L_2. Then the coefficients (a_k) are defined by

$$a_0 = -1, \quad a_k = -\sum_{i=0}^{k-1} a_i c_{k-i}, \quad k \geqslant 1 \tag{2.4}$$

and (2.3) may be rewritten

$$X_k = \sum_{i=1}^{\infty} a_i X_{k-i} + \varepsilon_k. \tag{2.5}$$

Eq. (2.5) is called the AR representation of (X_k). The coefficients (a_k) defined by (2.4) always exist and are called the AR parameters of (X_k) even if the series in (2.5) does not converge in L_2. The parameters (a_k) and (c_k) only depend on the Fourier coefficients of $\log f$, and to simplify the notations, we shall set $a_k = c_k = 0$ for any $k < 0$.

The process (X_k) is said to be minimal if $X_k \notin \overline{\mathrm{sp}}\{X_i;\ i \neq k\}$ for some k. Since $H_{k-1} \subset \overline{\mathrm{sp}}\{X_i;\ i \neq k\}$, every minimal process is nondeterministic. According to Kolmogorov (1941, Theorem 24) a stationary process (X_k) is minimal iff f is positive almost everywhere and f^{-1} is integrable with respect to $\mathrm{d}\lambda$, and this spectral characterization is equivalent in the time domain to the condition $\sum_{k=0}^{\infty} a_k^2 < \infty$, see Masani (1960, Lemma 2.7).

Throughout the paper, we shall assume that the data $X_{-n_1}, \ldots, X_{-n_N}$, where N is finite, are missing in the past, and we set

$$M = \{n_0, n_1, \ldots, n_N\}, \quad 0 = n_0 < n_1 < \cdots < n_N.$$

Let

$$H'_{-1} = \overline{\mathrm{sp}}\{X_{-k};\ k \in \mathbb{N} \setminus M\}$$

and we denote by $\hat{X}'_0$ the orthogonal projection of X_0 onto H'_{-1}.

Finally, for any $(k,n) \in \mathbb{Z}^2$, $k \wedge n$ stands for the minimum of k and n, $\delta_k = 0$ when $k \neq 0$, and $\delta_0 = 1$.

3. Prediction with incomplete past

In the following theorem, we establish a useful formula for $X_0 - \hat{X}'_0$.

Theorem 3.1. *Let (X_k) be a nondeterministic stationary process with the innovation process (ε_k) and the AR parameters (a_k). Then*

$$X_0 - \hat{X}'_0 = -\sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p} a_{n_p - j} \varepsilon_{-j}, \tag{3.1}$$

where the coefficients (ψ_p) satisfy the matrix equation

$$U(\psi_0, \psi_1, \ldots, \psi_N)' = (1, 0, \ldots, 0)', \tag{3.2}$$

U being the nonsingular $(N+1) \times (N+1)$ matrix with elements

$$U_{p,q} = \sum_{j=0}^{n_p \wedge n_q} a_{n_p - j} a_{n_q - j}, \quad p, q = 0, \ldots, N. \tag{3.3}$$

The prediction error variance is

$$\mathrm{var}(X_0 - \hat{X}'_0) = \sigma^2 \psi_0. \tag{3.4}$$

Proof. $\hat{X}'_0$ is the unique random variable in H'_{-1} such that $X_0 - \hat{X}'_0 \perp H'_{-1}$. First, let us check that $X_0 - \hat{X}'_0$ given by (3.1) is orthogonal to H'_{-1}. Let $k \in \mathbb{N}$, we have

$$\begin{aligned}
\langle X_0 - \hat{X}'_0, X_{-k} \rangle &= -\sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p} a_{n_p - j} \langle \varepsilon_{-j}, X_{-k} \rangle \quad \text{by (3.1)}, \\
&= -\sigma^2 \sum_{p=0}^{N} \psi_p \sum_{j=k}^{n_p} a_{n_p - j} c_{j-k} \quad \text{by (2.1)}, \\
&= \sigma^2 \sum_{p=0}^{N} \psi_p \delta_{n_p - k} \quad \text{by (2.4)}.
\end{aligned} \tag{3.5}$$

If $k \notin M$, the right-hand side of (3.5) is zero. Therefore, $X_0 - \hat{X}'_0 \perp X_{-k}$ for any $k \in \mathbb{N}\backslash M$, which implies that $X_0 - \hat{X}'_0 \perp H'_{-1}$. Now we verify that $\hat{X}'_0 \in H'_{-1}$. Let $j \in \{0,\dots,n_N\}$, we deduce from (2.1) that

$$X_{-j} = \sum_{i=0}^{n_N-j} c_i \varepsilon_{-j-i} + \sum_{i=n_N-j+1}^{\infty} c_i \varepsilon_{-j-i} + V_{-j}$$
$$= \sum_{i=j}^{n_N} c_{i-j} \varepsilon_{-i} + U_{-j}, \tag{3.6}$$

where $U_{-j} \in H_{-n_N-1}$. Let X, ε, and U be, respectively, the vectors with components X_{-j}, ε_{-j}, and U_{-j} for $j = 0,\dots,n_N$. It results from (3.6) that $X = C\varepsilon + U$ where C is the $(n_N+1)\times(n_N+1)$ upper triangular matrix with elements $C_{p,q} = c_{q-p}$ for $p,q = 0,\dots,n_N$. Since $c_0 = 1$, C is nonsingular and it follows from (2.4) that $C^{-1} = -A$ where A is the upper triangular matrix with elements $A_{p,q} = a_{q-p}$ for $p,q = 0,\dots,n_N$. Set $T = AU$ and denote by T_{-j} the components of T. We have $\varepsilon = -AX + T$, which is equivalent to

$$\varepsilon_{-j} = -\sum_{i=j}^{n_N} a_{i-j} X_{-i} + T_{-j}, \quad j = 0,\dots,n_N. \tag{3.7}$$

Inserting (3.7) in (3.1), we get $\hat{X}'_0 = R_0 + S_0$ where

$$R_0 = X_0 - \sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p} a_{n_p-j} \sum_{i=j}^{n_N} a_{i-j} X_{-i}, \tag{3.8}$$

$$S_0 = \sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p} a_{n_p-j} T_{-j}. \tag{3.9}$$

Let $i \in M$, i.e., $i = n_q$ where $q \in \{0,\dots,N\}$, and let α_{n_q} be the coefficient of X_{-n_q} in the expression of R_0. It follows from (3.8) that

$$\alpha_{n_q} = \delta_{n_q} - \sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p \wedge n_q} a_{n_p-j} a_{n_q-j} = \delta_{n_q} - \sum_{p=0}^{N} \psi_p U_{p,q},$$

where $U_{p,q}$ is defined by (3.3), and we deduce from (3.2) that $\alpha_{n_q} = 0$. Therefore, $R_0 \in \mathrm{sp}\{X_{-k};\ k \in \{0,\dots,n_N\}\backslash M\} \subset H'_{-1}$. Since each $U_{-j} \in H_{-n_N-1}$ and $T = AU$, each $T_{-j} \in H_{-n_N-1}$, and it results from (3.9) that $S_0 \in H_{-n_N-1} \subset H'_{-1}$. Hence, $\hat{X}'_0 = R_0 + S_0 \in H'_{-1}$. Now let us show that matrix U is positive definite which implies that U is nonsingular. Let $\alpha \in \mathbb{R}^{N+1}$, we deduce from (3.3) that

$$\alpha' U \alpha = \sum_{j=0}^{n_N} \left(\sum_{p=0}^{N} \alpha_p a_{n_p-j} \right)^2.$$

Thus, $\alpha' U \alpha = 0$ is equivalent to $\sum_{p=0}^{N} \alpha_p a_{n_p-j} = 0$ for $j = 0,\dots,n_N$. Taking successively $j = n_N, n_{N-1},\dots,n_0$ and using the fact that $a_0 = -1$, we obtain that $\alpha_N = \alpha_{N-1} = \cdots = \alpha_0 = 0$. Lastly, the expression of $\mathrm{var}(X_0 - \hat{X}'_0)$ is obtained from (3.5) where $k = 0$:

$$\mathrm{var}(X_0 - \hat{X}'_0) = \langle X_0 - \hat{X}'_0, X_0 \rangle = \sigma^2 \sum_{p=0}^{N} \psi_p \delta_{n_p} = \sigma^2 \psi_0. \qquad \square$$

Remark 3.1. (a) Formula (3.4) may be obtained using Pourahmadi (1994, Theorem 1). If $\hat{X}'_0 = X_0$, then $X_0 \in H'_{-1} \subset H_{-1}$, and thus, (X_k) is deterministic. Conversely, if (X_k) is deterministic, then $X_0 \in H_{-n_N-1} \subset H'_{-1}$, and thus, $\hat{X}'_0 = X_0$. Therefore, $\mathrm{var}(X_0 - \hat{X}'_0) > 0$ iff $\sigma^2 > 0$.

(b) Assume that $n_1 = 1, \dots, n_N = N$, so that $\hat{X}'_0$ is the $(N+1)$-step predictor of X_0. It follows from (3.3) that $U = A'A$, and since $A^{-1} = -C$, (3.2) is equivalent to

$$A(\psi_0, \psi_1, \dots, \psi_N)' = -(c_0, c_1, \dots, c_N)'. \tag{3.10}$$

We deduce from (3.1) and (3.10) that

$$X_0 - \hat{X}'_0 = -\sum_{j=0}^{N} \left(\sum_{p=0}^{N} \psi_p a_{p-j} \right) \varepsilon_{-j} = \sum_{j=0}^{N} c_j \varepsilon_{-j},$$

which is a well-known relation, see for instance Brockwell and Davis (1991, p. 189).

Since the innovation process (ε_k) is not directly observable, formula (3.1) cannot be used to calculate $\hat{X}'_0$ unless one can express the innovation ε_k in terms of the observations X_{k-i} for $i \geqslant 0$. This is equivalent to finding a mean square convergent AR series representation for $\hat{X}'_0$ in the time domain. In the following theorem, we establish such a representation.

Theorem 3.2. *Let (X_k) be a purely nondeterministic stationary process with the AR parameters (a_k). The predictor $\hat{X}'_0$ has an AR representation for any finite set of missing data iff (X_k) has the AR representation* (2.5). *In this case, the AR representation of $\hat{X}'_0$ is unique and is given by*

$$\hat{X}'_0 = \sum_{k \in \mathbb{N} \setminus M} h_k X_{-k}, \tag{3.11}$$

where

$$h_k = \delta_k - \sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p \wedge k} a_{n_p - j}\, a_{k-j}, \quad k \in \mathbb{N} \tag{3.12}$$

and the coefficients (ψ_p) are defined in Theorem 3.1.

Proof. First, note that if the AR representation of $\hat{X}'_0$ exists, it is unique because if (X_k) is nondeterministic and $\sum_{k=0}^{\infty} g_k X_{-k} = 0$, then $g_k = 0$ for all $k \geqslant 0$. Now, if $\hat{X}'_0$ has an AR representation for any finite set of missing data, $\hat{X}_0$ possesses the AR representation $\hat{X}_0 = \sum_{i=1}^{\infty} a_i X_{-i}$. By stationarity, we get $\hat{X}_k = \sum_{i=1}^{\infty} a_i X_{k-i}$ for any $k \in \mathbb{Z}$, which is equivalent to (2.5). Conversely, assume that (X_k) has the AR representation (2.5). By replacing ε_{-j} in (3.1) by its expression deduced from (2.5), we get

$$\hat{X}'_0 = X_0 - \sum_{k=0}^{\infty} \left(\sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p \wedge k} a_{n_p - j} a_{k-j} \right) X_{-k} = \sum_{k=0}^{\infty} h_k X_{-k},$$

where the coefficients (h_k) are defined by (3.12). Thanks to (3.2) and (3.3), we have $h_k = 0$ for any $k \in M$, and thus we obtain (3.11). □

Remark 3.2. (a) Under the condition that the spectral density f of (X_k) is bounded almost everywhere and f^{-1} is integrable with respect to $d\lambda$, it was established in Masani (1960, Theorem 5.2) that the $(N+1)$-step predictor of X_0 has an AR representation (3.11) with

$$h_k = \begin{cases} 0 & \text{if } 0 \leqslant k \leqslant N, \\ -\sum_{j=N+1}^{k} c_j a_{k-j} & \text{if } k \geqslant N+1. \end{cases} \tag{3.13}$$

It was shown in Bloomfield (1985, Theorem 1) that in fact, the $(N+1)$-step predictor has an AR representation once (X_k) has the representation (2.5). Theorem 3.2 generalizes this result to the prediction with any finite set of missing data. Note that taking $n_p = p$ in (3.12) and using (3.10), we obtain

$$h_k = \delta_k + \sum_{j=0}^{N\wedge k} c_j a_{k-j}, \quad k \in \mathbb{N} \tag{3.14}$$

and using (2.4), it is readily seen that (3.14) is equivalent to (3.13).

(b) The AR representation (3.11) was established in Bondon (2000, Corollary 1) under two different conditions, namely (i) f is bounded almost everywhere and f^{-1} is integrable with respect to $d\lambda$, (ii) the angle between the subspaces H_0 and $\overline{\mathrm{sp}}\{X_i;\ i \geqslant 1\}$ is positive. Neither of these conditions implies the other, but both imply that f^{-1} is integrable with respect to $d\lambda$, or in other words that the process (X_k) is minimal. Since the series in (2.5) may converge in L_2 while (X_k) is not minimal, see Bloomfield (1985, Theorem 4), Theorem 3.2 generalizes Bondon (2000, Corollary 1).

(c) According to (3.12), $h_0 = 0$, and for any $k \geqslant 1$, we have

$$h_k = -\sum_{j=0}^{n_N} \alpha_j a_{k-j}, \tag{3.15}$$

where $\alpha_j = \sum_{p=0}^{N} \psi_p a_{n_p - j}$. Therefore, $h_k^2 \leqslant (\sum_{j=0}^{n_N} \alpha_j^2)(\sum_{j=0}^{n_N} a_{k-j}^2)$ for any $k \in \mathbb{N}$, and if (X_k) is minimal, we have

$$\sum_{k=0}^{\infty} h_k^2 \leqslant (n_N + 1)\left(\sum_{j=0}^{n_N} \alpha_j^2\right) \sum_{k=0}^{\infty} a_k^2 < \infty.$$

(d) If $\sum_{k=0}^{\infty} |a_k| < \infty$, then the series in (2.5) converges in L_2, see for instance Brockwell and Davis (1991, Proposition 3.1.1). Thus, $\hat{X}_0'$ has the AR representation (3.11), and we deduce from (3.15) that

$$\sum_{k=0}^{\infty} |h_k| \leqslant \left(\sum_{j=0}^{n_N} |\alpha_j|\right) \sum_{k=0}^{\infty} |a_k| < \infty.$$

(e) For a given set of missing data, $\hat{X}_0'$ may have an AR representation while (X_k) does not. When this set is infinite, this property is obvious (assume for instance that $\{X_k;\ k \leqslant -2\}$ are missing so that $\hat{X}_0'$ involves only X_{-1}). An example with a finite set of missing data is given in Remark 4.2.

According to (3.4), the increase in variance of the prediction error of X_0 due to the missing data $X_{-n_1},\dots,X_{-n_N}$, is equal to $\sigma^2(\psi_0-1)$. In the following theorem, we characterize the processes for which the missing observations do not affect the prediction of X_0.

Theorem 3.3. *Let* (X_k) *be a nondeterministic stationary process with the AR parameters* (a_k). *Then* $\hat{X}_0'=\hat{X}_0$ *iff* $a_{n_i}=0$ *for* $i=1,\dots,N$.

Proof. It results from (3.1) that $\hat{X}_0'=\hat{X}_0$ iff

$$-\varepsilon_0=\left(\sum_{p=0}^{N}\psi_p a_{n_p}\right)\varepsilon_0+\sum_{j=1}^{n_N}\left(\sum_{p=1}^{N}\psi_p a_{n_p-j}\right)\varepsilon_{-j},$$

which is equivalent to

$$\sum_{p=0}^{N}\psi_p a_{n_p}=-1, \tag{3.16}$$

$$\sum_{p=1}^{N}\psi_p a_{n_p-j}=0,\quad j=1,\dots,n_N. \tag{3.17}$$

Taking successively $j=n_N,n_{N-1},\dots,n_1$ in (3.17) and using the fact that $a_0=-1$, we find that (3.16) and (3.17) are equivalent to $(\psi_0,\psi_1,\dots,\psi_N)=(1,0,\dots,0)$. Since the first column of U is $(1,-a_{n_1},\dots,-a_{n_N})'$, $(1,0,\dots,0)$ is the solution of (3.2) iff $a_{n_i}=0$ for $i=1,\dots,N$. □

Remark 3.3. (a) Under the assumption that (X_k) has the AR representation (2.5), Theorem 3.3 is easily proved as follows. If $\hat{X}_0'=\hat{X}_0$, we deduce from (3.11) and (2.5) that $\sum_{k\in\mathbb{N}\setminus M} h_kX_{-k}=\sum_{k=1}^{\infty} a_kX_{-k}$, which implies that $a_k=0$ for all $k\in M\setminus\{0\}$. Conversely, if $a_k=0$ for all $k\in M\setminus\{0\}$, $\hat{X}_0=\sum_{k\in\mathbb{N}\setminus M} a_kX_{-k}$, and thus, $\hat{X}_0\in H_{-1}'$ and $X_0-\hat{X}_0\perp H_{-1}\supset H_{-1}'$. Therefore, $\hat{X}_0=\hat{X}_0'$.

(b) Let $K=\mathrm{sp}\{X_{-k};1\leqslant k\leqslant n_1-1\}$ and $\hat{X}_{0,n_1}$ be the projection of X_0 onto K. Since $K\subset H_{-1}'$, we have $\mathrm{var}(X_0-\hat{X}_0')\leqslant\mathrm{var}(X_0-\hat{X}_{0,n_1})$. Since $\hat{X}_0-\hat{X}_0'\in H_{-1}$, $X_0-\hat{X}_0\perp\hat{X}_0-\hat{X}_0'$, and therefore, $\mathrm{var}(X_0-\hat{X}_0')=\sigma^2+\mathrm{var}(\hat{X}_0-\hat{X}_0')$. Hence, $\mathrm{var}(\hat{X}_0-\hat{X}_0')\leqslant\mathrm{var}(X_0-\hat{X}_{0,n_1})-\sigma^2$. When $n_1\to\infty$, $\hat{X}_{0,n_1}$ converges to $\hat{X}_0$ in L_2, and thus, $\hat{X}_0'$ converges to $\hat{X}_0$ for any stationary process.

4. Examples

Example 4.1. Assume that only the observation X_{-m}, $m>0$, is missing. We deduce from (3.3) that

$$U=\begin{pmatrix}1 & -a_m\\ -a_m & S_m\end{pmatrix},$$

where $S_m = \sum_{i=0}^{m} a_i^2$, so that the solution of (3.2) is $(\psi_0, \psi_1) = S_{m-1}^{-1}(S_m, a_m)$. From (3.1) and (3.4), we get

$$X_0 - \hat{X}'_0 = \varepsilon_0 - \frac{a_m}{S_{m-1}} \sum_{j=1}^{m} a_{m-j}\varepsilon_{-j},$$

$$\mathrm{var}(X_0 - \hat{X}'_0) = \sigma^2 \left(1 + \frac{a_m^2}{S_{m-1}}\right). \tag{4.1}$$

If (X_k) is purely nondeterministic and has the AR representation (2.5), we deduce from Theorem 3.2 that $\hat{X}'_0$ has the AR representation (3.11) where

$$h_k = a_k - \frac{a_m}{S_{m-1}} \sum_{j=1}^{m \wedge k} a_{m-j} a_{k-j}, \quad k \geqslant 1.$$

According to Theorem 3.3, $\hat{X}'_0 = \hat{X}_0$ iff $a_m = 0$. Finally, we deduce from Remark 3.3(b) and (4.1) that $\lim_{m\to\infty} a_m^2 S_{m-1}^{-1} = 0$ for any nondeterministic stationary process. Since $S_{m-1} \geqslant 1$, this result is obvious when $\lim_{m\to\infty} a_m = 0$, which is, in particular, the case when (X_k) has an AR representation and when (X_k) is minimal.

Example 4.2. Assume that (X_k) is a causal AR(r) process,

$$X_k = \sum_{i=1}^{r} a_i X_{k-i} + \varepsilon_k, \quad 1 \leqslant r < \infty. \tag{4.2}$$

Then
(a) $\hat{X}'_0$ has the AR representation (3.11) with $h_k = 0$ for all $k > n_N + r$,
(b) $\hat{X}'_0 = \hat{X}_0 = \sum_{i=1}^{r} a_i X_{-i}$ iff $a_k = 0$ for all $k \in \{(n_i)_{i=1,\dots,N} \mid n_i \leqslant r\}$.

Proof. Since (X_k) is causal, (ε_k) is the innovation process of (X_k) and (4.2) is the AR representation of (X_k). Hence, (a) results from Theorem 3.2 and (b) follows from Theorem 3.3. □

Remark 4.1. (a) According to Example 4.2(a), if (X_k) is a causal AR(r) process, then $\hat{X}'_0$ is the best linear mean square predictor of X_0 based on the finite past $\{X_{-k};\ k \in \{1,\dots,n_N + r\} \setminus M\}$. This property generalizes a result which is well known when the past is complete.

(b) Take $r = 1$ in (4.2) and set $a = a_1$ with $|a| < 1$. If $n_1 > 1$, we have $\hat{X}'_0 = \hat{X}_0 = aX_{-1}$. If $n_k = k$ for $1 \leqslant k < j$ and $n_j > j$ for some j, we have $a^j X_{-j} \in H'_{-1}$ and $X_0 - a^j X_{-j} = \sum_{k=0}^{j-1} a^k \varepsilon_{-k} \perp H_{-j} \supset H'_{-1}$. Accordingly, $\hat{X}'_0 = a^j X_{-j}$ and $\mathrm{var}(X_0 - \hat{X}'_0) = \sigma^2(1 - a^{2j})(1 - a^2)^{-1}$.

Example 4.3. Assume that (X_k) is an MA(r) process,

$$X_k = \varepsilon_k + \sum_{i=1}^{r} c_i \varepsilon_{k-i}, \quad 1 \leqslant r < \infty. \tag{4.3}$$

If at least r observations with consecutive indices are missing, i.e., $\{i, i+1, \dots, i+r-1\} \subset M$ for some $i \geqslant 1$, then $\hat{X}'_0$ is the best linear mean square predictor of X_0 based on the finite past $\{X_{-k};\ k \in \{1,\dots,i-1\} \setminus M\}$.

Proof. Set $K = \text{sp}\{X_{-k};\ k \in \{1,\dots,i-1\}\setminus M\}$ and $L = \text{sp}\{X_{-k};\ k \geqslant i+r, k \notin M\}$. We have $H'_{-1} = \overline{K+L} = \overline{K+\bar{L}}$. Since K is a finite dimensional subspace of L_2 and $\bar{L}$ is a closed subspace of L_2, $K+\bar{L}$ is closed in L_2, see for instance Bourbaki (1981, I, p. 15, Corollary 4). Hence, $H'_{-1} = K+\bar{L}$. Since (X_k) is an MA(r) process, $\langle X_k, X_n \rangle = 0$ when $|k-n| > r$, which implies that $K \perp \bar{L}$. Therefore, $\hat{X}'_0$ is the sum of the projections of X_0 onto K and $\bar{L}$. Since $i \geqslant 1$, $X_0 \perp H_{-i-r} \supset \bar{L}$. Hence, $\hat{X}'_0$ is the projection of X_0 onto K. □

Remark 4.2. Take $r=1$ in (4.3) and set $c = -c_1$ with $|c| \leqslant 1$. Then (ε_k) is the innovation process of (X_k) and the AR parameters of (X_k) are $a_k = -c^k$, $k \geqslant 0$. Thus, $a_k \neq 0$ for all $k \geqslant 0$ and we deduce from Theorem 3.3 that $\hat{X}'_0 \neq \hat{X}_0$ for any set of missing data. If $|c| < 1$, we have $\hat{X}_0 = -\sum_{k=1}^{\infty} c^k X_{-k}$. If $|c| = 1$, (X_k) does not have an AR representation since $a_k \nrightarrow 0$ when $k \to \infty$, but it may be shown that $\hat{X}_0 = -\lim_{k\to\infty} \sum_{i=1}^{k} c^i (1 - i/k) X_{-i}$. Therefore, for $|c| \leqslant 1$, $\hat{X}_0$ involves every X_k for $k \leqslant -1$, whereas according to Example 4.3, $\hat{X}'_0$ is the predictor of X_0 based on the finite past $\{X_{-1},\dots,X_{-n_1+1}\}$. If $n_1 = 1$, $\hat{X}'_0 = 0$. If $n_1 \geqslant 2$, $\hat{X}'_0 = \sum_{k=1}^{n_1-1} \alpha_k X_{-k}$, where $\alpha_k = -c^k S_{n_1-k} S_{n_1}^{-1}$ with $S_n = \sum_{i=0}^{n-1} c^{2i}$, $n \geqslant 1$, and

$$\text{var}(X_0 - \hat{X}'_0) = \langle X_0 - \hat{X}'_0, X_0 \rangle = \text{var}(X_0) - \sum_{k=1}^{n_1-1} \alpha_k \langle X_{-k}, X_0 \rangle.$$

Since $X_k = \varepsilon_k - c\varepsilon_{k-1}$, we have $\text{var}(X_0) = \sigma^2(1+c^2)$, $\langle X_{-1}, X_0 \rangle = -c\sigma^2$, and $\langle X_{-k}, X_0 \rangle = 0$ if $k > 1$. Therefore,

$$\text{var}(X_0 - \hat{X}'_0) = \sigma^2(1 + c^2 + \alpha_1 c) = \sigma^2 \left(1 + c^2 - c^2 \frac{S_{n_1-1}}{S_{n_1}}\right) = \sigma^2 \left(1 + \frac{c^{2n_1}}{S_{n_1}}\right).$$

References

Bloomfield, P., 1985. On series representations for linear predictors. Ann. Probab. 13 (1), 226–233.

Bondon, P., 2000. Représentation autorégressive du prédicteur à passé infini incomplet d'une série chronologique stationnaire. C. R. Acad. Sci. Paris Sér. I Math. 330 (10), 915–920.

Bourbaki, N., 1981. Éléments de Mathématique, Espaces Vectoriels Topologiques, Chapitres 1 à 5. Masson, Paris.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods., 2nd Edition. Springer, New York.

Cheng, R., Pourahmadi, M., 1997. Prediction with incomplete past and interpolation of missing values. Statist. Probab. Lett. 33, 341–346.

Grenander, U., Rosenblatt, M., 1954. An extension of a theorem of G. Szegö and its application to the study of stochastic processes. Trans. Am. Math. Soc. 76, 112–126.

Kolmogorov, A.N., 1941. Stationary sequences in Hilbert space. Bull. Moscow State Univ. 2 (6), 1–40.

Masani, P., 1960. The prediction theory of multivariate stochastic processes, III. Acta Math. 104, 141–162.

Pourahmadi, M., 1994. Two prediction problems and extensions of a theorem of Szegö. Bull. Iranian Math. Soc. 19 (2), 1–12.