Prediction of a stationary process with missing data

Pascal Bondon
CNRS UMR 8506
3 rue Joliot-Curie
91192 Gif-sur-Yvette, France
bondon@lss.supelec.fr

1. Main result

Let (X_k) be a zero-mean purely nondeterministic weakly stationary stochastic process with spectral density f. We denote by $\widehat{X}_0$ the best linear mean-square predictor of X_0 based on the past $\{X_k; k \leq -1\}$. In order to be computable in practice, $\widehat{X}_0$ should have a mean-convergent autoregressive (AR) representation,

$$\widehat{X}_0 = \sum_{k=1}^{\infty} a_k X_{-k}. \tag{1}$$

There are two well-known sufficient conditions for the existence (and the uniqueness) of (1) :

C1 f is bounded almost everywhere and f^{-1} is integrable with respect to Lebesgue measure on $(-\pi, \pi]$, Akutowicz (1957);

C2 $f = e^{u+\tilde{v}}$ where u, v are bounded real-valued functions with $\|v\|_\infty < \pi/2$, and $\tilde{v}$ is the harmonic conjugate of v, Helson and Szegö (1960).

C2 allows to relax the boundedness of f, but implies that f^{-p} is integrable for some $p > 1$. Hence, neither of these conditions implies the other and can be necessary for the existence of (1). A more general sufficient condition which includes C1 and C2 as special cases was derived by Bloomfield (1985). It should be noted that contrarily to C1 and C2, Bloomfield' condition does not imply that f^{-1} is integrable.

In various situations, all the data X_k for $k \leq -1$ are not observed, either because some X_k are missing or because prudent statistical analysis of the data leads to discard some observed suspect values. A particular case is the one where the N last data preceding X_0 are absent. In this case, the predictor of X_0 is the $(N+1)$-step predictor, and it was shown by Bloomfield (1985) that if (1) exists, then this predictor possesses also an AR representation which is obtained by means of well-known relations, given for instance in Akutowicz (1957).

In this paper, we consider the case of an arbitrary finite set of missing values in the past $\{X_{-n_1}, \ldots, X_{-n_N}\}$. We establish a formula for the AR representation of the predictor $\widehat{X}_0'$ of X_0 based on the incomplete past $\{X_k; k \leq -1, k \neq -n_1, \ldots, -n_N\}$. Specifically, the following theorem is proved, see Bondon (2000).

Theorem. *If C1 or C2 is satisfied, we have*

$$\widehat{X}_0' = \sum_{\substack{k=1 \\ k\neq n_1,\ldots,n_N}}^{\infty} h_k X_{-k} \tag{2}$$

where

$$h_k = -\sum_{p=0}^{N} \psi_p \sum_{j=0}^{n_p\wedge k} a_{n_p-j}\, a_{k-j},$$

$n_0 = 0$, $a_0 = -1$, *and the coefficients* (ψ_p) *are given by*

$$U(\psi_0, \psi_1, \ldots, \psi_N)' = (1, 0, \ldots, 0)',$$

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

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

The prediction error variance is $\mathrm{var}(X_0 - \widehat{X}_0') = \psi_0 \exp(\int_{-\pi}^{\pi} \log f(\lambda) d\lambda)$.

2. Extensions

The existence of (2) in the result presented above assumes that f^{-1} is integrable. But, in the case where $n_1 = 1, \ldots, n_N = N$, Bloomfield (1985) has shown that this condition is dispensable. Therefore, (2) should exist certainly under weaker conditions.

REFERENCES

Akutowicz, E. J. (1957). On an explicit formula in linear least squares prediction. *Math. Scand.*, **5**, 261–266.

Bloomfield, P. (1985). On series representations for linear predictors. *Ann. Probab.*, **13**, 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**, 915–920.

Helson, H. and Szegö, G. (1960). A problem in prediction theory. *Ann. Mat. Pura Appl.*, **51**, 107–138.

RÉSUMÉ

Nous établissons la représentation AR du meilleur prédicteur linéaire en moyenne quadratique d'une série stationnaire basé sur toutes les observations du passé à l'exception d'un nombre fini d'entre elles dont les indices temporels sont quelconques. Ce résultat est obtenu sous les hypothèses suffisantes classiques d'existence d'une représentation AR pour le prédicteur basé sur tout le passé.