INRIA Conf. Nice 1984 C21

ON THE USE OF VARIOGRAMS IN LEVINSON PREDICTORS

Michel Gevers
Department of Systems Engineering
Australian National University
Canberra, Australia

Abstract

We consider the prediction of stationary stochastic processes with non-zero mean. When the covariance of the process is known, but the mean is not, the classical approach is to first estimate the mean from the past data, and then apply an optimal predictor to the zero-mean residuals. Bastin and Henriet [1] showed that an alternative was to use a predictor based on "variograms" rather than covariance information, thus avoiding the estimation of the mean. We show here that the two predictors are identical, when the unknown mean is replaced by its minimum variance estimate. We also argue that in a practical situation, where the covariance is unknown, it is wiser to use the variogram approach.

1 Introduction

We consider a discrete scalar stochastic process $\{y_0, y_1, \ldots, y_t, \ldots\}$ with constant but not necessarily zero mean:

$$E\{y_t\} = m \tag{1}$$

We define the "variogram" of the process as the semi-variance of the increments

$$\gamma(t,t-\tau) = \tfrac{1}{2} E\{(y_t - y_{t-\tau})^2\} \tag{2}$$

We shall assume throughout that the variogram is stationary, i.e.

C21

$$\gamma(t,t-\tau) = \gamma(\tau) = \tfrac{1}{2} E\{y_t - y_{t-\tau})^2\} \tag{3}$$

A process with the properties (1) and (3) is called intrinsic (see e.g. [2]). If the process is wide-sense stationary, then the (auto)covariance can be defined as

$$R(\tau) = E\{(y_t - m)(y_{t-\tau} - m)\} \tag{4}$$

For a wide-sense stationary process, the covariance function and the variogram are related by

$$\gamma(\tau) = \sigma^2 - R(\tau) \text{ , with } \sigma^2 = R(0) \tag{5}$$

Note that an intrinsic random process is always wide-sense stationary, but the converse is not true. The class of intrinsic processes is larger and includes Wiener processes as a special case. Except when specifically stated, we shall from now on assume that the process $\{y_t\}$ is wide-sense stationary.

In this paper we shall derive different minimum variance unbiased (MVU) expressions for (d+1)-step ahead predictors of the $\{y_t\}$ process under a variety of assumptions. We shall not make any assumption about the existence of an underlying finite-dimensional model. Hence we shall consider Levinson predictors with growing memory: the predicted value at time t=N uses all available past data. In Section 2 we shall briefly recall the expressions of the classical Levinson predictor (CLP) for y_{N+d}, given $\{y_0, y_1, \ldots, y_{N-1}\}$, under the assumptions that the constant mean m and the covariance (or the variogram) are known. In Section 3 we consider the case where the mean is unknown. Two different MVU predictors can be used in this case:

1) one can compute an unbiased estimate $\hat{m}$, and then replace the mean by its estimate $\hat{m}$ in the expressions of the CLP of Section 2. This predictor will be called Approximate Levinson Predictor (ALP);

2) alternatively, one can use a MVU predictor of the form $\sum_{i=1}^{N} b_i y_{N-i}$, in which the b_i are computed from the variogram function. This predictor, derived by Bastin and Henriet [1], who called it the modified Levinson predictor, does not involve the mean or its estimate in any way.

In Section 4 we show that the two predictors are exactly identical if the MVU estimate is used for $\hat{m}$ in the ALP. This may appear as a surprising result, considering that the first involves an estimate of the mean, while the other does

not. Our result corrects a result of Bastin and Henriet, who claimed that the MLP leads to a strictly smaller prediction error variance than the ALP. We shall further derive a number of interesting expressions for the MLP and its prediction error variance, and show that all these quantities can be expressed in terms of either the covariance function or the variogram. In Section 5, we shall give a number of reasons why the MLP, written in terms of the variogram, is to be preferred in practical situations, where both the mean and the covariance function are unknown, and have to be estimated from the data.

2 Processes with known mean and covariance function

When the mean is known, the classical (d+1)-step ahead Levinson predictor (CLP) for y_{N+d} given $y_0,\dots,y_{N-1}$ has the form (see e.g. [3]):

$$\hat{y}_{N+d} = m + \sum_{i=1}^{N} a_i\,(y_{N-i}-m) = m + a^T(Y-mU) \tag{6}$$

where : $a^T = (a_1,\dots,a_N)$, $U^T = (1,1,\dots 1)$, $Y^T = (y_{N-1},\dots,y_0)$. The prediction error $\tilde{y}_{N+d} \triangleq \hat{y}_{N+d} - y_{N+d}$ is unbiased. The minimization of the prediction error variance $E\{(\tilde{y}_{N+d})^2\}$ with respect to the a_i leads to the following system of N linear equations:

$$Ra = R_0 \tag{7}$$

where

$$R = E\{(Y-mU)(Y-mU)^T\} \tag{8a}$$

$$R_0 = E\{(Y-mU)(y_{N+d}-m)\} \tag{8b}$$

Using (5), we can also write:

$$R = \sigma^2 UU^T - \Gamma\ ,\ R_0 = \sigma^2 U - \Gamma_0 \tag{9}$$

where

$$\Gamma = \begin{vmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(N-1) \\ \gamma(1) & \gamma(0) & & \gamma(N-2) \\ & & \ddots & \\ \gamma(N-1) & \gamma(N-2) & \dots & \gamma(0) \end{vmatrix}\ ,\ \Gamma_0 = \begin{vmatrix} \gamma(d+1) \\ \gamma(d+2) \\ \vdots \\ \gamma(d+N) \end{vmatrix} \tag{10}$$

Note that $\gamma(0) = 0$. Substituting (9) in (7) yields an alternative system of equations for a in terms of the variogram:

$$\begin{vmatrix} \Gamma & U \\ U^T & \frac{1}{\sigma^2} \end{vmatrix} \begin{vmatrix} a \\ \alpha\sigma^2 \end{vmatrix} = \begin{vmatrix} \Gamma_0 \\ 1 \end{vmatrix} \tag{11}$$

where $\alpha \triangleq 1 - a^T U = 1 - \sum_{i=1}^{N} a_i$. (12)

Using matrix and vector notations, (11) can be rewritten as $A_1\,\lambda_1 = c$, with obvious definitions for A_1, λ_1 and c. The optimal prediction error variance, $V_c \triangleq E\{(\tilde{y}_{N+d})^2\}$, can be written in a number of ways:

$$\begin{aligned} V_c &= \sigma^2 - a^T R_0 = \alpha\sigma^2 + a^T\Gamma_0 \\ &= \sigma^2 - R_0^T R^{-1} R_0 \\ &= c^T A_1^{-1} c \end{aligned} \tag{13}$$

3 Processes with unknown mean and known second order statistics

We now consider the case where the constant mean m is unknown, but where the covariance function $R(\tau)$ or the variogram $\gamma(\tau)$ are assumed known. Recall that they are related by (5).

3.1 The approximate Levinson predictor

The most obvious strategy is to replace the mean in (6) by a linear unbiased estimate based on the past data:

$$\hat{m} = \sum_{i=1}^{N} \zeta_i y_{N-i} = \zeta^T Y \tag{14}$$

with the unbiasedness condition:

$$\zeta^T U = 1 \tag{15}$$

The CLP of (6) is then replaced by the approximate Levinson predictor (ALP):

$$y^*_{N+d} = \hat{m} + a^T(Y-\hat{m}U) \tag{16}$$

where a is the solution of (7), as before. This estimator was called approximate classical Levinson predictor (ACLP) in [1]. The prediction error can be written

$$\tilde{y}_{N+d} \triangleq y^*_{N+d} - y_{N+d} = m + a^T(Y-mU) - y_{N+d} + \alpha(\hat{m}-m) \qquad (17)$$

The sum of the first two terms in (17) is the optimal CLP, so that the sum of the first three terms is the CLP prediction error. Therefore, the prediction error variance $V_a \triangleq E\{(y^*_{N+d} - y_{N+d})^2\}$ is:

$$V_a = V_c + \alpha^2\sigma^2_{\hat{m}} + 2\alpha E\{[m + a^T(Y-mU) - y_{N+d}][\hat{m}-m]\}$$

$$= V_c + \alpha^2\sigma^2_{\hat{m}} + 2\alpha E\{[a^T(Y-mU) - y_{N+d}][(Y-mU)^T\zeta]\}$$

since $\hat{m}$ is unbiased. $\sigma^2_{\hat{m}}$ denotes the variance of $\hat{m}$. The third term in the above expression is zero by (7). Therefore

$$V_a = V_c + \alpha^2\sigma^2_{\hat{m}} \qquad (18)$$

It follows from (18) that V_a will be minimum if $\sigma^2_{\hat{m}}$ is minimum, i.e. if $\hat{m}$ is chosen as the MVU estimate of m. The corresponding estimator will be called MVALP, for Minimum Variance Approximate Levinson Predictor. The MVU estimate for $\hat{m}$ is obtained as follows:

$$\sigma^2_{\hat{m}} = E\{(\hat{m} - m)^2\} = E\{[\zeta^T(Y-mU)]^2\}$$

$$= \zeta^T R\,\zeta = \sigma^2 - \zeta^T\Gamma\,\zeta \qquad (19)$$

Minimizing $\zeta^T R\zeta$ w.r.t. ζ subject to (15) yields

$$\zeta = (U^T R^{-1} U)^{-1} R^{-1} U \qquad (20)$$

After lengthy manipulations, using (5), ζ can also be expressed in terms of $\gamma(\tau)$:

$$\zeta = (U^T\Gamma^{-1}U)^{-1}\Gamma^{-1}U \qquad (21)$$

The corresponding minimum error variance for the mean estimate will be denoted $\underline{\sigma}^2_{\hat{m}}$:

$$\underline{\sigma}^2_{\hat{m}} = (U^T R^{-1} U)^{-1} = \sigma^2 - (U^T\Gamma^{-1}U)^{-1} \qquad (22)$$

The expression (18) shows that the ALP yields a larger error variance than the

CLP. This is to be expected, since the CLP is the optimal estimator. The two error variances coincide when $\sigma^2_{\hat{m}} \to 0$, as should be expected. We shall denote by $\underline{V}_a$ the prediction error variance of the MVALP. Using

$$\alpha = 1 - a^T U = 1 - R_0^T R^{-1} U \qquad (23)$$

together with (22) and (13), we can write

$$\underline{V}_a = V_c + \alpha^2(U^T R^{-1} U)^{-1} \qquad (24a)$$

$$= V_c + (1 - R_0^T R^{-1} U)^2 (U^T R^{-1} U)^{-1} \qquad (24b)$$

$$= \sigma^2 - R_0^T R^{-1} R_0 + (1 - R_0^T R^{-1} U)^2 (U^T R^{-1} U)^{-1} \qquad (24c)$$

Note that this last expression is entirely in terms of the covariance function of $\{y_t\}$.

3.2 The modified Levinson predictor (MLP)

When the mean m is unknown, an alternative strategy is to seek the MVU predictor of the following form:

$$\hat{y}_{N+d} = \sum_{i=1}^{N} b_i y_{N-i} = b^T Y \qquad (25)$$

together with the constraint, imposed by unbiasedness, that

$$\sum_{i=1}^{N} b_i = 1 \quad , \text{ i.e. } \quad b^T U = 1 \qquad (26)$$

Minimization of $E\{(\tilde{y}_{N+d})^2\}$ w.r.t. b subject to (26) yields the following linear system of (N+1) equations:

$$\begin{vmatrix} R & -U \\ U^T & 0 \end{vmatrix} \begin{vmatrix} b \\ \mu \end{vmatrix} = \begin{vmatrix} R_0 \\ 1 \end{vmatrix} \qquad (27)$$

where μ is a Lagrange coefficient. Alternatively, using (5), we get

$$\begin{vmatrix} \Gamma & U \\ U^T & 0 \end{vmatrix} \begin{vmatrix} b \\ \mu \end{vmatrix} = \begin{vmatrix} \Gamma_0 \\ 1 \end{vmatrix} \qquad (28a)$$

or, with obvious matrix and vector notations,

$$A_2\,\lambda_2 = c \qquad (28b)$$

The prediction error variance can be written in a number of ways

$$V_m = \sigma^2 + b^T R\, b - 2\, b^T R_0 = 2\, b^T\Gamma_0 - b^T\Gamma\, b \qquad (29a)$$

$$= \mu + \sigma^2 - b^T R_0 = \mu + b^T\Gamma_0 \qquad (29b)$$

$$= c^T A_2^{-1}\, c$$

Note that the last expression is the only one that does not involve **b** or μ. An expression similar to (29c), involving only the covariance function, can of course be obtained from (27).

The predictor MLP was proposed by Bastin and Henriet [1], who showed that

$$V_m = V_c + \alpha^2(U^T R^{-1} U)^{-1} \tag{30}$$

with α defined by (12). This, together with (24a), shows that

$$V_m = \underline{V}_a \tag{31}$$

In fact, we show in the next section that the MVALP and the MLP are identical, which of course explains (31). In the process, we shall derive a number of interesting expressions for **b** and μ.

4 The MVALP and the MLP are identical

Proof: It follows from (16) and (14) that the MVALP can be written:

$$\begin{aligned} y^*_{N+d} &= (1 - \mathbf{a}^T U)\hat{m} + \mathbf{a}^T Y \\ &= (\alpha\, \zeta^T + \mathbf{a}^T) Y \end{aligned} \tag{32}$$

Therefore, we need to prove that

$$\mathbf{a} + \alpha\, \zeta = \mathbf{b} \tag{33}$$

where **a**, ζ, **b** are solutions of (7), (20) and (27) respectively, and with α given by (12).

The solution of (27) is unique, because it is the vector **b** that minimizes $E\{[y_{N+d}]^2\} = b^T R b + \sigma^2 - 2b^T R_0$ subject to (26). It is also trivial to see that $\mathbf{a}+\alpha\zeta$ satisfies the last equation of (27). Therefore, to prove (33), it remains to be shown that there exists a μ such that

$$R(\mathbf{a} + \alpha\, \zeta) - \mu U = R_0 \tag{34}$$

It follows immediately from (7), (20) and (22) that this is so for

$$\mu = \alpha\, \underset{\sim}{g}^2_m \tag{35}$$

This concludes the proof.

In the process of proving our main result, we have shown that the solution μ of (27) and (28) is $\alpha\, \underset{\sim}{g}^2_m$. This yields some closed form expressions for **b** and

some new expressions for μ in terms of either $R(\tau)$ or $\gamma(\tau)$. From (35), (27) and (28) we get

$$\mathbf{b} = R^{-1}(R_0 + \alpha \underset{\sim}{g}^2_m U) = \Gamma^{-1}(\Gamma_0 - \alpha \underset{\sim}{g}^2_m U) \tag{36}$$

Using (35), (12) and (22), we have

$$\mu = \frac{1 - U^T R^{-1} R_0}{U^T R^{-1} U} = \frac{U^T \Gamma^{-1} \Gamma_0 - 1}{U^T \Gamma^{-1} U} \tag{37}$$

The second equality in (37) is obtained by multiplying (36) to the left by U^T, and using (26), (12) and (22). This then allows us to write two expressions for **b** in terms of $R(\tau)$ or $\gamma(\tau)$ only:

$$\mathbf{b} = R^{-1} R_0 + \frac{1 - U^T R^{-1} R_0}{U^T R^{-1} U} R^{-1} U \tag{38a}$$

$$= \Gamma^{-1} \Gamma_0 + \frac{1 - U^T \Gamma^{-1} \Gamma_0}{U^T \Gamma^{-1} U} \Gamma^{-1} U \tag{38b}$$

The formulas (38a) and (38b) are remarkable because the expressions in terms of (R, R_0) and (Γ, Γ_0) are identical. Note that the first term of (38a) is **a** while the second term is $\alpha\, \zeta$. (Recall (33)). This is not so, however, for (38b), because $R^{-1} R_0 \neq \Gamma^{-1} \Gamma_0$. Finally, note that the expressions (37) and (38) can also be obtained from (27) using a matrix inversion lemma.

In Bastin and Henriet [1], it was claimed that $\underline{V}_a - V_m = \alpha^2 \sigma^2 > 0$, and that therefore the MLP was better than the MVALP. This was based on an erroneous expression for V_a: the correct expression is (18) above. Note, however, that the equality $V_m = \underline{V}_a$ holds only if the unbiased estimator $\hat{m}$ is chosen to be the minimum variance estimator. If the sample mean estimator

$$\hat{m} = \frac{1}{N} \sum_{i=1}^{N} y_{N-i} \tag{39}$$

is used in (16), then $V_m < V_a$.

5 Motivation for the use of the variogram and the MLP

We have shown that when the mean is unknown, but the covariance or variogram is known, the MVALP and the MLP are identical. We have also shown that the predictor coefficients of all three predictors that we discussed can be computed either from $R(\tau)$ or from $\gamma(\tau)$.

In practice, neither the mean nor the covariance or the variogram are known, and they have to be estimated from the data. Two issues must then be raised:

a) what predictor should one use?

b) should one estimate its coefficients using sample estimates of $R(\tau)$ or $\gamma(\tau)$?

Since we have shown that the MVALP and the MLP are identical, the practical choice is between:

- the ALP (16) with $\hat{m}$ replaced by its sample mean (39), or
- the MLP (25).

In both cases, the parameter vectors (**a** for (16), **b** for (25)) can be computed either by covariance equations ((7) and (38.a)) or by variogram equations ((11) and (38.b)), and in practice covariances and variograms would be replaced by their sample estimates. Of course, other predictors could be used in practice, such as least-squares predictors derived directly from the data rather than from sample covariance or variogram estimates, but our aim in this discussion is only to compare the two predictors just mentioned.

We give a number of reasons why we believe that the MLP should be used with its coefficients b_i computed from the variogram formulas (38.b).

1) The MLP does not require an estimate of the mean. The ALP would use an estimate of the sample mean which, in almost all cases, would not have minimum variance. We have shown that, with known covariances, this would lead to an ALP that has larger error variance than the MLP. It is likely that this result will still hold when exact covariances are replaced by estimated ones, although this has not been proved.

2) So far, we have assumed that the process is known to be stationary. In practice, however, this is not always easy to validate. Failure to detect that a process is nonstationary may lead to completely erroneous results if covariances are used, as the following example, due to Matheron [5], shows. Consider a Wiener process with variogram $\gamma(\tau)=|\tau|$, and suppose we have observed $\{y_0,\ldots,y_{N-1}\}$. Now suppose that the user believes thet the process he or she observes is wide-sense stationary, and that he or she estimates $\hat{m}$ by (39) and $\hat{R}(\tau)$ by:

$$\hat{R}(\tau) = \frac{1}{N-\tau} \sum_{t-\tau}^{N-1} (y_t - \hat{m})(y_{t-\tau} - \hat{m}) \tag{40}$$

Then it can be shown that, for $\tau > 0$,

$$E\{\hat{R}(\tau)\} = \frac{N^2-1}{3N} - \frac{4}{3}\tau + \frac{2}{3}\frac{\tau^2}{N} \tag{41}$$

It is a parabola; an apparent variance of $\hat{R}(0) = \hat{\sigma}^2 = \frac{N^2-1}{3N}$ will be found, whereas the true variance is infinite. The experimental variogram, on the other hand, is unbiased:

$$\hat{\gamma}(\tau) = \frac{1}{2(N-\tau)} \sum_{t-\tau}^{N-1} (y_t - y_{t-\tau})^2 \tag{42}$$

3) For intrinsic (but not necessarily wide-sense stationary) processes, the equations (28) and the estimator (38b) can be derived directly without going through the relation (5). Note that in such case, the covariance $R(\tau)$ and the variance σ^2 may not exist. Therefore, the MLP (38b) covers a wider class of random processes.

6 Concluding remarks

All these elements seem to indicate that the MLP should be preferred in

situations where the mean and the covariance function of a process are unknown. Of course, a definite answer can only be given by comparing the expected predictors when estimated means, covariances and variograms are used. A theoretical comparison appears to be very difficult.

Acknowledgements

The author wishes to thank Georges Bastin for interesting comments on an earlier draft of this paper.

References

1. G. Bastin and E. Henriet, "A minimum variance property of Levinson predictors using variograms". Systems and Control Letters, Vol.2, No.1, July 1982, pp. 25-29.
2. G. Matheron, "Les variables regionalisees et leur estimation", Masson Ed., 1965.
3. B.D.O. Anderson and J.B. Moore, "Optimal Filtering", Prentice Hall, 1979.
4. T.W. Anderson, "The Statistical Analysis of Time Series", Wiley, 1958.
5. G. Matheron, "The Theory of Regionalized Variables and its Applications", Cahiers du Centre de Morphologie Mathematique de Fontainebleau, 1971.