On Difference Operation in Linear Prediction

M. Yanagida and O. Kakusho
Osaka, Japan

Abstract

The relationship between the predictors obtained on differenced data and those on original data is derived for both the covariance method and the autocorrelation method. The physical interpretation of the derived relationship is discussed in connection with spectral enhancement.

1. Difference Operation in the Covariance Method

The linear prediction model for a sampled $\{y_n\}$ is expressed in the form

$$y_n \doteqdot \sum_{i=1}^{p} \alpha_i \, y_{n-i} \tag{1}$$

where α_i denotes the ith predictor and p is the prediction order. In matrix form, eq.(1) can be written as

$$\underset{m\times 1}{y} \doteqdot \underset{m\times p}{Y} \; \underset{p\times 1}{\alpha} \quad \text{or} \quad \underset{m\times(p+1)}{[\, y \mid Y \,]} \; \underset{(p+1)\times 1}{a} \doteqdot \underset{m\times 1}{0} \tag{2}$$

where

$$y = \begin{bmatrix} y_n \\ \vdots \\ y_j \\ \vdots \\ y_{n-m+1} \end{bmatrix} \qquad Y = \begin{bmatrix} y_{n-1} & \cdots & y_{n-p} \\ \vdots & & \vdots \\ \cdot & & \cdot \\ \vdots & & \vdots \\ y_{n-m} & \cdots & y_{n-m-p+1} \end{bmatrix} \quad \text{and} \quad a = \begin{bmatrix} -1 \\ \\ \alpha \end{bmatrix} = \begin{bmatrix} -1 \\ \alpha_1 \\ \cdot \\ \vdots \\ \alpha_p \end{bmatrix} \tag{3}$$

Vector a is called the augmented predictor vector for α. The normal equation for the covariance method is obtained by premultiplying both sides of eq. (2) by Y^T.

$$Y^T y = Y^T Y \hat{\alpha} \tag{4}$$

where the product matrix $Y^T Y$ represents the covariance matrix of $\{y_n\}$.

The least-squares solution for α or a is derived from the normal equation (4) and is expressed as

$$\hat{\alpha} = Y^+ y \quad \text{or} \quad \hat{a} = \begin{bmatrix} -1 \\ \hat{\alpha} \end{bmatrix} \begin{bmatrix} -1 \\ Y^+ y \end{bmatrix} \tag{5}$$

where Y^+ denotes the generalized inverse of Y and is usually identical to $(Y^T Y)^{-1} y^T$ except for rank deficient cases. That was the formulation of the covariance method by the generalized inverse of matrices.

In the same way, the linear prediction model for the differenced sequence of the form $\{y_n - w y_{n-1}\}$ is expressed as

$$y_n - w y_{n-1} \doteqdot \sum_{i=1}^{p} \beta_i \, (y_{n-i} - w y_{n-i-1}) \tag{6}$$

where β_i denotes the ith predictor for the differenced data. Equation 6 is written in matrix form as

$$\underset{m\times(p+1)}{[\, \Delta y \mid \Delta Y \,]} \; \underset{(p+1)\times 1}{b} \doteqdot \underset{m\times 1}{0} \tag{7}$$

where

$$\Delta y = \begin{bmatrix} y_n - w y_{n-1} \\ \vdots \\ y_{n-m+1} - w y_{n-m} \end{bmatrix}, \qquad b = \begin{bmatrix} -1 \\ \beta \end{bmatrix} = \begin{bmatrix} -1 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

and

$$\Delta Y = \begin{bmatrix} y_{n-1} - w y_{n-2} & \cdots\cdots & y_{n-p} - w y_{n-p-1} \\ \vdots & & \vdots \\ y_{n-m} - w y_{n-m-1} & \cdots\cdots & y_{n-m-p+1} - w y_{n-m-p} \end{bmatrix} \tag{8}$$

Here again, **b** is the augmented predictor vector for $\boldsymbol{\beta}$.

In order to investigate the relationship between $\hat{a}$ and $\hat{b}$, it will be reasonable to start with the same number of prediction equations on the same number of data samples. Standing on that point, we assume $\beta_p = 0$ with the intention of preparing the same number of differenced data that can provide the same number of prediction equations as those on the original data. Under this assumption the pth column of ΔY is arbitrary and eq.(7) can be modified as follows:

$$[y \,\vdots\, Y]\, W b \doteq 0 \tag{9}$$

where

$$\underset{(p+1)\times(p+1)}{W} = \begin{bmatrix} 1 & & & & 0 \\ -w & 1 & & & \\ & -w & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ 0 & & & -w & 1 \end{bmatrix} \tag{10}$$

Comparing eq.(9) with eq.(2), we can derive the least-squares solution for Wb as

$$W\hat{b} = \begin{bmatrix} -1 \\ Y^{+}y \end{bmatrix} = \hat{a} \tag{11}$$

and we get

$$\hat{b} = W^{-1}\,\hat{a} \tag{12}$$

Equation (12) represents the relationship between the predictors obtained by the covariance method on differenced data and those on original data.

2. Difference Operation in the Auto-Correlation Method

The normal equation of the auto-correlation method for original data sequence is expressed as

$$R\,\hat{\alpha} = r \qquad \text{or} \qquad [r \,\vdots\, R]\,\hat{a} = 0 \tag{13}$$

where

$$\mathbf{R} \begin{bmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & \cdots & r_{p-2} \\ \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & \cdots & r_0 \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} r_1 \\ \vdots \\ \vdots \\ r_p \end{bmatrix} \tag{14}$$

and

$$r_i = \sum_n y_n\, y_{n+i}\,. \tag{15}$$

The least-squares solution for α or a is obtained as

$$\hat{\alpha} = R^{-1}\,r \qquad \text{or} \qquad \hat{a} = \begin{bmatrix} -1 \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} -1 \\ R^{-1}r \end{bmatrix} \tag{16}$$

On the other hand, the normal equation for the differenced data is expressed as

$$P\,\hat{\beta} = \rho \qquad\qquad [\rho \,\vdots\, P]\,\hat{b} = 0 \tag{17}$$

where

$$P = \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & \rho_0 \end{bmatrix}, \quad \rho = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_p \end{bmatrix} \tag{18}$$

and

$$\rho_i = \sum_n (y_n - w\,y_{n-1})\,(y_{n+i} - w\,y_{n+i-1}). \tag{19}$$

Since ρ_i is rewritten as

$$\rho_i = -w\,r_{|i-1|} + (1+w^2)\,r_i - w\,r_{i+1}, \tag{20}$$

we can rewrite $[\rho \,\vdots\, P]$ as

$$[\rho \,\vdots\, P] = W^{\mathrm{T}}\,[r \,\vdots\, R]\,W \tag{21}$$

and eq. (17) as

$$W^{\mathrm{T}}\,[r \,\vdots\, R]\,W\,\hat{b} = 0. \tag{22}$$

As $|W^{\mathrm{T}}| \neq 0$, we get

$$[r \,\vdots\, R]\,W\,\hat{b} = 0. \tag{23}$$

The least-squares solution for $W\,b$ is obtained as

$$W\,\hat{b} = \begin{bmatrix} -1 \\ R^{-1}r \end{bmatrix} = \hat{a} \tag{24}$$

and we get

$$\hat{b} = W^{-1}\,\hat{a}\,. \tag{25}$$

Equation (25) represents the relationship between the predictors obtained by the auto-correlation method on differenced data and those on original data.

3. Physical Interpretation of the Relation between $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$

The relations between $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ for the covariance method and that for the auto-correlation method are identical to each other as formulated in eqs. (12) and (25) or eqs. (11) and (24). The latter two equations express the following relation between the two sets of predictors:

$$\hat{\beta}_i - w\,\hat{\beta}_{i-1} = \hat{\alpha}_i, \quad i = 1,2,\ldots,p, \quad \hat{\beta}_0 = -1\,. \tag{26}$$

Modifying this equation, we get the following successive equation:

$$\hat{\beta}_i = \hat{\alpha}_i + w\,\hat{\beta}_{i-1} \tag{27}$$

In closed form, it is written as

$$\hat{\beta}_i = \sum_{j=1}^{i} w^{i-j}\,\hat{\alpha}_j - w^i \tag{28}$$

Equation (28) is another expression of eqs. (12) or (25), because

$$W^{-1} = \begin{bmatrix} 1 & & & & 0 \\ w & 1 & & & \\ w^2 & w & 1 & & \\ \vdots & \vdots & & \ddots & \\ w^p & w^{p-1} & \cdots & w & 1 \end{bmatrix}. \tag{29}$$

Equation (26) proclaims that $\{\hat{\alpha}_i\}$ is the differenced sequence of $\{\hat{\beta}_i\}$. Both the difference operations on sampled data and that on the predictor sequence are interpreted to have the same effects of spectral enhancement in higher frequency region as depicted in Fig. 1, where w is assumed to be unity for simplicity.

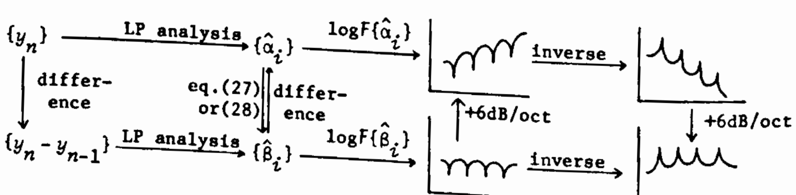

Figure 1. The relation between the difference operation on data and that on predictors.

4. Conclusions

The relationship between the predictors obtained on differenced data and those on original data has been derived. Although the way of derivation employed here is rather rough, the authors have already shown two ways of strict derivations: one (Yanagida et al., 1982a), equating the prediction errors at each sampling point for both sequences, and the other (Yanagida et al., 1982b), employing several theorems concerning the generalized inverse of matrices. Those, however, were only for the covariance method. This paper has discussed the difference operations in a linear prediction analysis aiming at a unified description for both the covariance method and the auto-correlation method.

In this paper, the discussion has been limited only to the first-order differencing, but the derived results are easily expanded to general higher order difference operations (Yanagida et al., 1982b).

Our present interest is to develop an efficient method to replace the sample differencing of fixed pre-emphasis factor with an adaptive inverse differencing on predictors, that is to replace the fixed pre-processing with an adaptive post-processing.

Acknowledgement

This work has been supported by Grant-in-Aid for Scientific Research for 1982 by the Ministry of Education, Science and Culture (No. 56550038).

References

Yanagida, M. et al. (1982a). Effects of the difference operation on linear predictive coefficients. *Proc. Autumn Meet. ASJ.* 3-6-6.

Yanagida, M. et al. (1982b). Analysis of the differencing operation in the linear prediction model by the generalized inverse matrix. *Trans. of the Committee on Speech Research, Acoust. Soc. Jpn.* S82-51.