ON AN UNBIASED LINEAR PREDICTOR OF A MULTIDIMENTIONAL NONSTATIONARY PROCESS

ON AN UNBIASED LINEAR PREDICTOR OF A MULTIDIMENTIONAL NONSTATIONARY PROCESS

By

Hajime YAMATO

(Received August, 1968)

§ 1. Summary

The prediction problem of a univariate nonstationary process represented by the sum of a polynomial mean value function and a stationary process has been discussed by G. E. P. Box and G. M. Jenkins [1]. It is our aim to consider the same problem in a multivariate case. The result may be regarded as a multivariate analogue of the result of E. Parzen [2] in the univariate case.

§ 2. Assumptions and notations

We deal with a n-dimentional nonstationary process $\boldsymbol{X}(t)$, $t=\cdots, -1, 0, 1, \cdots$. We shall consider a problem of predicting the value of $\boldsymbol{X}(t+\tau)$ on the basis of p previous m-dimentional observation vectors $\boldsymbol{Z}(t)$, $\boldsymbol{Z}(t-1)$, $\cdots$, $\boldsymbol{Z}(t-p+1)$. We shall restrict ourselves to the linear predictor of the form

$$\sum_{s=0}^{p-1} \boldsymbol{\mu}(s)\boldsymbol{Z}(t-s)$$

with the condition of unbiasedness

$$E[\boldsymbol{X}(t+\tau)]=E[\sum_{s=0}^{p-1} \boldsymbol{\mu}(s)\boldsymbol{Z}(t-s)],$$

where $\boldsymbol{\mu}(s)$ is a $n\times m$ matrix, $p>n$ and $m\geqq n$. We need the following assumptions and notations.

Assumptions

A 1. The process $\boldsymbol{X}(t)$ is represented as the sum of two functions:

$$\boldsymbol{X}(t)=\boldsymbol{M}(t)+\boldsymbol{\xi}(t).$$

We call a n-dimentional vector $\boldsymbol{M}(t)$ the mean value vector, and the i-th component $M_i(t)$ of $\boldsymbol{M}(t)$ can be represented by a polynomial of t of known degree $0\leqq d_i<p-1$ for $i=1, 2, \cdots, n$. $\boldsymbol{\xi}(t)$ is a n-dimentional stationary process with zero mean vector and known $n\times n$ stationary covariance matrices $\boldsymbol{R}(t)$, $t=\cdots -1, 0, 1, \cdots$, where $\boldsymbol{R}(t-s)$

$=E[\xi(t)\xi'(s)]$ for any t and s.

A 2. The observation vector $\boldsymbol{Z}(t)$ is represented as the sum of the signal $\boldsymbol{X}(t)$ multiplied by a known $m\times n$ matrix $\boldsymbol{H}$ and the noise $\boldsymbol{U}(t)$:

$$\boldsymbol{Z}(t)=\mathbf{H}\boldsymbol{X}(t)+\boldsymbol{U}(t).$$

The rank of $\mathbf{H}$ is n and $\boldsymbol{U}(t)$ is a m-dimentional stationary process with zero mean vector and known $m\times m$ stationary covariance matrices $\mathrm{S}(t)$, $t=\cdots, -1, 0, 1, \cdots$, where $\mathrm{S}(t-s)=E[\boldsymbol{U}(t)\boldsymbol{U}'(s)]$.

A 3. Process $\xi(t)$ and $\boldsymbol{U}(t)$ are independent.

A 4. A $mp\times mp$ matrix $\mathbf{T}$ defined by

$$\mathbf{T}=\begin{pmatrix} \mathbf{T}(0), & \mathbf{T}(1), & \cdots, & \mathbf{T}(p-1) \\ \mathbf{T}(-1), & \mathbf{T}(0), & \cdots, & \mathbf{T}(p-2) \\ & \cdots & & \\ \mathbf{T}(-p+1), & \mathbf{T}(-p+2), & \cdots, & \mathbf{T}(0) \end{pmatrix}$$

is positive definite, where $\mathbf{T}(\cdot)$ is the covariance matrix of the process $\boldsymbol{Z}(t)$, i. e. $\mathbf{T}(t-s)=\mathrm{Cov}(\boldsymbol{Z}(t), \boldsymbol{Z}'(s))$.

Notations

N 1. A m-dimentional vector $\boldsymbol{\mu}_i'(\cdot)$ is the i-th row of the matrix $\boldsymbol{\mu}(\cdot)$ and

$$\boldsymbol{\mu}_i[mp\times 1]=\begin{pmatrix} \boldsymbol{\mu}_i(0) \\ \boldsymbol{\mu}_i(1) \\ \vdots \\ \boldsymbol{\mu}_i(p-1) \end{pmatrix}$$

for all $i=1, 2, \cdots, n$.

N 2. A n-dimentional vector $\mathbf{r}_i'(\cdot)$ is the i-th row of the matrix $\mathbf{R}(\cdot)$,

$$\boldsymbol{R}_i[np\times 1]=\begin{pmatrix} \mathbf{r}_i(\tau) \\ \mathbf{r}_i(\tau+1) \\ \vdots \\ \mathbf{r}_i(\tau+p-1) \end{pmatrix} \text{ and } \mathbf{R}[np\times n]=\begin{pmatrix} \mathbf{R}'(\tau) \\ \mathbf{R}'(\tau+1) \\ \vdots \\ \mathbf{R}'(\tau+p-1) \end{pmatrix}$$

for all $i=1, 2, \cdots, n$.

N 3.

$$\mathbf{B}\Big[mp\times \sum_{i=1}^{n}(d_i+1)\Big]=\begin{pmatrix} h_1(1, & 0, & \cdots, & 0\ ), & h_2(1, & 0, & \cdots, & 0\ ) & \cdot & \cdot & \cdot & , h_n(1, & 0, & \cdots, & 0\ ) \\ h_1(1, & 1, & \cdots, & 1^{d_1}), & h_2(1, & 1, & \cdots, & 1^{d_2}) & \cdot & \cdot & \cdot & , h_n(1, & 1, & \cdots, & 1^{d_n}) \\ & \cdot & \cdot & \cdot & & & & & & & & \cdot & \cdot & \cdot \\ h_1(1, & p-1, & \cdots, & (p-1)^{d_1}), & h_2(1, & p-1, & \cdots, & (p-1)^{d_2}) & \cdot & \cdot & \cdot & , h_n(1, & p-1, & \cdots, & (p-1)^{d_n}) \end{pmatrix},$$

where a m-dimentional vector h_i is the i-th columm of the matrix $\mathbf{H}$ for all $i=1, 2, \cdots, n$.

N 4.

$$\boldsymbol{c}(s, i)\Big[\sum_{j=1}^{n}(d_j+1)\times 1\Big]=\begin{pmatrix} \mathbf{0}_{d_1+1} \\ \vdots \\ \mathbf{0}_{d_{i-1}+1} \\ 1 \\ s \\ s^2 \\ \vdots \\ s^{d_i} \\ \mathbf{0}_{d_{i+1}+1} \\ \vdots \\ \mathbf{0}_{d_n+1} \end{pmatrix}$$

where $\mathbf{0}_{d_j+1}$ is a (d_j+1)-dimentional zero vector for all $j=1, 2, \cdots, n$.

N 5. $\mathbf{I}_p$ is a $p\times p$ unite matrix.

N 6.

$$\mathbf{C}\Big[\sum_{i=1}^{n}(d_i+1)\times n\Big]=(\boldsymbol{c}(-\tau, 1), \boldsymbol{c}(-\tau, 2), \cdots, \boldsymbol{c}(-\tau, n))$$

N 7.

$$\mathbf{Z}[mp\times 1]=\begin{pmatrix} \boldsymbol{Z}(t) \\ \boldsymbol{Z}(t-1) \\ \vdots \\ \boldsymbol{Z}(t-p+1) \end{pmatrix}$$

§ 3. The best linear unbiased predictor

Lemma 1. *The covariance matrix* $\mathbf{T}(\cdot)$ *of the process* $\boldsymbol{Z}(t)$ *is given by*

$$\mathbf{T}(t)=\mathbf{H}\mathbf{R}(t)\mathbf{H}'+\mathbf{S}(t) \qquad \textit{for all} \quad t.$$

Also the following relations hold:

$$\mathrm{Cov}(\boldsymbol{X}(t), \boldsymbol{Z}'(s))=\mathbf{R}(t-s)\mathbf{H}',$$

$$\mathrm{Cov}(\boldsymbol{Z}(t), \boldsymbol{X}'(s))=\mathbf{H}\mathbf{R}(t-s),$$

$$\mathbf{R}(t-s)=\mathbf{R}'(s-t),$$

$$\mathbf{S}(t-s)=\mathbf{S}'(s-t)$$

and

$$\mathbf{T}(t-s)=\mathbf{T}'(s-t).$$

Lemma 2. *The covariance matrix of the prediction error*

$$\boldsymbol{\varepsilon}(t,\tau)=\boldsymbol{X}(t+\tau)-\sum_{s=0}^{p-1}\boldsymbol{\mu}(s)\boldsymbol{Z}(t-s)$$

is given by

$$\mathbf{R}(0)-\sum_{j=0}^{p-1}\mathbf{R}(\tau+j)\mathbf{H}'\boldsymbol{\mu}'(j)-\sum_{k=0}^{p-1}\boldsymbol{\mu}(k)\mathbf{H}\mathbf{R}'(\tau+k)+\sum_{k=0}^{p-1}\sum_{j=0}^{p-1}\boldsymbol{\mu}(k)\mathbf{T}(j-k)\boldsymbol{\mu}'(j).$$

An unbiased linear predictor is said to be the best when the trace of the covariance matrix of the prediction error is minimized. In order to derive the best one, we evaluate the trace of the error matrix.

Lemma 3. *Under the unbiasedness* $E[\boldsymbol{X}(t+\tau)]=E[\sum_{s=0}^{p-1}\boldsymbol{\mu}(s)\boldsymbol{Z}(t-s)]$ *we have*

$$\begin{aligned}&\operatorname{tr}\operatorname{Cov}(\boldsymbol{\varepsilon}(t,\tau),\boldsymbol{\varepsilon}'(t,\tau))\\&\quad=\operatorname{tr}\mathbf{R}(0)-2\sum_{i=1}^{n}\boldsymbol{\mu}_i'(\mathbf{I}_p\otimes\mathbf{H})\boldsymbol{R}_i+\sum_{i=1}^{n}\boldsymbol{\mu}_i'\mathbf{T}\boldsymbol{\mu}_i,\end{aligned}$$

where $\otimes$ *denotes the Kronecker product.*

Lemma 4. *A necessary and sufficient condition for the linear predictor* $\sum_{s=0}^{n}\boldsymbol{\mu}(s)\boldsymbol{Z}(t-s)$ *to be unbiased is*

$$\mathbf{B}'\boldsymbol{\mu}_i=\boldsymbol{c}(-\tau,i)\quad\text{for all}\quad\boldsymbol{\mu}_i,\ i=1,2,\cdots,n.$$

It is easy to show that rank $(\mathbf{B})=\sum_{i=1}^{n}(d_i+1)$ and $\mathbf{B}'\mathbf{T}^{-1}\mathbf{B}$ is positive definite, so we can obtain the best unbiased linear predictor by the Lagrange's multiplier.

Theorem 1. *The best unbiased linear predictor* $\hat{\mathbf{X}}(t+\tau)$ *which minimizes the trace of the covariance matrix of the prediction error subject to the unbiasedness is given by*

$$(1)\quad \hat{\mathbf{X}}(t+\tau)=\mathbf{R}'(\mathbf{I}_p\otimes\mathbf{H})'\mathbf{T}^{-1}\mathbf{Z}+(\mathbf{C}'-\mathbf{R}'(\mathbf{I}_p\otimes\mathbf{H})'\mathbf{T}^{-1}\mathbf{B})(\mathbf{B}'\mathbf{T}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{T}^{-1}\mathbf{Z}$$

with a covariance matrix

$$\begin{aligned}(2)\quad \operatorname{Cov}(\hat{\mathbf{X}}(t+\tau),\hat{\mathbf{X}}'(t+\tau))&=\mathbf{C}'(\mathbf{B}'\mathbf{T}^{-1}\mathbf{B})^{-1}\mathbf{C}\\&\quad+\mathbf{R}'(\mathbf{I}_p\otimes\mathbf{H})'(\mathbf{T}^{-1}-\mathbf{T}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{T}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{T}^{-1})(\mathbf{I}_p\otimes\mathbf{H})\mathbf{R}.\end{aligned}$$

We now consider the estimate of the mean value function.

Theorem 2. *Let us put*

$$M_i(t-s)=\sum_{j=1}^{d_i+1}a_{ij}s^{j-1}$$

where t *is the present time. Then the minimum variance unbiased linear estimate* $\hat{M}_i(t-s)$ *of* $M_i(t-s)$ *is given by*

$$\hat{M}_i(t-s)=\boldsymbol{c}'(s,\ i)(\mathbf{B}'\mathbf{T}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{T}^{-1}\mathbf{Z} \tag{3}$$

and

$$\hat{\boldsymbol{M}}(t-s)=\begin{pmatrix}\hat{M}_1(t-s)\\ \hat{M}_2(t-s)\\ \vdots\\ \hat{M}_n(t-s)\end{pmatrix}=\begin{pmatrix}\boldsymbol{c}'(s,\ 1)\\ \boldsymbol{c}'(s,\ 2)\\ \vdots\\ \boldsymbol{c}'(s,\ n)\end{pmatrix}(\mathbf{B}'\mathbf{T}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{T}^{-1}\mathbf{Z}. \tag{4}$$

On the other hand, in case when the mean value functions were known, we have:

Theorem 3. *In the case where all mean vectors* $\boldsymbol{M}(t)$, $t=\cdots,\ -1,\ 0,\ 1,\ \cdots$ *are known, the best unbiased linear predictor* $\tilde{\boldsymbol{X}}(t+\tau)$ *is given by*

$$\tilde{\boldsymbol{X}}(t+\tau)=\boldsymbol{M}(t+\tau)+\mathbf{R}'(\mathbf{I}_p\otimes\mathbf{H})'\mathbf{T}^{-1}(\mathbf{Z}-\mathbf{HM}) \tag{5}$$

with the covariance matrix

$$\mathrm{Cov}(\tilde{\boldsymbol{X}}(t+\tau),\ \tilde{\boldsymbol{X}}'(t+\tau))=\mathbf{R}'(\mathbf{I}_p\otimes\mathbf{H})'\mathbf{T}^{-1}(\mathbf{I}_p\otimes\mathbf{H})\mathbf{R}, \tag{6}$$

where

$$\mathbf{HM}=\begin{pmatrix}\mathbf{H}\boldsymbol{M}(t)\\ \mathbf{H}\boldsymbol{M}(t-1)\\ \vdots\\ \mathbf{H}\boldsymbol{M}(t-p+1)\end{pmatrix}.$$

If we substitute the best linear unbiased estimate of the mean value function (4) in place of the mean value function in (5), it turns out to be equal to (1). Thus we have:

Theorem 4. *This result may be regarded as a multivariate analogue of the minimum variance linear unbiased predictor derived by E. Parzen* [2].

ACKNOWLEDGEMENT

I would like to express my thanks to Dr. S. Kano and Dr. A. Kudo for their very helpful comments.

References

[1] G. E. P. Box and G. M. Jenkins (1962), Some statistical aspects of adaptive optimization and control, J. R. Statist. Soc. B, 24, 297-343.

[2] E. Parzen (1963), A new approach to the synthesis of optimal smoothing and prediction systems, Mathematical Optimization Techniques, University of California Press, Berkeley, Calif., 75-108.

[3] E. Parzen (1961), An approach to time series analysis, Ann. Math. Statist., Vol. **32**, No. 4, 951-989.