ON THE VARIANCE OF A SIMPLIFIED ESTIMATE OF CORRELOGRAM

MITUAKI HUZII

(Received Jan. 13, 1970; Revised May 12, 1970)

1. Introduction

Let $X(n)$, $n=0, \pm 1, \pm 2, \cdots$, be a real valued weakly stationary process. We assume $\mathrm{E}X(n)=0$ for simplicity. Now we put

$$\mathrm{E}(X(n)^2)=\sigma^2 \qquad \text{and} \qquad \mathrm{E}(X(n)X(n+h))=\sigma^2\rho_h \ .$$

In the following, we assume the variance σ^2 to be known and discuss the estimation of ρ_h. Let $X(n)$ be observed at $n=1, 2, \cdots, N, \cdots, N+h$. Then we usually use the statistic

$$\tilde{\gamma}_h=\frac{1}{\sigma^2}\frac{1}{N}\sum_{n=1}^{N} X(n)X(n+h)$$

to estimate ρ_h. Obviously $\tilde{\gamma}_h$ is an unbiased estimate of ρ_h.

On the other hand, if $X(n)$ is a Gaussian process, the statistic

$$\gamma_h=\sqrt{\frac{\pi}{2}}\frac{1}{\sigma}\frac{1}{N}\sum_{n=1}^{N} X(n)\,\mathrm{sgn}\,(X(n+h))$$

is an unbiased estimate of ρ_h (see [3], [4]), where $\mathrm{sgn}\,(y)$ means $1, 0, -1$ correspondingly as $y>0$, $y=0$, $y<0$. We have shown (see [3], [4]) that the variance of γ_h is smaller than that of $\tilde{\gamma}_h$ for a small $|h|$, if

$$\rho_\tau=ae^{-b|\tau|}\cos(c\,|\tau|+d) \ ,$$

where a, b, c, d are constants and $b>0$.

In this paper, we introduce an estimate which is a generalization of γ_h. Let

$$G_\delta(x)=\begin{cases} 0 & \text{for } x>\delta\sigma \ , \\ 1 & \text{for } 0<x\leqq\delta\sigma \ , \\ 0 & \text{for } x=0 \ , \\ -1 & \text{for } -\delta\sigma\leqq x<0 \ , \\ 0 & \text{for } x<-\delta\sigma \ , \end{cases}$$

where, δ is a positive constant. If $\delta=\infty$, we have $G_\infty(x)=\mathrm{sgn}\,(x)$. Our new estimate is obtained by replacing $\mathrm{sgn}\,(X(n+h))$ in the statistic γ_h by $G_\delta(X(n+h))$. We discuss the relation between the value of δ and the variance of this new estimate, and we numerically investigate the estimate with minimum variance.

This result shows that the optimum value $\delta(h)$ of δ, which gives the minimum variance estimate, takes a finite value for small h and that it increases indefinitely as h increases.

2. The unbiased estimate using $G_\delta(X(n+h))$ and its variance

Let $X(n)$ be a stationary Gaussian process. Then we have

$$\mathrm{E}(X(n)G_\delta(X(n+h)))=\sqrt{\frac{2}{\pi}}\sigma\rho_h(1-e^{-\delta^2/2}) .$$

So, the statistic

$$\gamma_h(\delta)=\frac{1}{\alpha}\frac{1}{N}\sum_{n=1}^{N} X(n)G_\delta(X(n+h))$$

is an unbiased estimate of ρ_h, where

$$\alpha=\sqrt{\frac{2}{\pi}}\sigma(1-e^{-\delta^2/2}) .$$

Now, we evaluate the variance of $\gamma_h(\delta)$.

$$(1)\quad \mathrm{Var}\,(\gamma_h(\delta))=\frac{1}{\alpha^2}\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathrm{E}(X(n)G_\delta(X(n+h))X(m)G_\delta(X(m+h)))-\rho_h^2)$$
$$=\frac{1}{\alpha^2}\frac{1}{N}\Big\{\mathrm{E}(X(0)^2G_\delta(X(h))^2)$$
$$+\frac{2}{N}\sum_{k=1}^{N-1}(N-k)\mathrm{E}(X(0)G_\delta(X(h))X(k)G_\delta(X(h+k)))\Big\}-\rho_h^2 .$$

In the expression of $\mathrm{Var}\,(\gamma_h(\delta))$, we have

$$\mathrm{E}(X(0)^2G_\delta(X(h))^2)=\mathrm{E}\,[\rho_hX(h)+(X(0)-\rho_hX(h))]^2G_\delta(X(h))^2$$
$$=\rho_h^2\mathrm{E}(X(h)^2G_\delta(X(h))^2)+\mathrm{E}[(X(0)-\rho_hX(h))^2\mathrm{E}(G_\delta(X(h))^2)]$$
$$=\sqrt{\frac{2}{\pi}}\sigma^2\rho_h^2\left\{\int_0^\delta e^{-t^2/2}dt-\delta e^{-\delta^2/2}\right\}$$
$$+\sqrt{\frac{2}{\pi}}\sigma^2(1-\rho_h^2)\int_0^\delta e^{-t^2/2}dt .$$

In the following of this section and in Section 3, we consider the case when $X(n)$ is a simple Markov process. Then we have

$$\rho_\tau=a^{|\tau|}$$

where a is a constant and $0<a<1$.

Let us first evaluate the value of

$$E(X(0)X(k)G_\delta(X(h))G_\delta(X(k+h)))\ .$$

(i) When $1\leqq k\leqq h-1$, we have

$$\begin{aligned}&E(X(0)X(k)G_\delta(X(h))G_\delta(X(k+h)))\\&=\rho_k E(X(k)^2G_\delta(X(h))G_\delta(X(k+h)))\\&=\rho_k\rho_{h-k}^2 E(X(h)^2G_\delta(X(h))G_\delta(X(k+h)))\\&\quad+\sigma^2\rho_k(1-\rho_{h-k}^2)E(G_\delta(X(h))G_\delta(X(k+h)))\\&=\rho_k\rho_{h-k}^2 E(X(0)^2G_\delta(X(0))G_\delta(X(k)))\\&\quad+\sigma^2\rho_k(1-\rho_{h-k}^2)E(G_\delta(X(0))G_\delta(X(k)))\ .\end{aligned}$$

(ii) When $k=h$, it holds

$$\begin{aligned}&E(X(0)X(h)G_\delta(X(h))G_\delta(X(2h)))\\&=\rho_h E(X(h)^2G_\delta(X(h))G_\delta(X(2h)))\\&=\rho_h E(X(0)^2G_\delta(X(0))G_\delta(X(h)))\ .\end{aligned}$$

(iii) When $h+1\leqq k\leqq N-1$, we have

$$\begin{aligned}&E(X(0)G_\delta(X(h))X(k)G_\delta(X(k+h)))\\&=\rho_h E(X(h)G_\delta(X(h))X(k)G_\delta(X(k+h)))\\&=\frac{(\rho_k-\rho_k\rho_h^2)}{1-\rho_k^2}E(X(h)^2G_\delta(X(h))G_\delta(X(k+h)))\\&\quad+\frac{(\rho_h^2-\rho_k^2)}{1-\rho_k^2}E(X(h)G_\delta(X(h))X(k+h)G_\delta(X(k+h)))\\&=\frac{\rho_k(1-\rho_h^2)}{1-\rho_k^2}E(X(0)^2G_\delta(X(0))G_\delta(X(k)))\\&\quad+\frac{(\rho_h^2-\rho_k^2)}{1-\rho_k^2}E(X(0)G_\delta(X(0))X(k)G_\delta(X(k)))\ .\end{aligned}$$

Consequently, we get

$$\begin{aligned}\mathrm{Var}\,(\gamma_h(\delta))=\frac{1}{a^2}\frac{1}{N}\Big\{&E(X(0)^2G_\delta(X(h))^2)+\frac{2}{N}\sum_{k=1}^{h-1}(N-k)V_1(k)\\&+\frac{2}{N}(N-h)V_2(h)+\frac{2}{N}\sum_{k=h+1}^{N-1}(N-k)V_3(k)\Big\}-\rho_h^2\ ,\end{aligned}$$

where

$$\begin{aligned}V_1(k)&=\rho_k\rho_{h-k}^2E(X(0)^2G_\delta(X(0))G_\delta(X(k)))\\&\quad+\sigma^2\rho_k(1-\rho_{h-k}^2)E(G_\delta(X(0))G_\delta(X(k)))\ ,\\V_2(h)&=\rho_hE(X(0)^2G_\delta(X(0))G_\delta(X(h)))\ ,\end{aligned}$$

$$V_3(k)=\frac{\rho_k(1-\rho_h^2)}{1-\rho_k^2}\mathrm{E}(X(0)^2G_\delta(X(0))G_\delta(X(k)))$$
$$+\frac{(\rho_h^2-\rho_k^2)}{1-\rho_k^2}\mathrm{E}(X(0)G_\delta(X(0))X(k)G_\delta(X(k)))\ .$$

3. Numerical results

First, we shall derive the formulas which will conveniently be used for the evaluation of each term of Var $(\gamma_h(\delta))$. In general, we have

$$\mathrm{E}(X(0)^iX(k)^jG_\delta(X(0))G_\delta(X(k)))$$
$$=\frac{1}{2\pi\sigma^2\sqrt{1-\rho_k^2}}\Big[\{1+(-1)^{i+j}\}$$
$$\cdot\int_0^{\delta\sigma}\int_0^{\delta\sigma}x^iy^j\exp\left\{-\frac{1}{2\sigma^2(1-\rho_k^2)}(x^2-2\rho_kxy+y^2)\right\}dxdy$$
$$-\{(-1)^i+(-1)^j\}$$
$$\cdot\int_0^{\delta\sigma}\int_0^{\delta\sigma}x^iy^j\exp\left\{-\frac{1}{2\sigma^2(1-\rho_k^2)}(x^2+2\rho_kxy+y^2)\right\}dxdy\Big]\ ,$$

and

$$I_{i,j}(A,\rho_k)=\int_0^A\int_0^A x^iy^je^{-(x^2-2\rho_kxy+y^2)/2}dxdy$$
$$=\sum_{l=0}^{\infty}\frac{(\rho_k)^l}{l!}\left(\int_0^A x^{i+l}e^{-x^2/2}dx\right)\left(\int_0^A y^{j+l}e^{-y^2/2}dy\right)$$
$$=\sum_{l=0}^{\infty}\frac{(\rho_k)^l}{l!}U_{i+l}(A)U_{j+l}(A)\ ,$$

where

$$U_p(A)=\int_0^A x^pe^{-x^2/2}dx\ .$$

On the other hand, we have

$$U_0(A)=\int_0^A e^{-x^2/2}dx\ ,$$
$$U_1(A)=1-e^{-A^2/2}\ ,$$

and

$$U_p(A)=-A^{p-1}e^{-A^2/2}+(p-1)U_{p-2}(A)\qquad\text{for } p\geqq 2\ .$$

Using the above relations, we get

$$\mathrm{E}(X(0)^2G_\delta(X(0))G_\delta(X(k)))$$
$$=\frac{\sigma^2(1-\rho_k^2)^{3/2}}{\pi}\left[I_{2,0}\left(\frac{\delta}{\sqrt{1-\rho_k^2}},\ \rho_k\right)-I_{2,0}\left(\frac{\delta}{\sqrt{1-\rho_k^2}},\ -\rho_k\right)\right],$$

$$\mathrm{E}(X(0)X(k)G_\delta(X(0))G_\delta(X(k)))$$
$$=\frac{\sigma^2(1-\rho_k^2)^{3/2}}{\pi}\left[I_{1,1}\left(\frac{\delta}{\sqrt{1-\rho_k^2}},\ \rho_k\right)+I_{1,1}\left(\frac{\delta}{\sqrt{1-\rho_k^2}},\ -\rho_k\right)\right],$$

and

$$\mathrm{E}(G_\delta(X(0))G_\delta(X(k)))$$
$$=\frac{(1-\rho_k^2)^{1/2}}{\pi}\left[I_{0,0}\left(\frac{\delta}{\sqrt{1-\rho_k^2}},\ \rho_k\right)-I_{0,0}\left(\frac{\delta}{\sqrt{1-\rho_k^2}},\ -\rho_k\right)\right].$$

Now we shall numerically discuss the relation between δ and $\mathrm{Var}\,(\gamma_h(\delta))$. For this, we treat the case where

$$\rho_\tau=(0.8)^{|\tau|}.$$

The results are shown in Tables 1 to 5.

For a fixed h, we denote by $\delta(h)$ the value of δ which minimizes $\mathrm{Var}\,(\gamma_h(\delta))$. The results show that $\delta(h)$ increases as h increases.

Table 1

$h=1$ $(\rho_1=0.8)$			
δ	$\mathrm{Var}\,(\gamma_1(\delta))$	δ	$\mathrm{Var}\,(\gamma_1(\delta))$
0.50	0.16205	1.51	0.02029
1.00	0.02938	1.55	0.02035
1.30	0.02121	1.60	0.02049
1.40	0.02047	1.70	0.02100
1.45	0.02032	2.00	0.02361
1.47	0.02029	3.00	0.03229
1.48	0.02028	4.00	0.03427
1.49	0.02028	5.00	0.03438
1.50	0.02028	∞	0.03438

Table 2

$h=2$ $\left(\begin{smallmatrix}\rho_2=(0.8)^2\\=0.64\end{smallmatrix}\right)$	
δ	Var $(\gamma_2(\delta))$
1.0	0.05058
1.5	0.03228
1.6	0.03176
1.7	0.03168
1.8	0.03191
2.0	0.03298
3.0	0.03954
4.0	0.04113
5.0	0.04122
∞	0.04122

Table 3

$h=5$ $\left(\begin{smallmatrix}\rho_5=(0.8)^5\\=0.32768\end{smallmatrix}\right)$	
δ	Var $(\gamma_5(\delta))$
1.0	0.09958
1.5	0.05953
2.0	0.05071
2.4	0.04943
2.5	0.04940
2.6	0.04942
3.0	0.04973
4.0	0.05017
5.0	0.05020
∞	0.05020

Table 4

$h=10$ $\left(\begin{smallmatrix}\rho_{10}=(0.8)^{10}\\=0.10737\end{smallmatrix}\right)$	
δ	Var $(\gamma_{10}(\delta))$
1.0	0.12599
2.0	0.06012
3.0	0.05351
4.0	0.05306
5.0	0.05305
∞	0.05305

Table 5

$h=20$ $\left(\begin{smallmatrix}\rho_{20}=(0.8)^{20}\\=0.01153\end{smallmatrix}\right)$	
δ	Var $(\gamma_{20}(\delta))$
1.0	0.13090
2.0	0.06194
3.0	0.05407
4.0	0.05341
5.0	0.05339
∞	0.05339

4. Var $(\gamma_h(\delta))$ for a general Gaussian process

In Section 2 and Section 3, we have evaluated Var $(\gamma_h(\delta))$, mainly, for a simple Markov Gaussian process. In this section, we shall do it for a general Gaussian process. We can do this evaluation analogously as in a previous paper (see [4]), and we shall only show the outline of this evaluation.

We assume that the process $X(n)$ has the correlogram satisfying the following two conditions (A.1) and (A.2):

(A.1) It holds

$$\Delta=\begin{vmatrix} 1 & \rho_{k-h} & \rho_h \\ \rho_{k-h} & 1 & \rho_k \\ \rho_h & \rho_k & 1 \end{vmatrix} \neq 0\,, \qquad D=\begin{vmatrix} 1 & \rho_k \\ \rho_k & 1 \end{vmatrix} \neq 0\,, \qquad D_1=\begin{vmatrix} 1 & \rho_{h-k} \\ \rho_{h-k} & 1 \end{vmatrix} \neq 0\,,$$

$$D_2=\begin{vmatrix} 1 & \rho_{h+k} \\ \rho_{h+k} & 1 \end{vmatrix} \neq 0 \qquad \text{for } k\geqq 1 \text{ and } k\neq h\,.$$

(A.2)

$$D_h=\begin{vmatrix} 1 & \rho_h \\ \rho_h & 1 \end{vmatrix} \neq 0 \qquad \text{and} \qquad D_{2h}=\begin{vmatrix} 1 & \rho_{2h} \\ \rho_{2h} & 1 \end{vmatrix} \neq 0\,.$$

Let us investigate the value of $\mathrm{E}(X(0)G_\delta(X(h))X(k)G_\delta(X(k+h)))$. When $k\neq h$, we put

$$X(0)=AX(k)+BX(h)+CX(k+h)+\varepsilon(0)\,,$$

where A, B and C are constants such that the random variable $\varepsilon(0)$ is uncorrelated with $X(k)$, $X(h)$ and $X(k+h)$. We further put

$$X(k)=FX(h)+GX(k+h)+\eta(k)\,,$$

where F and G are constants such that $\eta(k)$ is uncorrelated with $X(h)$ and $X(k+h)$.

By using these representation, we have

$$\begin{aligned} &\mathrm{E}(X(0)G_\delta(X(h))X(k)G_\delta(X(k+h))) \\ &=(AF^2+BF)\mathrm{E}(X(0)^2G_\delta(X(0))G_\delta(X(k))) \\ &\quad+(2AFG+BG+CF)\mathrm{E}(X(0)X(k)G_\delta(X(0))G_\delta(X(k))) \\ &\quad+(AG^2+CG)\mathrm{E}(X(k)^2G_\delta(X(0))G_\delta(X(k))) \\ &\quad+A\frac{\sigma^2\Delta}{D}\mathrm{E}(G_\delta(X(0))G_\delta(X(k)))\,, \end{aligned}$$

where Δ and D are given in (A.1).

When $k=h$, we obtain, in the same way,

$$\begin{aligned} &\mathrm{E}(X(0)X(h)G_\delta(X(h))G_\delta(X(2h))) \\ &=H\mathrm{E}(X(0)^2G_\delta(X(0))G_\delta(X(h)))+K\mathrm{E}(X(0)X(h)G_\delta(X(0))G_\delta(X(h)))\,, \end{aligned}$$

where H and K are constants such that $X(0)-HX(h)-KX(2h)$ is uncorrelated with $X(h)$ and $X(2h)$.

Using the above results and the method discussed in Section 3, we can numerically evaluate the variance of the estimate $\gamma_h(\delta)$. If the correlogram is given by

$$\rho_\tau=ae^{-b|\tau|}\cos(c\,|\,\tau\,|+d)\,,$$

we will obtain similar results as in Section 3 (see [3], [4]).

Acknowledgement

The author wishes to express his gratitude to Dr. H. Akaike for helpful comments and advices.

TOKYO INSTITUTE OF TECHNOLOGY

REFERENCES

[1] Brillinger, D. R. (1968). Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeros, *J. R. Statist. Soc.* **B**, **30**, 145-159.

[2] Doob, J. L. (1953). *Stochastic processes*, New York.

[3] Huzii, M. (1962). On a simplified method of the estimation of the correlogram for a stationary Gaussian process, *Ann. Inst. Stat. Math.*, **14**, 259-268.

[4] Huzii, M. (1964). On a simplified method of the estimation of the correlogram for a stationary Gaussian process, II, *Kōdai Math. Sem. Rep.*, **16**, 199-212.

[5] Rice, S. O. (1944-45). Mathematical analysis of random noise, *Bell. Syst. Tech. J.*, **23**, **24**.

[6] Rodemich, E. R. (1966). Spectral estimates using nonlinear functions, *Ann. Math. Statist.*, **37**, 1237-1256.