48. *Observed Value of the Autocorrelation Function*

By Hirosi ITÔ
Department of Applied Physics, Faculty of Engineering, Osaka University
(Comm. by M. MASIMA, M.J.A., May 13, 1953)

1. Introduction

The characteristics of the time series are given by its autocorrelation function or its Fourier transform, spectral density. Then, it is necessary to get its autocorrelation function by the observation for the study of time series. However, it is inevitable for some errors to be introduced in spite of our efforts to avoid them. In the measurement of the autocorrelation function, some sorts of fluctuations are considered, one of which is that of delay time. Other fluctuations, for example, that of amplitude seems to be based upon the mechanism of the instrument, but the author is not qualified to judge it, because he cannot see this instrument in his neighbourhood. The fluctuation of the delay time is very interesting from the theoretical standpoint and we can understand it from the uncertainty relation between time and frequency. The observed value of the spectrum has the definite relation with the mathematical spectrum and is correctable by the idealized experiment. This fact tells us also that ideal white noise cannot be observed in any case. The prove has been made on the line of ergodic theorem which is rewritten into the form which is easily handled.

It is often discussed how to choose the time interval of the measurement. This problem seems to be unimportant from the intuitive consideration, but was added for the completeness of this theory.

2. Ergodic Theorem

The definition of the autocorrelation function is given as:

$$\varphi(\tau)=\lim_{T=\infty}\frac{1}{2T}\int_{-T}^{T}f(t+\tau)f(t)\,dt. \tag{2.1}$$

The discrete form of (2.1) is

$$\varphi(\tau)=\lim_{N\to\infty}\frac{1}{N+1}\sum_{i=0}^{N}f(t_i+\tau)f(t_i) \tag{2.2}$$

where τ is the delay time.

When we choose the delay times successively, they fluctuate about the mean value τ_0, so that we have the observed value as follows:

$$\varphi^*(\tau_0)=\frac{1}{N+1}\sum_{i=0}^{N}f(t_i+\tau_0+s_i)f(t_i), \tag{2.3}$$

where s_i is the fluctuation of the i th time point. What is the relation between the mathematical autocorrelation function (2.2) and the observed value (2.3)? Or, how much information loss is there when we use the result (2.3)?

The difficulty in solving this problem is that, the law of the randomness of the time series is different from that to which the error distribution is obeyed. Nevertheless, the single summation with respect to the index i requires the both distribution laws when we use the formula (2.3). To avoid this confusion we use the ergodic property of the time series.

Let $f(t)$ be an almost periodic function which is real or complex, and bounded under the condition

$$|f(t)| < C. \tag{2.4}$$

We can prove the following theorem:

We can find for all $k(k=0, 1, \dots N)$ a positive integer M and an almost periodicity T such that

$$\left| \frac{1}{M+1} \sum_{j=0}^{M} \{f(t_k+\tau_0+s_{kj}+jT)f(t_k+jT)-f(t_k+\tau_0+s_{kj})f(t_k)\} \right| < \varepsilon,$$

where ε is the given positive number.

To prove this theorem, we use the ergodic property of $f(t)$. We can choose T such that

$$|f(t_k+T)-f(t_k)| < \frac{\varepsilon}{CM}$$

for all k. Now the following relation is easily proved.

$$\begin{aligned} &|f(t_k+2T)-f(t_k)| \\ &= |f(t_k+2T)-f(t_k+T)+f(t_k+T)-f(t_k)| \\ &< \big| |f(t_k+2T)-f(t_k+T)| + |f(t_k+T)-f(t_k)| \big| \\ &< \frac{2}{CM}\varepsilon. \end{aligned}$$

Generally we have

$$|f(t_k+jT)-f(t_k)| < \frac{j}{CM}\varepsilon. \quad (j=1, 2, \dots M)$$

Then

$$\begin{aligned} &|f(t_k+\tau_0+s_{kj}+jT)f(t_k+jT)-f(t_k+\tau_0+s_{kj})f(t_k)| \\ &= |\{f(t_k+\tau_0+s_{kj}+jT)-f(t_k+\tau_0+s_{kj})\}f(t_k+jT) \\ &\qquad + f(t_k+\tau_0+s_{kj})\{f(t_k+jT)-f(t_k)\}| \\ &< \frac{j}{MC}\varepsilon\,|f(t_k+jT)+f(t_k+\tau_0+s_{kj})| < \frac{2j}{M}\varepsilon, \end{aligned}$$

where the last relation is obtained from the bounded condition (2.4). Hence

$$\left|\frac{1}{M+1}\sum_{j=0}^{M}\{f(t_k+\tau_0+s_{kj}+jT)f(t_k+jT)-f(t_k+\tau_0+s_{kj})f(t_k)\}\right|$$
$$<\frac{2\varepsilon}{M(M+1)}\sum_{j=0}^{M} j=\varepsilon, \qquad \left(\because \sum_{j=0}^{M} j=\frac{M(M+1)}{2}\right)$$

which is what we must prove.

3. Observed Spectral Density

This theorem teaches us that, instead of the $M+1$ observations on the time points each of which differs by period T, we may adopt the same time observation of $M+1$ samples. From this theorem, if we use the formula

$$\varphi^*(\tau_0)=\frac{1}{(M+1)(N+1)}\sum_{k=0}^{N}\sum_{j=0}^{M} f(t_k+\tau_0+s_{kj})f(t_k)$$

instead of

$$\varphi(\tau_0)=\frac{1}{(M+1)(N+1)}\sum_{k=0}^{N}\sum_{j=0}^{M} f(t_k+\tau_0+s_{kj}+jT)f(t_k+jT),$$

the absolute value of the difference can be made as small as desired, by properly choicing of M and T. This relation exists even when the complex function is adopted, so that, we define the observed value of the autocorrelation function as follows:

$$\varphi^*(\tau_0)=\frac{1}{N+1}\sum_{k=0}^{N}\frac{1}{M+1}\sum_{j=0}^{M} f(t_k+\tau_0+s_{kj})\overline{f(t_k)}.$$

Multiplying $e^{-i\omega\tau_0}$, integrating over the domain $(-T, T)$, we have

$$\Phi_T^*(\omega)=\frac{1}{N+1}\sum_{k=0}^{N}\frac{1}{M+1}\sum_{j=0}^{M}\int_{-T}^{T} f(t_k+\tau_0+s_{kj})\overline{f(t_k)}\,e^{-i\omega\tau_0}d\tau_0$$
$$=\frac{A_T(\omega)}{(N+1)(M+1)}\sum_{k=0}^{N}\sum_{j=0}^{M}\overline{f(t_k)}\,e^{i\omega(t_k+s_{kj})},$$

where $A_T(\omega)$ is the Fourier transform of the function $f_T(t)$ which satisfies

$$f_T(t)=f(t) \qquad (-T<t<T)$$

and

$$f_T(t)=0 \qquad (|t|>T)$$

$s_{kj}\,(j=0,\dots M)$ may be considered as the $M+1$ samples of the fluctuation on the k th time point, and these $N+1$ fluctuations on $N+1$ time points may be assumed to be independent mutually. Therefore, it may be concluded that s_{kj} is independent of the suffix k, and is rewritten as s_j hereafter. When the number N becomes larger, summation may be replaced by integration.

$$\Phi_T^*(\omega)=\frac{A_T(\omega)}{M+1}\sum_{j=0}^{M} e^{i\omega s_j}\frac{1}{T}\int_0^T\overline{f(t)}\,e^{i\omega t}dt$$
$$=\frac{|A_T(\omega)|^2}{T}\,\frac{1}{M+1}\sum_{j=0}^{M} e^{i\omega s_j}$$
$$=\frac{|A_T(\omega)|^2}{T}\,\frac{1}{M+1}\sum_{j=0}^{M} s_j^k\sum_{k=0}^{\infty}\frac{(i)^k}{(k)!}\,\omega^k$$

Let us consider the case where M approaches ∞, and assume the Gaussian distribution; then we have

$$\lim_{M\to\infty}\frac{1}{M+1}\sum_{j=0}^{M}s_j^k=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}s^k e^{-\frac{s^2}{2\sigma^2}}ds$$
$$=\frac{(2n)!}{2^n n!}\sigma^{2n}\qquad (k=2n)$$

Hence
$$\Phi_T^*(\omega)=\frac{|A_T(\omega)|^2}{T}\sum_{n=0}^{\infty}(-)^n\left(\frac{\omega^2\sigma^2}{2}\right)^n$$
$$=\frac{|A_T(\omega)|^2}{T}e^{-\frac{\omega^2\sigma^2}{2}}$$

In the limit of $T\to\infty$, this becomes
$$\Phi^*(\omega)=\Phi(\omega)\,e^{-\frac{\omega^2\sigma^2}{2}},$$

where $\Phi(\omega)$ is the exact spectral density.

In Fig. 1 the full line denotes $\Phi(\omega)$ and dotted line $\Phi^*(\omega)$, and $AB/AC=\exp\left(-\frac{\sigma^2\omega^2}{2}\right)$. If we observe the ideal white noise whose spectrum is constant over all frequencies, the observed value is
$$\Phi^*(\omega)=ce^{-\frac{\omega^2\sigma^2}{2}}$$

In Fig. 2, the full line is the observed value and its effective bandwidth is $1/\sigma$.

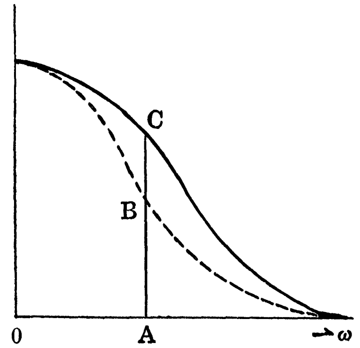

Fig. 1

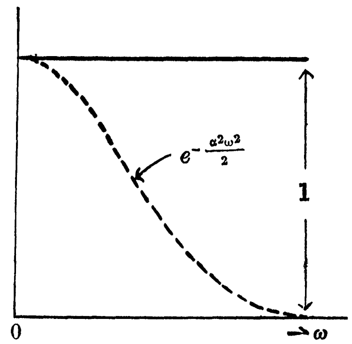

Fig. 2

This result is based on the uncertainty relation
$$\Delta\omega\Delta t=1,$$
and tells us that we cannot observe the ideal white noise, because the fluctuation of the delay time hinders the measurement of the high frequency region. In order to complete the theory, we consider the problem of random sampling.

The autocorrelation function in random sampling is given by the ergodic theorem as follows:
$$\varphi'(\tau_0)=\frac{1}{M+1}\sum_{j=0}^{M}\frac{1}{2N+1}\sum_{k=-N}^{N}f(ks_0+s_{kj}+\tau_0)\overline{f(ks_0+s_{kj})},$$

where s_0 is the uniform time interval. The Fourier transform of the both sides gives

$$\Phi'_T(\omega) = \frac{1}{M+1}\sum_{j=0}^{M} e^{is_{kj}(\omega-\omega')} \frac{A_T(\omega)}{2N+1} \sum_{k=-N}^{N} e^{iks_0(\omega-\omega')} \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{A_T(\omega')}\, d\omega',$$

where s_{kj} is independent of k as in the calculation of the delay time. By the relation

$$\sum_{k=-N}^{N} e^{iks_0(\omega-\omega')} = \frac{\sin (N+\frac{1}{2})\, s_0(\omega-\omega')}{\sin \frac{1}{2}\, s_0(\omega-\omega')},$$

we have

$$\Phi'_T(\omega) = \frac{1}{M+1}\sum_{j=0}^{M} e^{is_j\omega} \frac{A_T(\omega)}{(2N+1)s_0} \frac{s_0}{2\pi}\int_{-\infty}^{\infty} \frac{\sin(N+\frac{1}{2})s_0(\omega-\omega')}{\sin\frac{1}{2}s_0(\omega-\omega')} e^{-is_j\omega'}\overline{A_T(\omega')}d\omega'.$$

By the relation $T=(2N+1)\,s_0$, it becomes in the limit $T\to\infty$,

$$\Phi'_T(\omega) = \lim_{T\to\infty}\frac{|A_T(\omega)|^2}{T} \longrightarrow \Phi(\omega)\,.$$

Now we may conclude that it does not affect the spectral density to use the random sampling so long as the number of data is large.

4. Conclusions

The result obtained in §3 is necessary for us to estimate the loss of the information by the observation. If we measure the spectral density of the ensemble which is seemed to have perfect randomness theoretically, the observed spectral density curve will have the tendency to go down towards the high frequency region. From this data we can calculate the standard deviation of the delay time which may be definite for its instrument. Further measurements will be corrected using this value. The author hopes that this result will be supported by the experiments.