A NOTE ON DETERMINATION OF SAMPLE SIZE

By M. N. MURTHY

Indian Statistical Institute

SUMMARY. In this note a procedure of determining the sample size is proposed, where the idea is to fix the sample size in such a way that the probability (P) of the length (L) of the confidence interval (associated with a specified confidence coefficient) for the parameter (μ) being less than a given value ($k\mu$) is a pre-specified quantity.

Let y be normally distributed with mean μ and standard deviation σ. Suppose a sample of N units is drawn with equal probability. Then the mean $\bar{y}$ based on the N observations is normally distributed with mean μ and standard error $\sigma/\sqrt{N}$. Let s^2 be an unbiased estimator of σ^2 based on a sub-sample of n units (or on n random sub-groups of $\frac{N}{n}$ units),

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

It is well known that the statistic

$$t = \frac{\bar{y} - \mu}{s/\sqrt{N}}$$

is distributed as Student's t with $(n-1)$ degrees of freedom.

Using the tabulated values of the t-distribution, we can set up confidence interval for μ at any specified level of confidence $(1-\alpha)$. That is, if t_a is the $\alpha\%$ point of t, then

$$P\left[\frac{|\bar{y} - \mu|}{s/\sqrt{N}} < t_a \right] = 1-\alpha. \qquad \ldots \quad (1)$$

The length L of the confidence interval is given by

$$L = 2t_a s/\sqrt{N}. \qquad \ldots \quad (2)$$

Suppose the sample size is to be so fixed that

$$P(L < k\mu) = 1-\beta, \qquad \ldots \quad (3)$$

where k is a pre-specified quantity and $(1-\beta)$ may be taken as the *second level of confidence*, the first level of confidence being $(1-\alpha)$ in (1). It may be noted that $P(L < k\mu)$ is a function of the sample size and increases with increase in sample size.

For finding the sample size which could satisfy both the levels of confidence given in (1) and (3), we may proceed as follows.

$$P(L < k\mu) = 1-\beta,$$

381

that is,
$$P(2t_\alpha s/\sqrt{N} < k\mu) = 1-\beta,$$

that is,
$$P\left[\frac{(n-1)s^2}{\sigma^2} < \frac{k^2}{c^2} \frac{N(n-1)}{4t_\alpha^2} \right] = 1-\beta, \qquad \ldots \quad (4)$$

where c is the population coefficient of variation (σ/μ) and $(n-1)s^2/\sigma^2$ is a χ^2 with $(n-1)$ degrees of freedom. Reducing (4) to an incomplete Γ-function which is already tabulated, we get

$$P(L < k\mu) = I(u, p) \qquad \ldots \quad (5)$$

where
$$I(u, p) = \frac{1}{\Gamma(p+1)} \int_0^{u\sqrt{(p+1)}} e^{-x} x^p dx,$$

$$p = (n-1)/2, \text{ and } u = \frac{k^2}{c^2} \frac{N\sqrt{(n-1)}}{4\sqrt{2}t_\alpha^2}.$$

For given values of $(1-\alpha)$, $(1-\beta)$, c, n, and k we can first get the value of u such that
$$I(u, p) = 1-\beta$$

and then get the required sample size
$$N = u \frac{c^2}{k^2} \frac{4\sqrt{2}\, t_\alpha^2}{\sqrt{(n-1)}}. \qquad \ldots \quad (6)$$

The usual procedure of determining the sample size consisted in finding the value of N such that $E(L)$ is equal to a specified value $k\mu$. The proposed procedure given in this note is a generalization of the usual procedure in the sense that it ensures a pre-specified value for the probability that L is less than $k\mu$.

Paper received : February, 1962.

ON SPECTRAL ANALYSIS WITH MISSING OBSERVATIONS AND AMPLITUDE MODULATION*

By EMANUEL PARZEN

Stanford University

SUMMARY. The notion of an asymptotically stationary time series and its spectral analysis was considered by the author (Parzen, 1961b). An important example of an asymptotically stationary time series is an amplitude modulated stationary time series. In this note, the problem of spectral analysis of stationary normal time series with missing observations, recently treated by Jones (1962), is treated as a special case of the problem of spectral analysis of an amplitude modulated stationary normal time series.

1. INTRODUCTION

Let $\{X(t), t = 1, 2, ...\}$ be a discrete parameter time series with zero means and finite second moments. It is said to be *weakly* (see Doob, 1953) *or covariance* (see Parzen, 1962) *stationary* if there exists a function, denoted $R(v)$ and called the covariance function of the time series, such that for $v = 0, 1, 2, ...$,

$$R(v) = E[X(t) \ X(t+v)] \qquad ... \ (1.1)$$

independently of $t = 1, 2, ...$. It is said to be asymptotically (weakly) stationary if instead of (1.1) it holds that

$$R(v) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-v} E[X(t) \ X(t+v)]. \qquad ... \ (1.2)$$

If either (1.1) or (1.2) hold, the time series is said to be *ergodic* if the sample covariance function

$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T-v} X(t) \ X(t+v) \qquad ... \ (1.3)$$

is, for $v = 0, 1, ...$, a consistent in quadratic mean estimate of $R(v)$. In order for this to be the case it is necessary and sufficient that for each v

$$\lim_{T \to \infty} \text{var}[R_T(v)] = 0. \qquad ... \ (1.4)$$

One important way in which asymptotically stationary time series arise is by *amplitude modulating a (covariance) stationary process*.

Let $\{Y(t), t = 1, 2, ...\}$ be a stationary time series with zero means and covariance function

$$R_Y(v) = E[Y(t) \ Y(t+v)]. \qquad ... \ (1.5)$$

Let $\{g(t),\ t = 1,\ 2,\ ...\}$ be a non-random bounded function possessing a generalized harmonic analysis in the sense that for $v = 0,\ 1,\ ...$

$$R_g(v) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-v} g(t)\, g(t+v) \qquad \ldots \ (1.6)$$

exists. The time series

$$X(t) = g(t)\ Y(t) \qquad \ldots \ (1.7)$$

may be called the original time series $Y(\bullet)$ amplitude modulated by the function $g(\bullet)$. Since

$$E[X(t)\ X(t+v)] = g(t)\ g(t+v)\ R_Y(v) \qquad \ldots \ (1.8)$$

it is clear that while $X(\bullet)$ is not covariance stationary, it is asymptotically stationary with covariance function $R_X(v)$ given by

$$R_X(v) = R_g(v)\ R_Y(v). \qquad \ldots \ (1.9)$$

It is shown by the author (Parzen, 1961b) that if $Y(\bullet)$ is an ergodic normal process, then $X(\bullet)$ is ergodic. Consequently, given observations $\{X(t), t = 1,\ 2,\ ...,\ T\}$ a consistent (in quadratic mean) estimate of $R_X(v)$ is given by the sample covariance function

$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T} X(t)\ X(t+v). \qquad \ldots \ (1.10)$$

A consistent estimate of $R_Y(v)$ is then available at all lags v for which $R_g(v) \neq 0$, namely

$$\hat{R}_Y(v) = R_T(v)/R_g(v). \qquad \ldots \ (1.11)$$

From these facts we obtain immediately the following theorem.

Theorem 1A : Let $\{Y(t),\ t = 1,\ 2,\ ...\}$ be *stationary* and *normal with zero means* and *covariance function* $R_Y(v)$ *satisfying*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} R^2(v) = 0, \qquad \ldots \ (1.12)$$

so that $Y(\bullet)$ *is ergodic.*

Suppose that the time series $Y(\bullet)$ *is not directly observed. Rather one observes a time series* $X(\bullet)$ *which is an amplitude modulated version of* $Y(\bullet)$:

$$X(t) = g(t)\ Y(t), \qquad t = 1,\ 2,\ ..., \qquad \ldots \ (1.13)$$

where $g(\bullet)$ *is a non-random function possessing a covariance function* $R_g(v)$ *defined by* (1.6). *If*

$$R_g(v) \neq 0, \quad v = 0,\ 1,\ ..., \qquad \ldots \ (1.14)$$

a consistent in quadratic mean estimate of the covariance function $R_Y(v)$ *of the unobserved time series* $Y(\bullet)$ *is given by* (1.11).

Assume next that the series $Y(\cdot)$ possesses a spectral density function $f_Y(\omega)$ so that

$$R_Y(v) = \int_{-\pi}^{\pi} \cos v\omega f_Y(\omega) \, d\omega. \qquad \dots \ (1.15)$$

Given consistent estimates $\hat{R}_Y(v)$ of $R_Y(v)$, one may construct consistent estimates $\hat{f}_Y(\omega)$ of $f_Y(\omega)$ in a multitude of ways by suitably choosing the weights $k_T(v)$ in the formula

$$\hat{f}_Y(\omega) = \frac{1}{2\pi} \hat{R}_Y(0) + \frac{1}{\pi} \sum_{v=1}^{T} \cos v\omega \, k_T(v) \, \hat{R}_Y(v); \qquad \dots \ (1.16)$$

proofs of this assertion are essentially given in Parzen (1961a) and Parthasarathy (1960). We do not discuss this assertion further here since we will actually obtain a formula for the asymptotic variance of the estimate $f_Y(\omega)$.

2. MISSING OBSERVATIONS

There exist time series $\{Y(t), t = 1, 2, \dots\}$, defined at equally spaced intervals of time, which are systematically unobservable. For example, in radar studies of the surface of the moon, one observes a time series $Y(\cdot)$ which represents the *echo* (reflection from the moon) of a radar signal transmitted to the moon. In order to receive the echo, one must systematically cease transmission during the intervals in which one is receiving the echo. Another example of missing observations is the case of a time series which can be observed only during certain hours of the day.

A time series with missing observations seems to be best regarded as an amplitude modulated version of the original time series :

$$X(t) = g(t) \, Y(t), \quad t = 1, 2, \dots, \qquad \dots \ (2.1)$$

where (i) $Y(\cdot)$ is the time series under study, assumed to be defined at successive equally spaced points of time, (ii) $g(\cdot)$ is defined by

$$
\begin{aligned}
g(t) &= 0 \text{ if } Y(t) \text{ is missing at time } t, \\
&= 1 \text{ if } Y(t) \text{ is observed at time } t, \qquad \dots \ (2.2)
\end{aligned}
$$

and (iii) $X(\cdot)$ represents the actually observed values of $Y(\cdot)$, with 0 inserted in the series whenever the value of $Y(t)$ is missing.

A case of particular interest is the case of systematically missing observations. Suppose that the time series $Y(\cdot)$ is periodically observed for α time points, then not observed for β time points; then $g(\cdot)$ is a periodic function with period $\alpha + \beta$, and

$$
\begin{aligned}
g(t) &= 1 \quad \text{if} \quad t = 1, 2, \dots, \alpha, \\
&= 0 \quad \text{if} \quad t = \alpha+1, \dots, \alpha+\beta. \qquad \dots \ (2.3)
\end{aligned}
$$

It may be shown that a periodic function $g(\cdot)$ possesses a generalized harmonic analysis. If the period of $g(\cdot)$ is θ (for $g(\cdot)$ defined by (2.3), $\theta = \alpha + \beta$), then its covariance function $R_g(\cdot)$ has period θ and is given by

$$R_g(v) = \frac{1}{\theta} \sum_{t=1}^{\theta} g(t) \, g(t+\theta). \qquad \qquad \dots \ (2.4)$$

Thus for $g(\cdot)$ defined by (2.3), the covariance function $R_g(\cdot)$ has period $\alpha + \beta$. To determine its values for $0 \leqslant v \leqslant \alpha + \beta - 1$, we distinguish two cases : (i) $\alpha \leqslant \beta$ and (ii) $\alpha > \beta$.

TABLE 1. VALUES OF $R_g(v)$

case (i) : $\alpha \leqslant \beta$	case (ii) : $\alpha > \beta$
$\dfrac{\alpha-v}{\alpha+\beta}$, $v = 0, \dots, \alpha$,	$\dfrac{\alpha-v}{\alpha+\beta}$, $v = 0, \dots, \beta$
0, $v = \alpha, \dots, \beta$	$\dfrac{\alpha-\beta}{\alpha+\beta}$, $v = \beta, \dots, \alpha$
$\dfrac{v-\beta}{\alpha+\beta}$, $v = \beta, \dots, \alpha+\beta$	$\dfrac{v-\beta}{\alpha+\beta}$, $v = \alpha, \dots, \alpha+\beta$.

Only in the case $\alpha > \beta$ (one observes more values than one misses) does $R_g(v)$ never vanish. Therefore in order to be able to estimate $R_Y(v)$ we must assume that $\alpha > \beta$.

3. Variance of spectral estimates

In this section we find the variance of the estimated spectral density function $\hat{f}_Y(\omega)$ when it is formed from observations of an amplitude modulated time series $X(t)$ satisfying the assumptions of Theorem 1A. We first note that $\hat{f}_Y(\omega)$, defined by (1.16), can be written

$$\hat{f}_Y(\omega) = \frac{1}{2\pi T} \sum_{s,t=1}^{T} g(s) \, g(t) \, Y(s) \, Y(t)$$

$$\cos \omega(s-t) \, k_T(s-t) \, \{R_g(s-t)\}^{-1}. \qquad \dots \ (3.1)$$

In words, $\hat{f}_Y(\omega)$ is a quadratic form in the time series $Y(\cdot)$.

Let $a(s, t)$ be a symmetric function of two variables, and let

$$J_T |a(s, t)| = \sum_{s,t=1}^{T} a(s, t) \, Y(s) \, Y(t) \qquad \dots \ (3.2)$$

denote a quadratic form in the stationary normally distributed random variables $Y(\cdot)$ with covariance function $R_Y(\cdot)$ and spectral density function $f_Y(\cdot)$. It may be verified that

$$\mathrm{var}\,[J_T\,|\,a(s,\,t)\,|] = \sum_{s,t,u,\,v=1}^{T} a(s,\,t)\,a(u,\,v)\,\{R_Y(s-u)\,R_Y(t-v) + R_Y(s-v)\,R_Y(t-u)\}$$

$$= 2\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d\lambda_1\, d\lambda_2\, f_Y(\lambda_1) f_Y(\lambda_2)\,|\,A(\lambda_1,\,\lambda_2)\,|^2 \qquad \ldots \quad (3.3)$$

defining
$$A(\lambda_1,\,\lambda_2) = \sum_{s,t=1}^{T} a(s,\,t)\,\exp\,[i(s\lambda_1 + t\lambda_2)]. \qquad \ldots \quad (3.4)$$

The usual case considered in the theory of spectral analysis of stationary time series is the case where

$$a(s,\,t) = \frac{1}{2\pi T}\cos\omega(s-t)\,k_T(s-t) \qquad \ldots \quad (3.5)$$

and
$$k_T(v) = k(B_T v) \qquad \ldots \quad (3.6)$$

for a suitable weighting function $k(v)$ and constants B_T satisfying

$$B_T \to 0, \quad T B_T \to \infty \ \text{ as } \ T \to \infty; \qquad \ldots \quad (3.7)$$

for the exact conditions to be satisfied by the covariance averaging kernel $k(v)$ (see Parzen, 1957, p. 336). Define

$$K_T(\omega_1,\,\omega_2) = \frac{1}{2\pi T}\sum_{s,\,t=1}^{T} \exp\,[i(s\omega_1 + t\omega_2)]k_T(s-t). \qquad \ldots \quad (3.8)$$

By the argument employed in Parzen (1957, p. 342), one may show that for suitable functions $f(\lambda_1,\,\lambda_2)$ which are symmetric in the sense that

$$f(-\lambda_1,\,-\lambda_2) = f(\lambda_1,\,\lambda_2)$$

it holds that

$$\lim_{T\to\infty}\ TB_T \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(\lambda_1,\,\lambda_2)\ K_T(\lambda_1 + \omega_1,\,\lambda_2 + \omega_2)K_T(-\lambda_1 - \omega_3,\,-\lambda_2 - \omega_4)d\lambda_1 d\lambda_2$$

$$= \begin{cases} f(\omega_1,\,\omega_2)\int_{-\infty}^{\infty} k^2(u)du & \text{if}\quad \omega_1 = \omega_3,\, \omega_2 = \omega_4. \\ \\ 0 & \text{otherwise.} \end{cases} \qquad \ldots \quad (3.9)$$

In particular, (3.9) holds for a function $f(\omega_1,\,\omega_2)$ of the form

$$f(\omega_1,\,\omega_2) = \frac{1}{2\pi}\sum_{v_1,\,v_2 = -\infty}^{\infty} \exp\,[i(v_1\omega_1 + v_2\omega_2)]\,R(v_1,\,v_2) \qquad \ldots \quad (3.10)$$

where
$$\sum_{v_1,\,v_2 = -\infty}^{\infty} |\,R(v_1,\,v_2)\,| < \infty. \qquad \ldots \quad (3.11)$$

A careful derivation of (3.9) would unduly lengthen the present paper. However, let us sketch a proof. It suffices to show that (3.9) holds for

$$f(\lambda_1, \lambda_2) = \exp\left[i(\lambda_1 v_1 + \lambda_2 v_2)\right] \qquad \dots \text{(3.12)}$$

for arbitrary integers v_1 and v_2. Under (3.12), the double integral in (3.9) can be written

$$(4\pi^2 T)^{-1} B_T \sum_{s,\, t,\, u,\, v=1}^{T} k_T(s-t)\, k_T(u-v)\, J(s-u+v_1) J(t-v+v_2)\, \exp[i(s\omega_1 + t\omega_2 - u\omega_3 - v\omega_4)],$$

$$\dots \text{(3.13)}$$

defining $\qquad\qquad J(\alpha) = \int_{-\pi}^{\pi} e^{i\alpha\lambda}\, d\lambda = 2\,\dfrac{\sin \pi\alpha}{\alpha}. \qquad\qquad \dots \text{(3.14)}$

In (3.13) make the change of variables

$$x = s-u, \quad y = t-v, \quad z = u-v.$$

so that $\qquad\qquad s = x+u, \quad t = y-z+u, \quad v = u-z.$

Then (3.13) becomes

$$(4\pi^2)^{-1} B_T \sum_{x,\, y,\, z} J(x+v_1)\, J(y+v_2) k_T(z) k_T(z+x-y)$$

$$\exp\left[i\{x\omega_1 + y\omega_2 + z(\omega_4 - \omega_2)\}\right] \times \frac{1}{T} \sum_{u=1}^{T} \exp\left[iu(\omega_1 - \omega_3 + \omega_2 - \omega_4)\right]. \qquad \dots \text{(3.15)}$$

As T tends to ∞, (3.15) has the following limiting values : if $\omega_1 = \omega_3$ and $\omega_2 = \omega_4$, then it has the value

$$(4\pi^2)^{-1} \sum_{x,\, y=-\infty}^{\infty} J(x+v_1)\, J(y+v_2) \exp\left[i(x\omega_1 + y\omega_2)\right] \int_{-\infty}^{\infty} k^2(u) du \qquad \dots \text{(3.16)}$$

and 0 otherwise. To conclude the proof of (3.9) we need only note that

$$(2\pi)^{-1} \sum_{x=-\infty}^{\infty} J(x+v_1) \exp\left[ix\omega_1\right] = \exp\left[-iv_1\omega_1\right]. \qquad \dots \text{(3.17)}$$

We next show how using (3.9) one may derive an expression for the variance of the spectral density function of an amplitude modulated normal time series. We are then considering quadratic forms corresponding to

$$a(s, t) = \frac{1}{2\pi T} \cos\, \omega(s-t)\, k_T(s-t)\, h(s, t), \qquad \dots \text{(3.18)}$$

defining $\qquad\qquad h(s, t) = g(s)\, g(t)\, \{R_\varrho(s-t)\}^{-1}. \qquad\qquad \dots \text{(3.19)}$

We consider only the important special case that $g(t)$ is a periodic function. If $g(t)$ has period θ, then it possesses the harmonic representation

$$g(t) = \sum_{n=-N}^{N} e_n \exp\left[it\, \lambda_n\right]\, G_n \qquad \dots \text{(3.20)}$$

where $\lambda_n = 2\pi n/\theta$, $N = \theta/2$ or $(\theta-1)/2$ according as θ is even or odd, and

$$G_n = \frac{1}{\theta} \sum_{s=1}^{\theta} \exp\left[-is\lambda_n\right]\, g(s) \quad \text{for } n = 0,\, \pm 1,\, \dots,\, \pm[\theta/2], \qquad \dots \text{(3.21)}$$

while $e_n = 1$ for all n except that if θ is even $e_{+N} = 1/2$. It may be verified that $R_\rho(v)$ is an even function of period θ given by

$$R_\rho(v) = \sum_{n=-N}^{N} e_n G_n G_{-n} \exp[-iv\lambda_n]. \qquad \ldots \text{ (3.22)}$$

Now
$$g(s)\, g(t) = \sum_{m,\, n=-N}^{N} e_m e_n \exp\left[i(s\lambda_m + t\lambda_n)\right] G_m G_n$$

$$\{R_\rho(s-t)\}^{-1} = \sum_{m,\, n=-N}^{N} e_m e_n \exp\left[i(s\lambda_m + t\lambda_n)\right] W_{m,n} \qquad \ldots \text{ (3.23)}$$

where, defining
$$W_n = \frac{1}{\theta} \sum_{s=1}^{\theta} \exp\left[-is\,\lambda_n\right] \{R_\rho(s)\}^{-1},$$

$$W_{m,n} = W_n \quad \text{if } m = -n, \qquad \ldots \text{ (3.24)}$$
$$= 0 \quad \text{otherwise.}$$

Consequently
$$h(s,\, t) = \sum_{m,n=-N}^{N} e_m e_n \exp\left[i(s\lambda_m + t\lambda_n)\right] H_{m,n} \qquad \ldots \text{ (3.25)}$$

where
$$H_{m,n} = \sum_{j,k} e_j\, e_k\, W_{j,k}\, G_{j-m}\, G_{k-n} = \sum_j e_j\, W_j\, G_{j-m}\, G_{-j-n}. \quad \ldots \text{ (3.26)}$$

It should be noted that

$$H_{0,0} = \sum_j e_j\, W_j\, G_j\, G_{-j}$$
$$= \frac{1}{\theta} \sum_{s=1}^{\theta} R_\rho(s)\, \{(R_\rho(s)\}^{-1}$$
$$= 1. \qquad \ldots \text{ (3.27)}$$

We next write

$$A(\lambda_1,\, \lambda_2) = \frac{1}{2\pi T} \sum_{s,t} \cos \omega(s-t) k_T(s-t)\, h(s,\, t) \exp\left[i(s\lambda_1 + t\lambda_2)\right]$$

$$= \frac{1}{2\pi T} \sum_{m,n=-N}^{N} e_m\, e_n\, H_{m,n} \sum_{s,t} \cos \omega(s-t)\, k_T(s-t)$$
$$\exp\left[i(s\{\lambda_1 + \lambda_m\} + t\{\lambda_2 + \lambda_n\})\right]$$

$$= \sum_{m,n=-N}^{N} e_m\, e_n\, H_{m,n}\, \tfrac{1}{2}\{K_T(\lambda_1 + \lambda_m + \omega,\, \lambda_2 + \lambda_n - \omega)$$
$$+ K_T(\lambda_1 + \lambda_m - \omega,\, \lambda_2 + \lambda_n + \omega)\}. \qquad \ldots \text{ (3.28)}$$

We are now in a position to evaluate

$$TB_T \operatorname{var} [\hat{f}_Y(\omega)] = 2TB_T \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d\lambda_1\, d\lambda_2\, f_Y(\lambda_1) f_Y(\lambda_2)\, |A(\lambda_1,\, \lambda_2)|^2.$$

By (3.28) and (3.9) one sees that, as T tends to ∞,

$$TB_T \text{ var } [\hat{f}_Y(\omega)] \rightarrow \tfrac{1}{2} \sum_{m,n} e_m e_n \, |H_{m,n}|^2 \{f_Y(\lambda_m+\omega)f_Y(\lambda_n-\omega)+f_Y(\lambda_m-\omega)f_Y(\lambda_n+\omega)\} \int_{-\infty}^{\infty} k^2(u)du$$

$$= \left\{ \sum_{m,n} e_m e_n \, |H_{m,n}|^2 \, f_Y(\omega+\lambda_m) \, f_Y(\omega+\lambda_n) \right\} \int_{-\infty}^{\infty} k^2(u)du$$

$$= \left\{ f_Y^2(\omega) + \sum_{m \neq n} e_m e_n \, |H_{m,n}|^2 f_Y(\omega+\lambda_m) \, f_Y(\omega+\lambda_n) \right\} \int_{-\infty}^{\infty} k^2(u)du. \quad \ldots \quad (3.29)$$

The foregoing formula is valid for $0 < \omega < \pi$; it should be multiplied by 2 in the case that $\omega = 0$ or $\omega = \pi$.

If the spectral density function $f_Y(\omega)$ had been directly estimated from observations of the time series $Y(\cdot)$, the variance of the estimate $\hat{f}_Y(\omega)$ would satisfy (for $0 < \omega < \pi$)

$$\lim_{T \to \infty} TB_T \text{ var } [\hat{f}_Y(\omega)] = f_Y^2(\omega) \int_{-\infty}^{\infty} k^2(u)du. \quad \ldots \quad (3.30)$$

Consequently one can infer from (3.29) the effect on the variance of the estimate $\hat{f}_Y(\omega)$ due to the fact that it is formed from an amplitude modulated version of the time series $Y(\cdot)$. An upper bound to this variance is

$$TB_T \text{ var } [\hat{f}_Y(\omega)] \leqslant \bar{H} \{ \max_\omega f_Y^2(\omega) \} \int_{-\infty}^{\infty} k^2(u)du \quad \ldots \quad (3.31)$$

where

$$\bar{H} = \sum_{m,n} e_m e_n \, |H_{m,n}|^2. \quad \ldots \quad (3.32)$$

Thus $\bar{H}$ may be taken as a measure of the increase in variance due to amplitude modulation.

One may verify that

$$\bar{H} = \frac{1}{\theta^2} \sum_{s,t=1}^{\theta} h^2(s,t) = \frac{1}{\theta^2} \sum_{s,t=1}^{\theta} g^2(s) \, g^2(t) \, \{R_g(s-t)\}^{-2}. \quad \ldots \quad (3.33)$$

An upper bound for $\bar{H}$ can be obtained as follows. Let ρ be a lower bound for $R_g(v)$:

$$|R_g(v)| \geqslant \rho, \quad v = 0, 1, \ldots, \theta. \quad \ldots \quad (3.34)$$

Then

$$\bar{H} \leqslant \rho^{-2} \left\{ \frac{1}{\theta} \sum_{t=1}^{T} g^2(t) \right\}^2. \quad \ldots \quad (3.35)$$

An exact evaluation of $\bar{H}$ can be obtained from the formula

$$\bar{H} = \frac{1}{\theta} \sum_{v=-|\theta-1|}^{\theta-1} \{R_g(v)\}^{-2} \frac{1}{\theta} \sum_{t=1}^{\theta-|v|} g^2(t) \, g^2(t+|v|). \quad \ldots \quad (3.36)$$

To illustrate the use of these expressions we consider the modulating function $g(\cdot)$, defined by (3.3), which corresponds to the case of periodically missing observations. Then $\theta = \alpha + \beta$,

$$\rho = \frac{\alpha-\beta}{\alpha+\beta}, \quad \frac{1}{\theta}\sum_{t=1}^{\theta} g^2(t) = \frac{\alpha}{\alpha+\beta}. \qquad \ldots \quad (3.37)$$

By (3.35)

$$\bar{H} \leqslant \left(\frac{\alpha}{\alpha-\beta}\right)^2 = \left(\frac{1}{1-r}\right)^2 \qquad \ldots \quad (3.38)$$

defining

$$r = \frac{\beta}{\alpha} \qquad \ldots \quad (3.39)$$

to be the ratio of the number of observations missed to the number observed. An exact expression for $\bar{H}$ can be obtained from (3.36) : for $v = 0, 1..., \theta$

$$\frac{1}{\theta}\sum_{t=1}^{\theta-v} g^2(t)\, g^2(t+v) = \frac{\alpha-v}{\alpha+\beta}, \quad v < \alpha,$$

$$= 0, \qquad v \geqslant \alpha. \qquad \ldots \quad (3.40)$$

Consequently,

$$\{R_g(v)\}^{-2} \frac{1}{\theta}\sum_{t=1}^{\theta-v} g^2(t)\, g^2(t+v) = \frac{\alpha+\beta}{\alpha-v}, \quad v = 0, 1, ..., \beta;$$

$$= \frac{(\alpha-v)(\alpha+\beta)}{(\alpha-\beta)^2}, \quad v = \beta, ..., \alpha;$$

$$= 0, \quad v > \alpha.$$

Thus

$$(\alpha+\beta)\,\bar{H} = \frac{\alpha+\beta}{\alpha} + 2\sum_{v=1}^{\beta}\frac{\alpha+\beta}{\alpha-v} + 2\sum_{v=\beta+1}^{\alpha}\frac{(\alpha-v)(\alpha+\beta)}{(\alpha-\beta)^2}.$$

Finally, one obtains

$$\bar{H} = \frac{1}{\alpha} + \frac{\alpha-\beta-1}{\alpha-\beta} + 2\left\{\frac{1}{\alpha-1} + \frac{1}{\alpha-2} + \cdots + \frac{1}{\alpha-\beta}\right\}$$

$$= \frac{1}{\alpha} + \frac{2}{\alpha-1} + \frac{2}{\alpha-2} + \cdots + \frac{2}{\alpha-\beta+1} + \frac{\alpha-\beta+1}{\alpha-\beta}. \qquad \ldots \quad (3.41)$$

By replacing every denominator by $\alpha - \beta$, one obtains the following upper bound for $\bar{H}$:

$$\bar{H} \leqslant \frac{\alpha+\beta}{\alpha-\beta} = \frac{1+r}{1-r}. \qquad \ldots \quad (3.42)$$

One easily verifies that (3.42) provides a lower upper bound than does (3.38). In any event, both (3.38) and 3.42) provide some measure of how rapidly the variance of the spectral estimates increases as the ratio r tends to 1.

It may be of interest to express the variance of $\hat{f}_Y(\omega)$ in terms of the number T_0 of observations actually observed : approximately,

$$T_0 = \frac{\alpha}{\alpha+\beta} T. \qquad \qquad \dots \quad (3.43)$$

Combining (3.31) and (3.42) one sees that

$$\frac{T_0 B_T \text{ var } [\hat{f}_Y(\omega)]}{\left\{ \max_\omega \hat{f}_Y(\omega) \right\}^2 \int_{-\infty}^{\infty} k^2(u)du} \leqslant \frac{\alpha}{\alpha+\beta} \bar{H} = \frac{1}{1-r} \cdot$$

References

Doob, J. L. (1953) : *Stochastic Processes*, Wiley, New York.

Jones, R. H. (1962) : Spectral analysis with regularly missed observations. *Ann. Math. Stat.*, **32**, 455-461.

Parthasarathy, K. R. (1960) : On the estimation of the spectrum of a stationary stochastic process. *Ann. Math. Stat.*, **31**, 568-573.

Parzen, E. (1957) : On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.*, **28**, 329-348.

——— (1961a) : Mathematical considerations in the estimation of spectra *Technometrika*, **3**, 167-190.

——— (1961b) : Spectral analysis of asymptotically stationary time series. *Bulletin of the International Statistical Institute*, 33rd Session, Paris.

——— (1962) : *Stochastic Processes*, Holden-Day, San Francisco.

Paper received : October, 1962.