SPECTRAL ANALYSIS WITH REGULARLY MISSED OBSERVATIONS[1]

By Richard H. Jones

University of Stockholm

0. Summary. Estimating the spectral density of a discrete stationary stochastic process is studied for the case when the observations consist of repeated groups of α equally spaced observations followed by β missed observations, $(\alpha > \beta)$. The asymptotic variance of the estimate is derived for normally distributed variables. It is found that this variance depends not only on the value of the spectral density being estimated, but also on the spectral density at the harmonic frequencies brought in by the periodic method of sampling. Curves are presented for $\beta = 1$ showing the increase in the standard deviation and effective decrease in sample size as a function of α.

1. Introduction. When observing a stationary stochastic process at equally spaced intervals of time, it is sometimes necessary to occasionally miss observations for calibration or other purposes. The difficulty of estimating the spectral density in this case is not greatly increased, but in order to determine what is lost by this method of sampling, it is necessary to determine the increase in variance.

Given a sample of size N, $x_1, x_2, \cdots, x_N$, from a real stationary stochastic process of mean zero and continuous spectral density,

$$f(\lambda) = r_0 + 2\sum_{\nu=1}^{\infty} r_\nu \cos \nu\lambda, \qquad -\pi \leqq \lambda \leqq \pi, \qquad r_0 = Ex_t^2, \qquad r_\nu = Ex_t x_{t+\nu},$$

the usual method of estimating the spectrum is to form the empirical covariances,

$$r_\nu^* = \frac{1}{N-\nu}\sum_{i=1}^{N-\nu} x_i\, x_{i+\nu},$$

and use them in an estimate of the form $f^*(\lambda_0) = w_0^{(N)} r_0^* + 2\sum_{\nu=1}^{N-1} w_0^{(N)} r_\nu^* \cos \nu\lambda_0$, (Grenander and Rosenblatt, [2]). This is asymptotically the same as the quadratic form $f^*(\lambda_0) = (1/N)\sum_{\nu,\mu=1}^{N} x_\nu x_\mu w_{(\nu-\mu)}^{(N)} \cos(\nu - \mu)\lambda_0$, since it just replaces r_ν^* by $(1/N)\sum_{i=1}^{N-\nu} x_i x_{i+\nu}$. $w_N(\lambda) = w_0^{(N)} + 2\sum_{\nu=1}^{N-\nu} w_\nu^{(N)} \cos \nu\lambda$ is called the spectral window, and the expected value of the estimate is asymptotically equal to the convolution of the spectral density and the spectral window,

$$\lim_{N\to\infty} Ef^*(\lambda_0) = \lim_{N\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda) w_N(\lambda_0 - \lambda)\, d\lambda.$$

Choosing $w_N(\lambda)$ to be symmetric about $\lambda = 0$, the estimate of the spectral

Received September 8, 1961.

[1] This research was supported by the National Science Foundation.

density will be asymptotically unbiased and consistent if:

(a) $$\frac{1}{2\pi}\int_{-\pi}^{\pi} w_N(\lambda)\, d\lambda = 1$$

(b) for any $\epsilon > 0$, $w_N(\lambda) \to 0$ uniformly for $|\lambda| > \epsilon$ as $N \to \infty$.

(c) $$\lim_{N\to\infty} \frac{1}{N}\int_{-\pi}^{\pi} w_N^2(\lambda)\, d\lambda = 0.$$

(For discussions of these questions and how to choose the spectral windows, see [1], [2], and [3].)

2. Regularly missed observations. If the time series is sampled in groups of α equally spaced time points separated by β missed observations ($\alpha > \beta$), the empirical covariances can still be estimated, $r_\nu^* = (1/N_\nu) \sum x_i x_{i+\nu}$ where the summation is over the i's for which both x's are present, and N_ν is the number of terms. Then by proceeding in the same manner as before, an asymptotically unbiased and consistent estimate is obtained, but in general the variance will be greater.

One method of studying this situation is to assume that all the samples are present, but that the weight is zero when one of the samples is, in fact, missing. When this is done, the summations are over all consecutive values of the indices and can be summed in the usual manner. The estimate will then be the quadratic form $f^*(\lambda_0) = (1/N) \sum_{\nu,\mu=1}^{N} b_{\nu\mu} x_\nu x_\mu \cos (\nu - \mu)\lambda_0$, where $b_{\nu\mu} = w_{(\nu-\mu)} c_{(\nu-\mu)} a_\nu a_\mu$, $w_{(\nu-\mu)}$ being the usual weight, $c_{(\nu-\mu)}$ the modification necessary to make the estimate asymptotically unbiased, and a_ν equal to zero when x_ν is missing and one when x_ν is present. c_ν may be defined as

$$c_\nu = \lim_{N\to\infty} N \Big/ \left(\sum_{i=1}^{N-\nu} a_i\, a_{i+\nu} \right),$$

i.e., the limit of the ratio of N to the number of pairs available for estimating r_ν. $1 < c < \infty$ if $0 < \beta < \alpha$. Now

$$\begin{aligned} c_{\nu+\alpha+\beta} &= \lim_{N\to\infty} N \Big/ \left(\sum_{i=1}^{N-\nu-\alpha-\beta} a_i\, a_{i+\nu+\alpha+\beta} \right) = \lim_{N\to\infty} N \Big/ \left(\sum_{i=1}^{N-\nu-\alpha-\beta} a_i\, a_{i+\nu} \right) \\ &= \lim_{N\to\infty} N \Big/ \left(\sum_{i=1}^{N-\nu} a_i\, a_{i+\nu} - \sum_{i=N-\nu-\alpha-\beta}^{N-\nu} a_i\, a_{i+\nu} \right) = c_\nu \end{aligned}$$

since $a_{i+\alpha+\beta} = a_i$ and $|\sum_{i=N-\nu-\alpha-\beta}^{N-\nu} a_i a_{i+\nu}| \leqq \alpha + \beta$. Therefore, c_ν is cyclic with period $\alpha + \beta$, and it is only necessary to determine $c_0, \cdots c_{\alpha+\beta-1}$. Writing $N = k(\alpha + \beta)$, the number of pairs available for estimating r_ν,

$$0 \leqq \nu \leqq \alpha + \beta - 1,$$

from within the same group is $k(\alpha - \nu)$ for $\nu \leqq \alpha$ and 0 otherwise. From adjoining groups there are $(k - 1)$ $(\nu - \beta)$ for $\nu \geqq \beta$ and 0 otherwise. This covers all possibilities so

$$c_\nu = \lim_{k\to\infty} \frac{k(\alpha+\beta)}{k(\alpha-\nu)} = \frac{\alpha+\beta}{\alpha-\nu} \qquad 0 \leqq \nu \leqq \beta$$

$$c_\nu = \lim_{k\to\infty} \frac{k(\alpha+\beta}{k(\alpha-\nu)+(k-1)(\nu-\beta)} = \frac{\alpha+\beta}{\alpha-\beta} \qquad \beta \leqq \nu \leqq \alpha$$

$$c_\nu = \lim_{k\to\infty} \frac{k(\alpha+\beta)}{(k-1)(\nu-\beta)} = \frac{\alpha+\beta}{\nu-\beta}\, \alpha \leqq \nu \leqq \alpha+\beta.$$

c_ν is real with $c_\nu = c_{-\nu}$, and has the spectral representation

$$c_\nu = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos \nu\lambda \, dW(\lambda).$$

Because of the cyclic nature of the c's, the function $W(\lambda)$ is a step function with $\alpha+\beta$ steps of height W_k at $\lambda_k = 2\pi k/(\alpha+\beta)$, $k = -(\alpha+\beta-1)/2$, $-(\alpha+\beta-3)/2, \cdots (\alpha+\beta-1)/2$ if $\alpha+\beta$ is odd, or $k = -(\alpha+\beta-2)/2$, $-(\alpha+\beta-4)/2, \cdots (\alpha+\beta)/2$ if $\alpha+\beta$ is even. The steps are symmetric about $\lambda = 0$, and may take any real value with $\sum_k W_k = 1 + \beta/\alpha$. Inverting this representation gives

$$(\alpha+\beta)W_k = c_0 + 2c_1 \cos\frac{2\pi k}{\alpha+\beta} + 2c_2\cos\frac{4\pi k}{\alpha+\beta}$$

$$+\cdots+\begin{cases} 2c_{\frac{1}{2}(\alpha+\beta-1)} \cos \dfrac{(\alpha+\beta-1)\pi k}{\alpha+\beta} & \text{if } \alpha+\beta \text{ is odd} \\ 2c_{\frac{1}{2}(\alpha+\beta-2)} \cos \dfrac{(\alpha+\beta-2)\pi k}{\alpha+\beta} + c_{\frac{1}{2}(\alpha+\beta)} \cos \pi k & \text{if } \alpha+\beta \text{ is even}. \end{cases}$$

Using

$$(1) \qquad \frac{1}{\alpha+\beta}\sum_{k=m}^{m+\alpha+\beta} e^{2\nu k\pi i/(\alpha+\beta)} = \begin{cases} 0 \text{ for } \nu \neq n(\alpha+\beta) \\ 1 \text{ for } \nu = n(\alpha+\beta) \end{cases},$$

where ν, m and n are integers, it may be written

$$W_k = \frac{\alpha+\beta}{\alpha-\beta}\delta_{k0} - \frac{1}{\alpha-\beta}\left[\frac{\beta}{\alpha} + 2\sum_{\nu=1}^{\beta-1}\frac{\beta-\nu}{\alpha-\nu}\cos\frac{2\nu k\pi}{\alpha+\beta}\right], \qquad \delta_{\nu\mu} = \begin{cases} 0 \text{ if } \nu \neq \mu \\ 1 \text{ if } \nu = \mu \end{cases}.$$

Similarly, $a_\nu a_\mu$ has a two-dimensional spectral representation which is also discrete. Writing $a_\nu = \sum_k A_k e^{-i\nu\lambda_k}$, this may be written $a_\nu a_\mu = \sum_{kl} A_k \bar{A}_l e^{-i\nu\lambda_k + i\mu\lambda_l}$. Since this depends on the choice of the origin it is not stationary. Taking a_1, $a_2, \cdots, a_\alpha = 1$, and $a_{\alpha+1}$, $a_{\alpha+2}, \cdots, a_{\alpha+\beta} = 0$, the inverse becomes

$$(\alpha+\beta)A_k = \sum_{\nu=1}^{\alpha} e^{2\nu k\pi i/(\alpha+\beta)},$$

which may be written $A_k = \delta_{k0} - \sum_{\nu=0}^{\beta-1} e^{-2\nu k\pi i/(\alpha+\beta)}$.

The spectral representation of $b_{\nu\mu} = w_{(\nu-\mu)}c_{(\nu-\mu)}a_\nu a_\mu$ is the convolution of the spectral representations of the three factors, and will be two-dimensional because of the non-stationarity. The two-dimensional representation for the c's is

$$c_{(\nu-\mu)} = \sum_{kl} W_{kl} e^{-i\nu\lambda_k + i\mu\lambda_l'}, \qquad W_{kl} = \delta_{kl} W_k ,$$

i.e., $\alpha + \beta$ discrete steps which fall along the diagonal $\lambda = \lambda'$. The convolution of the two discrete components, $H_{kl} = W_{kl} * A_k \bar{A}_l$, is the finite sum

$$H_{kl} = \sum_{\nu,\mu} W_{\nu\mu} A_{k-\nu} \bar{A}_{l-\mu} = \sum_{\nu} W_\nu A_{k-\nu} \bar{A}_{l-\nu} ,$$

where $\nu,\ \mu = -(\alpha + \beta - 1)/2,\ -(\alpha + \beta - 3)/2, \cdots,\ (\alpha + \beta - 1)/2$ if $\alpha + \beta$ is odd, or $-(\alpha + \beta - 2)/2,\ -(\alpha + \beta - 4)/2, \cdots,\ (\alpha + \beta)/2$ if $\alpha + \beta$ is even. Substituting the values for A_k and W_k gives

$$H_{kl} = \sum_\nu \frac{1}{\alpha - \beta}\left[(\alpha + \beta)\lambda_{\nu 0} - \frac{\beta}{\alpha} - 2\sum_{m=1}^{\beta-1} \frac{\beta - m}{\alpha - m} \cos\frac{2\nu m\pi}{\alpha + \beta}\right]$$
$$\cdot \left[\delta_{k\nu} - \frac{1}{\alpha + \beta}\sum_{m=0}^{\beta-1} e^{2\pi i m(k-\nu)/(\alpha+\beta)}\right]\left[\delta_{l\nu} - \frac{1}{\alpha + \beta}\sum_{m=0}^{\beta-1} e^{2\pi i m(l-\nu)/(\alpha+\beta)}\right].$$

This may be simplified using equation (1) giving

$$\begin{aligned}
H_{kl} = {} & \frac{\alpha + \beta}{\alpha - \beta}\,\delta_{k0}\,\delta_{l0} - \frac{\delta_{l0}}{\alpha - \beta}\sum_{m=0}^{\beta-1} e^{-2\pi i m k/(\alpha+\beta)} - \frac{\delta_{k0}}{\alpha - \beta}\sum_{m=0}^{\beta-1} e^{2\pi i m l/(\alpha+\beta)} \\
& - \frac{\delta_{kl}}{\alpha - \beta}\left(\frac{\beta}{\alpha} + 2\sum_{m=1}^{\beta-1}\frac{\beta - m}{\alpha - m}\cos\frac{2km\pi}{\alpha + \beta}\right) \\
& + \frac{1}{(\alpha + \beta)(\alpha - \beta)}\sum_{m,n=0}^{\beta-1} e^{-2\pi i(mk - nl)/(\alpha+\beta)} \\
& + \frac{\beta}{\alpha(\alpha + \beta)(\alpha - \beta)}\sum_{m=0}^{\beta-1} e^{-2\pi i m(k-l)/(\alpha+\beta)} \\
& + \frac{2}{(\alpha + \beta)(\alpha - \beta)}\sum_{m=0}^{\beta-1} e^{-2\pi i m(k-l)/(\alpha+\beta)} \sum_{n=1}^{\beta-1}\frac{\beta - n}{\alpha - n}\left(\cos\frac{2\pi n k}{\alpha + \beta} + \cos\frac{2\pi n l}{\alpha + \beta}\right) \\
& - \frac{1}{(\alpha + \beta)(\alpha - \beta)}\sum_{m=1}^{\beta-1}\left[\frac{\beta - m}{\alpha - m}\left(e^{2\pi i m l/(\alpha+\beta)} + e^{-2\pi i m k/(\alpha+\beta)}\right)\right. \\
& \qquad \left. \cdot \sum_{n=0}^{\beta-1-m} e^{-2\pi i n(k-l)/(\alpha+\beta)}\right].
\end{aligned}$$

When $k = l$, this reduces to $H_{kk} = \delta_{k0}$. For $\beta = 1$, $k \neq l$,

$$H_{kl} = \frac{1}{\alpha - 1}\left(\frac{1}{\alpha} - \delta_{k0} - \delta_{l0}\right).$$

Two schematic examples of these calculations are shown in Figure 1, for $\alpha + \beta$ even, and $\alpha + \beta$ odd. It should be noted that $A_k \bar{A}_l$ and H_{kl} take real values only when $\beta = 1$, as this is the only case where $\alpha_\nu = a_{-\nu}$. Although the

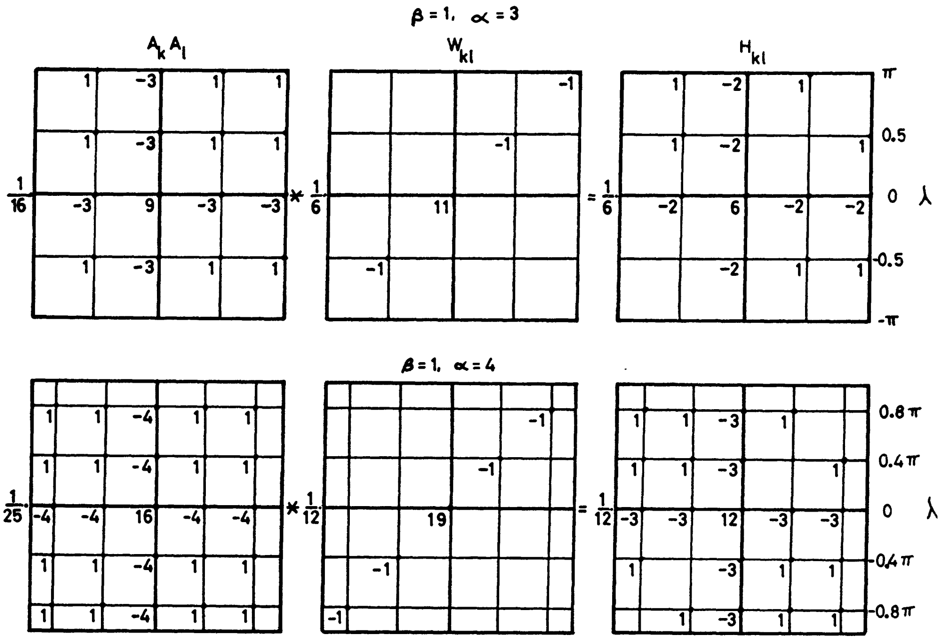

FIG. 1. Schematic representation of discrete convolutions

expression for H_{kl} appears rather formidable for $\beta > 1$, the calculations would not be difficult for modern computers.

3. Variance of the estimate. When using the two-dimensional spectral representation,

$$w_N(\lambda, \lambda') = \sum_{\nu,\mu=1}^{N} w_{(\nu-\mu)}^{(N)} e^{i\nu\lambda - i\mu\lambda'},$$

the asymptotic variance of the estimate

$$f^*(\lambda_0) = \frac{1}{N} \sum_{\nu,\mu=1}^{N} x_\nu \, x_\mu \, w_{(\nu-\mu)}^{(N)} \, e^{i(\nu-\mu)\lambda_0}$$

is (assuming normally distributed variables)

$$\frac{2}{(2\pi N)^2} \iint_{-\pi}^{\pi} | \, w_N(\lambda - \lambda_0 \, , \lambda' - \lambda_0) \, |^2 f(\lambda) f(\lambda') \, d\lambda d\lambda'$$

$$= \frac{1}{\pi N} \int_{-\pi}^{\pi} | \, w_N(\lambda - \lambda_0) \, |^2 f^2(\lambda) \, d\lambda \sim \frac{f^2(\lambda_0)}{\pi N} \int_{-\pi}^{\pi} w_N^2(\lambda) \, d\lambda,$$

(Grenander and Rosenblatt, [2]). The convolution of $w_N(\lambda, \lambda')$ with H_{kl} is simply $H_{kl} w_N(\lambda - \lambda_k \, , \lambda' - \lambda_l)$ if $w_N(\lambda)$ has all its area inside $-\pi/(\alpha + \beta) < \lambda < \pi/(\alpha + \beta)$. This will always be the case for large N if the estimate is to be asymptotically unbiased. The contribution to the variance from this term is

$$\frac{2|H_{kl}|^2}{(2\pi N)^2}\iint_{-\pi}^{\pi} | w_N(\lambda - \lambda_k - \lambda_0, \lambda' - \lambda_l - \lambda_0) |^2 f(\lambda)f(\lambda')\, d\lambda d\lambda'$$

$$\sim \frac{|H_{kl}|^2 f(\lambda_0 + \lambda_k)f(\lambda_0 + \lambda_l)}{\pi N}\int_{-\pi}^{\pi} w_N^2(\lambda)\, d\lambda.$$

Therefore,

$$D^2 f^*(\lambda_0) = \frac{1}{\pi N}\left[\sum_{kl} |H_{kl}|^2 f(\lambda_0 + \lambda_k)f(\lambda_0 + \lambda_l)\right]\int_{-\pi}^{\pi} w_N^2(\lambda)\, d\lambda,$$

or since $H_{kk} = \delta_{k0}$, and $H_{kl} = \overline{H_{lk}}$, this may be written

$$D^2 f^*(\lambda_0) = \frac{1}{\pi N}\left[f^2(\lambda_0) + 2\sum_{k>l} |H_{kl}|^2 f(\lambda_0 + \lambda_k)f(\lambda_0 + \lambda_l)\right]\int_{-\pi}^{\pi} w_N^2(\lambda)\, d\lambda.$$

The first term is the same as the variance when no observations are missing, and the other terms involve harmonic frequencies brought in by the periodic method of sampling. An upper bound to this variance which is of interest is

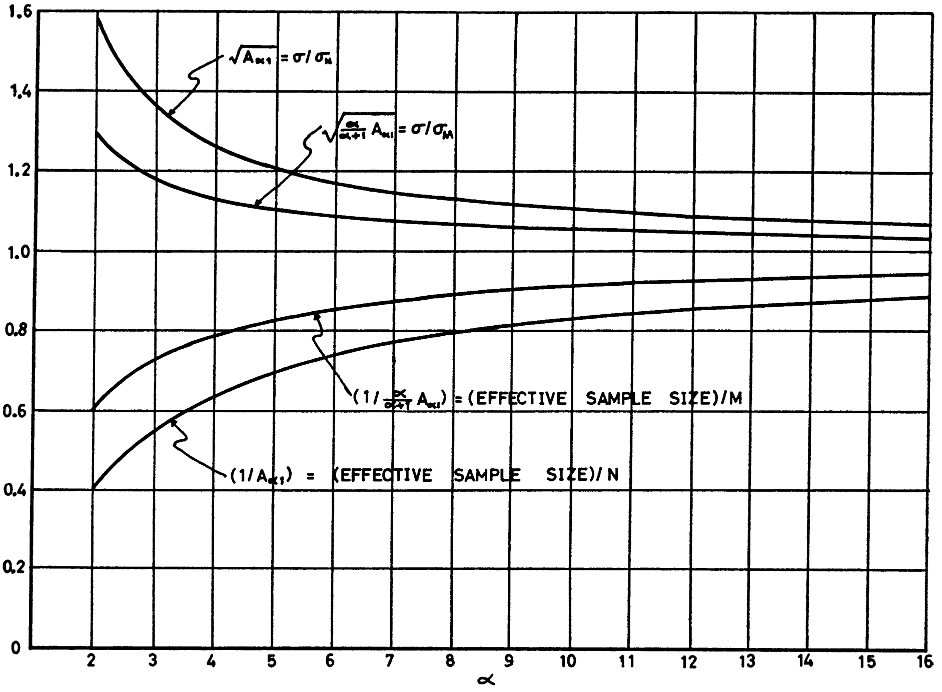

FIG. 2. $\beta = 1$, $A_{\alpha 1} = 1 + (2\alpha - 1)/\alpha(\alpha - 1)$. N = sum of sampled and missed points. $M = \alpha N/(\alpha + 1)$ = actual sample size. σ = "average" s.d. with missed observations. σ_N = s.d. of estimate with all N samples present. σ_M = s.d. of estimate with M equally spaced observations.

$$D^2 f^*(\lambda_0) \leqq \frac{1}{\pi N}\left[1 + 2\sum_{k>l}|H_{kl}|^2\right] f^2_{\max}(\lambda)\int_{-\pi}^{\pi} w_N^2(\lambda)\, d\lambda.$$

$A_{\alpha\beta} = [1 + 2\sum_{k>l}|H_{kl}|^2]$ is the important quantity, as it is an "average" value for the increase in variance, or the effective decrease in sample size. The simplest case, but probably the most important, is $\beta = 1$. This arises when it is necessary to periodically miss an observation in order to observe a standard for calibration purposes. For $\beta = 1$, $k \neq l$, $A_{\alpha 1} = 1 + (2\alpha - 1)/\alpha(\alpha - 1)$. Figure 2 shows the standard deviation and effective sample size compared to N, the total of observed and missed points, and compared to $N\alpha/(\alpha + 1)$, the actual sample size.

REFERENCES

[1] Blackman, R. B. and Tukey, J. W. (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering*. Dover Pub., New York.

[2] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[3] Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics*, **3** 167–190.