

SPECTRUM ESTIMATION WITH MISSING OBSERVATIONS*

RICHARD H. JONES

(Received Dec. 2, 1969; revised May 10, 1971)

1. Introduction

An observed stationary time series may contain missing values, and yet it may be necessary to estimate the spectral density. If a very small percentage of the observations is missing, any logical method of allowing for the missing values may produce acceptable results. The simplest procedure would be to subtract the sample mean from the available data and insert zeroes for the missing values. A standard method of estimating the spectral density could then be used. When a larger percentage of the observations is missing, this technique may be inadequate.

The approach considered here assumes the data consist of observations of a discrete time stationary process together with the times of the observations. Since the observation times are known, it is irrelevant whether the missed data are caused by some random mechanism or by some deterministic procedure. Some remarks on this problem appear in Jones [4], and the case of regularly missed observations is discussed in Jones [5]. Parzen [6] generalized regularly missed observations to include amplitude modulated time series where the random process is multiplied by a known non-random function of time.

The results of this paper are also generalized in the case where the process is multiplied by a known non-random function of time. Therefore, the results can be applied when the data is first multiplied by a data window [1].

2. Preliminaries

Let x_t be a discrete time, weakly stationary complex stochastic process with mean zero and covariance function

$$(2.1) \quad r_t = E\{x_{s+t}\bar{x}_s\},$$

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Contract No. F44620-69-C-0035.

for all integers s and t . The unit of time is taken to be the sampling interval. It will be assumed that r_t is absolutely summable so that there exists a continuous spectral density,

$$(2.2) \quad s(f) = \sum_{t=-\infty}^{\infty} r_t e^{2\pi i t f}, \quad |f| \leq \frac{1}{2}.$$

Data are collected over a span of n time points, except that some of the observations are missing. Let I_t be an indicator sequence defined as

$$(2.3) \quad I_t = \begin{cases} 1 & \text{if } x_t \text{ is observed} \\ 0 & \text{if } x_t \text{ is missing or unobserved.} \end{cases}$$

It is assumed that the observation times are known.

It is possible to obtain an unbiased estimate of each covariance for which at least one pair of observations separated by the proper lag exists,

$$(2.4) \quad \hat{r}_t = \frac{1}{c_t} \sum_s x_{s+t} I_{s+t} \bar{x}_s I_s, \quad t=0, 1, 2, \dots,$$

where

$$(2.5) \quad c_t = \sum_s I_{s+t} I_s$$

assuming that $c_t > 0$. Also, $c_{-t} = c_t$ and \hat{r}_{-t} is the complex conjugate of \hat{r}_t . If for the given sequence of observations it is possible to choose a truncation point m , such that

$$c_t > 0 \quad \text{for all } t \leq m,$$

it is possible to use a standard lag window to obtain an estimate of the spectral density,

$$(2.6) \quad \hat{s}(f) = \sum_{t=-\infty}^{\infty} \hat{r}_t w_t e^{2\pi i t f}$$

where w_t is a sequence of real weights with

$$w_0 = 1,$$

$$w_{-t} = w_t$$

and

$$w_t = 0 \quad \text{for } |t| > m.$$

It follows that

$$(2.7) \quad E\{\hat{s}(f)\} = \sum_{t=-\infty}^{\infty} r_t w_t e^{2\pi i t f} \\ = \int_{-1/2}^{1/2} s(u) W(f-u) du,$$

where

$$(2.8) \quad W(f) = \sum_{t=-\infty}^{\infty} w_t e^{2\pi i t f}$$

is the spectral window. Subject to the assumptions which have been stated, the analysis has proceeded as when no observations are missed. However, additional care is necessary when studying the variance of the estimate.

3. Variance of the estimate

The usual asymptotic results for the variance of spectrum estimates do not apply in the present situation. In fact, no asymptotics are considered here, the sample size and times of the observations being assumed fixed. The estimate of the spectral density can be written as a quadratic form,

$$(3.1) \quad \hat{s}(f) = \sum_{s=1}^n \sum_{t=1}^n x_s \bar{x}_t a_{st} e^{2\pi i (s-t)f}$$

where

$$(3.2) \quad a_{st} = \frac{I_s w_{s-t} I_t}{c_{s-t}}.$$

Using the spectral representation of a stationary process

$$(3.3) \quad x_t = \int_{-1/2}^{1/2} e^{-2\pi i t f} dZ(f),$$

where $Z(f)$ is a process with mean zero, orthogonal increments and

$$(3.4) \quad E\{|dZ(f)|^2\} = s(f)df,$$

gives

$$(3.5) \quad \hat{s}(f) = \int \int_{-1/2}^{1/2} A(f-u, f-v) dZ(u) \overline{dZ(v)}$$

where

$$(3.6) \quad A(u, v) = \sum_{s=1}^n \sum_{t=1}^n a_{st} e^{2\pi i (su-tv)}.$$

The expected value of (3.5) is

$$(3.7) \quad E\{\hat{s}(f)\} = \int_{-1/2}^{1/2} A(f-u, f-u)s(u)du$$

which is identical to equation (2.7) since

$$(3.8) \quad W(f) = A(f, f).$$

For a Gaussian process, the covariance of two estimates is

$$(3.9) \quad \text{Cov}\{\hat{s}(f_1), \hat{s}(f_2)\} = \int \int_{-1/2}^{1/2} A(f_1-u, f_1-v)\overline{A(f_2-u, f_2-v)}s(u)s(v)dudv.$$

For a real process, the symmetry condition

$$(3.10) \quad dZ(u) = \overline{dZ(-u)}$$

introduces another term on the right-hand side of (3.9),

$$(3.11) \quad \int \int_{-1/2}^{1/2} A(f_1-u, f_1-v)\overline{A(f_2+v, f_2+u)}s(u)s(v)dudv.$$

This term has the effect of doubling the variance of the spectrum estimate of a real process at the frequencies 0 and $\pm 1/2$, and is essentially zero when f_1+f_2 is not in the neighbourhood of 0 or ± 1 .

The variance is

$$(3.12) \quad \text{Var}\{\hat{s}(f)\} = \int \int_{-1/2}^{1/2} |A(f-u, f-v)|^2 s(u)s(v)dudv.$$

This equation for the variance is also given in Parzen [6]. Again, for a real process there is a second term

$$(3.13) \quad \int \int_{-1/2}^{1/2} A(f-u, f-v)\overline{A(f+v, f+u)}s(u)s(v)dudv,$$

which contributes only near the frequencies 0 and $\pm 1/2$.

$A(u, v)$ is a two-dimensional Hermitian kernel which has the spectral window along the main diagonal. The spectral window has unit area with the area concentrated near the origin. When estimating a spectral density without missing observations, often positive definite estimates are used. In this case $A(u, v)$ will be a positive definite kernel so that any off diagonal point has an absolute value which can not be greater than the square root of the product of the corresponding diagonal values.

$$(3.14) \quad |A(u, v)| \leq [A(u, u)A(v, v)]^{1/2}.$$

When observations are missing, as long as a sufficient number of lags can be estimated, the shape of the spectral window ($A(f, f)$ or $W(f)$) can be controlled; however, the estimate is not necessarily positive definite so the inequality (3.14) may not hold. This could be called *vari-*

ance leakage. This phenomenon was noted for regularly missed observations Jones [4], where harmonic frequencies of the sampling period entered into the variance.

The expression for the variance (3.12) is exact, but contains the unknown spectral density. The asymptotic theory without missing observations handles this problem by using a sequence of kernels which approaches a delta function at zero as the sample size becomes large. This enables the factor $s^2(f)$ to be moved outside the integral sign. For finite sample size, it is necessary to argue that the spectral density can be assumed constant where $W(f)$ is significantly different from zero so that it can be moved outside the integral sign. With missing observations, this argument becomes less convincing due to the variance leakage. Therefore, it is necessary to set some standard for comparison purposes. What is often called the asymptotic variance will be called here the white noise variance since it is the variance obtained by assuming that $s(f)$ is constant. Therefore we define

$$(3.15) \quad \text{Var}_w \{\hat{s}(f)\} = s^2(f) \int \int_{-1/2}^{1/2} |A(u, v)|^2 du dv \\ = s^2(f) \sum_{s=1}^n \sum_{t=1}^n a_{st}^2 .$$

Substituting the values of a_{st} from (3.2) gives

$$(3.16) \quad \text{Var}_w \{\hat{s}(f)\} = s^2(f) \sum_{t=-m}^m w_t^2 / c_t .$$

At this point it is possible to relate the above result to the case when no observations are missed. In this case $c_t = n - |t|$, so

$$(3.17) \quad \text{Var}_w \{\hat{s}(f)\} = s^2(f) \sum_{t=-m}^m w_t^2 / (n - |t|) .$$

When the positive definite estimate of the covariance functions is used,

$$(3.18) \quad \frac{1}{n} \sum_{s=1}^{n-t} x_{s+t} \bar{x}_s = \frac{n-t}{n} \hat{r}_t .$$

the estimate is then

$$(3.19) \quad \hat{s}(f) = \sum_{t=-m}^m w_t \left(\frac{n-|t|}{n} \right) \hat{r}_t e^{2\pi i t f}$$

so

$$(3.20) \quad \text{Var}_w \{\hat{s}(f)\} = s^2(f) \frac{1}{n} \sum_{t=-m}^m \left(\frac{n-|t|}{n} \right) (w_t)^2 .$$

These could be considered finite sample versions of the usual asymptotic

formula.

Equation (3.16) also gives some insight into the missing observation problem. Suppose there are n' actual observations, $n-n'$ being the number of missing observations. If c_t is close to $n'-t$ for $t=0, 1, \dots, m$, the properties of the estimate of the spectrum will not differ much from an estimate obtained using a sample of size n without missing observations. However, if some of the c_t are low, the estimate will suffer from variance leakage even though the first order properties (spectral window) can be held unchanged. Missing observations always causes an increase in equation (3.16) since $c_t < n'-t$ for $t=1, 2, \dots$ if an observation is missing within the data span.

Following Blackman and Tukey [1], the equivalent degrees of freedom can be defined as

$$(3.21) \quad edf = 2 / \left(\sum_{t=-m}^m w_t^2 / c_t \right).$$

For a Gaussian process it is also possible to calculate what could be called the white noise covariance between estimates at different frequencies.

$$(3.22) \quad \text{Cov}_w \{ \hat{s}(f_1), \hat{s}(f_2) \} = s(f_1)s(f_2) \sum_t \frac{w_t^2}{c_t} \cos [2\pi t(f_1 - f_2)].$$

It is easy to plot

$$(3.23) \quad \rho(f) = \left[\sum_t (w_t^2 / c_t) \cos 2\pi t f \right] / \left[\sum_t w_t^2 / c_t \right]$$

to give an idea of the correlation between two spectrum estimates separated by a frequency interval f .

4. Data window

Suppose that before calculating the sum of lagged products, the data had been multiplied by a known real function of time b_t . This function is assumed to take the value zero if the observation is missing.

Therefore, it replaces I_t of Section 2. Now c_t can be calculated as before

$$(4.1) \quad c_t = \sum_s b_{s+t} b_s,$$

but to complete the analysis, it is also necessary to calculate

$$(4.2) \quad d_t = \sum_s b_{s+t}^2 b_s^2.$$

Note that $d_t = c_t$ when $b_t = I_t$. Now

$$(4.3) \quad a_{st} = \frac{b_s w_{s-t} b_t}{c_{s-t}}$$

so that (3.16) becomes

$$(4.4) \quad \text{Var}_w \{\hat{s}(f)\} = s^2(f) \sum_{t=-m}^m \frac{d_t}{c_t^2} w_t^2 .$$

Also

$$(4.5) \quad edf = 2 / \left(\sum_{t=-m}^m d_t w_t^2 / c_t^2 \right)$$

and

$$(4.6) \quad \rho(f) = \left[\sum_t (d_t w_t^2 / c_t^2) \cos 2\pi t f \right] / \left[\sum_t d_t w_t^2 / c_t^2 \right] .$$

5. Numerical considerations

An efficient method for calculating c_t from equation (2.5) is to use fixed point arithmetic with a short word length. The word length would only need to be long enough to accommodate an integer the size of the number of observations.

The analysis can be approached through the direct Fourier transformation of the data. When using the fast Fourier transform [2] it is often faster to Fourier transform twice rather than calculate a numerical convolution. It is well known that the periodogram

$$(5.1) \quad p(f) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{2\pi i t f} \right|^2$$

is also equal to

$$(5.2) \quad p(f) = \sum_{t=-(n-1)}^{n-1} \hat{r}'_t e^{2\pi i t f}$$

where

$$(5.3) \quad \hat{r}'_t = \frac{1}{n} \sum_{s=1}^{n-t} x_s \overline{x_{s+t}} \quad t=0, 1, \dots, n-1$$

$$\hat{r}'_{-t} = \overline{\hat{r}'_t} .$$

It is possible to calculate $p(f)$ using equation (5.1) and then calculate \hat{r}'_t using the inverse transform. To obtain the exact inverse transform it is necessary to calculate $p(f)$ at the frequencies

$$f=0, \pm \frac{1}{2n-1}, \pm \frac{2}{2n-1}, \dots, \pm \frac{n-1}{2n-1}$$

Cooley, Welsh and Lewis [3],

$$(5.4) \quad \hat{r}'_t = \frac{1}{2n-1} \sum_{\nu=-(n-1)}^{n-1} p\left(\frac{\nu}{2n-1}\right) e^{-2\pi i t(\nu/2n-1)}.$$

When using the fast Fourier transform it is necessary to annex $n-1$ zeros to the data in order to obtain the exact inverse transform. In practice it is often not necessary to add this many zeros to the data, but as was pointed out by Cooley, Welsh and Lewis [3], at least m zeros should be added, where m is the largest lag used in the estimation of the spectrum. This avoids a circular definition of the estimated covariances.

With missing observations it is also necessary to calculate the Fourier transform of I_t in order to calculate c_t , unless this calculation has been carried out in the time domain. For a real time series, x_t and I_t can be transformed simultaneously by forming the complex series $y_t = x_t + iI_t$ and using the complex Fourier transform

$$(5.5) \quad \alpha_\nu = \sum_{t=1}^n y_t e^{2\pi i t \nu / n}, \quad \nu = 0, 1, \dots, n-1$$

where n denotes the total length of the series with the zeros included. Now

$$(5.6) \quad p\left(\frac{\nu}{n}\right) = \frac{1}{4n} \{[\operatorname{Re}(\alpha_\nu) + \operatorname{Re}(\alpha_{n-\nu})]^2 + [\operatorname{Im}(\alpha_\nu) - \operatorname{Im}(\alpha_{n-\nu})]^2\}$$

$$\nu = 0, 1, \dots, n-1,$$

and

$$(5.7) \quad P_c\left(\frac{\nu}{n}\right) = \frac{1}{4n} \{[\operatorname{Im}(\alpha_\nu) + \operatorname{Im}(\alpha_{n-\nu})]^2 + [\operatorname{Re}(\alpha_{n-\nu}) - \operatorname{Re}(\alpha_\nu)]^2\}$$

is the periodogram of the indicator series. For real data,

$$(5.8) \quad p\left(\frac{\nu}{n}\right) = p\left(\frac{n-\nu}{n}\right).$$

Let

$$(5.9) \quad L_t = \sum_{\nu=0}^{n-1} p\left(\frac{\nu}{n}\right) e^{-2\pi i t \nu / n}.$$

Also

$$(5.10) \quad c_t = \sum_{\nu=0}^{n-1} P_c\left(\frac{\nu}{n}\right) e^{-2\pi i t \nu / n}.$$

Then the estimated spectrum is

$$(5.11) \quad \hat{s}(f) = \sum_{t=-m}^m \frac{L_t}{c_t} w_t e^{2\pi i t f}$$

where $L_{-t} = \bar{L}_t$ and $c_{-t} = c_t$.

When a data window is used both c_t and d_t from equations (4.1) and (4.2) can be calculated via the direct Fourier transform by forming the complex series $b_t + i b_t^2$ and proceeding as above.

6. Simulations

First order autoregressions were generated on a computer,

$$(6.1) \quad x_t = a x_{t-1} + \varepsilon_t,$$

where $|a| < 1$ and ε_t are uncorrelated normal variables with mean zero and constant variance σ^2 . It is well known that the spectral density of this process is

$$(6.2) \quad s(f) = \sigma^2 (1 + a^2 - 2a \cos 2\pi f)^{-1}, \quad |f| \leq 1/2.$$

For each simulation, 1000 consecutive time points were generated, but were randomly observed with probability 1/2. A maximum lag was chosen, and a spectral window with weights proportional to binomial coefficients was used. These weights produce lag windows and spectral windows very Gaussian in shape, an aim of many spectral windows since they fall off fast in both the time domain and frequency domain. The lag window is generated as follows

$$(6.3) \quad \begin{aligned} W_0 &= 1 \\ W_t &= \frac{L+1-t}{L+t} W_{t-1}, \quad t=1, 2, \dots \end{aligned}$$

where L is an integer which remains to be determined. For L large, w_t becomes very small with increasing t , so small that computer underflow may be encountered. Therefore, the procedure is truncated when w_t becomes less than some small number, say $w_t < \varepsilon$. Let this truncation point be $t=m$, as before. Using the normal approximations to the binomial coefficients,

$$(6.4) \quad W_t \simeq e^{-t^2/L},$$

it follows that, for a given ε ,

$$(6.5) \quad L = [m^2 / (-\ln \varepsilon)],$$

where the brackets denote "the integer part." For $\varepsilon = 0.001$, the value used for these simulations,

$$(6.6) \quad L = [m^2/6.9], \quad m \geq 7.$$

The spectral window is given, to a good approximation, by

$$(6.7) \quad W(f) \approx m\sqrt{\pi/6.9} e^{(\pi m f)^2/6.9}.$$

Figure 1 shows the time and estimated spectrum for a process with autoregressive coefficient, $a=0.25$, and maximum lag, $m=50$. 513 time points out of 1000 were actually observed. The equivalent degrees of freedom for this estimate was calculated by the program, using equation (3.21), as $edf=22.5$. The spectral window and 99% confidence interval for a single point are also shown.

A problem which can occur when estimating a spectral density with missing observations, is negative estimates, particularly in frequency regions where the spectral density is small. This is because the estimate is not positive definite, and shows that the chi-square approximation to the distribution of spectrum estimates is not always a good approximation. In Figure 2, the autoregressive coefficient is 0.75, and 488 observations out of 1000 were actually observed. The spectrum was

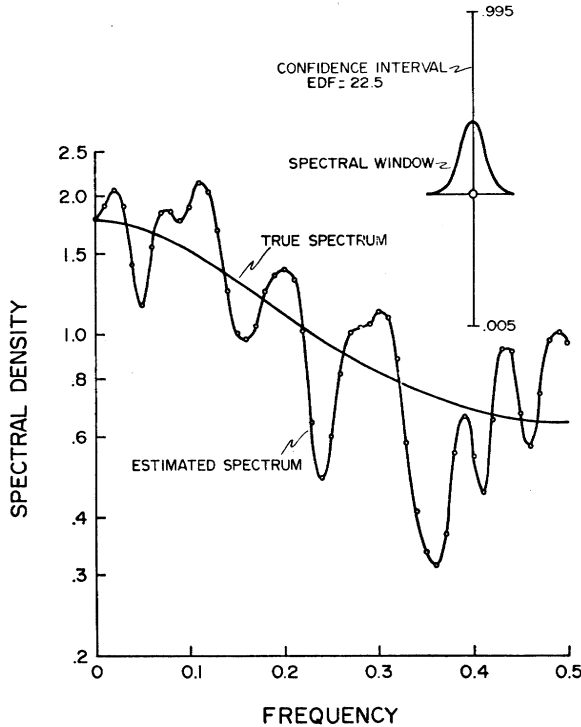


Fig. 1. Actual and estimated spectrum from 1000 observations with observations randomly unobserved with probability 1/2. 513 observations were actually observed.

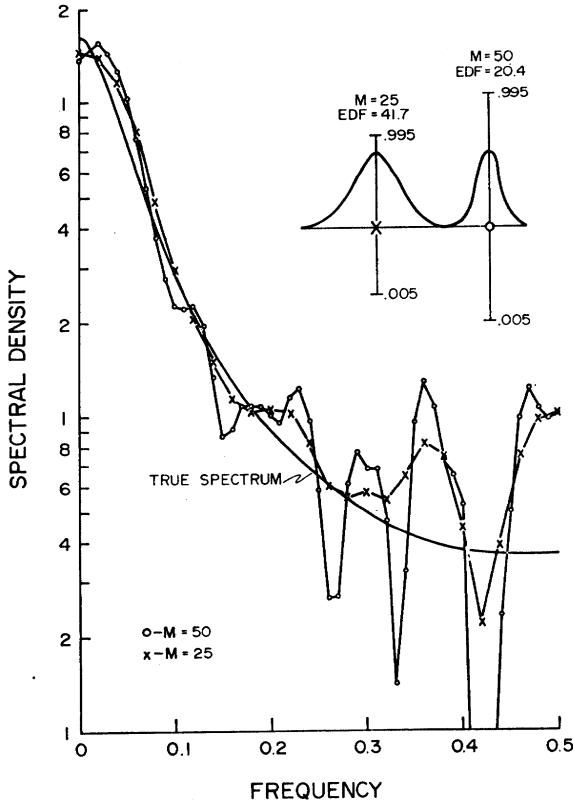


Fig. 2. Actual and estimated spectrum from 1000 observations with observations randomly unobserved with probability $1/2$. 488 observations were actually observed. Two degrees of smoothing were used. The less smoothed estimate is negative at frequency 0.42 cycles per unit time.

estimated twice, using two different degrees of smoothing, $m=25$ and $m=50$. The estimate was negative at the frequency 0.42 cycles per unit time when $m=50$. However, when the width of the spectral window was increased ($m=25$), the estimate became positive.

6. Conclusion

The spectral density of a stationary time series can be estimated when there are missing observations by first forming unbiased estimates of the covariances of lag 0 to m . Assuming that there is at least one data pair separated by each lag needed, an estimate with the usual first order properties can be formed. This is done by dividing the sum of lagged products by the number of terms in the sum. However, if the number of terms in some of the covariance estimates is low compared

to the number of observations, the variance will be increased. Exact expression for the "white noise" variance and covariance for a finite sample of a Gaussian process with missing observations have been derived and generalized to the case when the time series is multiplied by a data window before the analysis. The estimation can be approached by first estimating the covariances in the time domain, or by the direct Fourier transformation of the data.

UNIVERSITY OF HAWAII

REFERENCES

- [1] Blackman, R. D. and Tukey, J. W. (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover, New York.
- [2] Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation*, **19**, 297-301.
- [3] Cooley, J. W., Lewis, P. A. W. and Welch, P. D. (1967). The fast Fourier transform algorithm and its applications, IBM Research Paper RC-1743, IBM Waston Research Center, Yorktown Heights, New York.
- [4] Jones, R. H. (1962). Spectral estimates and their distributions, Part II, *Skandinavisk Aktuarietidskrift*, 135-153.
- [5] Jones, R. H. (1962). Spectral analysis with regularly missed observations, *Ann. Math. Statist.*, **33**, 455-461.
- [6] Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation, *Sankhya*, Series A, **25**, 383-392.