SPECTRAL ANALYSIS WITH RANDOMLY MISSED OBSERVATIONS: THE BINOMIAL CASE

By Perry A. Scheinok[1]

Hahnemann Medical College

0. Summary. Estimation of the spectral density of a discrete stationary process is considered under the assumption that some of the observations are missing due to some binomially distributed random mechanism. The asymptotic variance of the estimate is derived for normally distributed random variables. As in the author's dissertation [7], the extension to processes which are stationary of fourth order is fairly standard. It is hoped that one will be able to extend these results to more complicated random mechanisms.

1. Introduction. When observing a stationary stochastic process at equally spaced intervals of time, it might happen that the device being used to observe the process will miss an observation because of some random failure. The extension to estimating the spectral density in this case utilizes an idea introduced by R. Jones [3], and generalized by Parzen [5].

Let us be given a sample of size N, $x_1, \cdots, x_N$ from a Gaussian stationary process with mean zero, and absolutely continuous spectral distribution function $F(\lambda)$. Let $f(\lambda) = F'(\lambda)$ represent the spectral density. Then we have by a classic theorem of Herglotz [2] that

$$r_j = \int_{-\pi}^{\pi} f(u)e^{iju}\,du. \tag{1.1}$$

Here $r_j = E\{x_t x_{t+j}\}$. As is well-known, (Grenander and Rosenblatt [1]), one usually estimates the spectrum by forming the empirical auto-covariances

$$r_j^* = (N - |j|)^{-1}\sum_{i=1}^{N-|j|} x_i x_{i+j}, \quad j = 0, \pm 1, \pm 2, \cdots, \pm(N-1) \tag{1.2}$$

and then one forms the weighted sums

$$f_N^*(\lambda) = (2\pi)^{-1}\sum_{j=-(N-1)}^{N-1} r_j^* w_j^{(N)}(0)e^{-ij\lambda} \tag{1.3}$$

The function $W_N(x - \lambda) = (2\pi)^{-1}\sum_{k=1}^{N} w_k^{(N)}(\lambda)e^{ikx}$ is called the spectral window, and is chosen in such a way as to cause the estimate (1.3) to be asymptotically unbiased and consistent, i.e., $\lim_{N\to\infty} Ef_N^*(u) = f(u)$, and

$$\lim_{n\to\infty} \operatorname{Var} f_N^*(u) = 0.$$

One way of accomplishing this is to choose a basic kernel function $W(x)$ having the following properties (a) $W(x) \geqq 0$ in $-\pi \leqq x \leqq \pi$, (b) $W(x) = 0$ for $|x| > \pi$, (c) $W(x)$ is continuous in $-\pi \leqq x \leqq \pi$, and (d) $\int_{-\pi}^{\pi} W(x)\,dx = 1$. For convenience we choose $W(x)$ to be symmetric about $x = 0$.

One can now obtain a sequence of weighting functions from $W(x)$ by defining

Received 20 April 1964; revised 10 November 1964.

[1] Work begun while author was at Radio Corporation of America.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

$W_N(x) = B_N^{-1}W(xB_N^{-1})$. Here $\{B_N\}$ denotes a sequence of positive real numbers which has the property that $B_N \to 0$, $NB_N \to \infty$ as $N \to \infty$, and might be considered as the band-width of the functions $W_N(x)$.

The way this sequence of weighting functions has been defined assures one that $W_N(x)$ accumulates mass in the neighborhood of $x = 0$ at a slower rate than the Fejér kernel. This will be very important in obtaining the results of Section 5.

2. Randomly missed observations. Let there be a random process attached to the device sampling the time series which causes it to miss observations at times. As in [3], assume that all the samples are present, and attach a new set of weights in the composition of the estimate to make it asymptotically unbiased and consistent.

Let

$$a_j = 1 \quad \text{if } x_j \text{ is read}$$
$$= 0 \quad \text{if } x_j \text{ is } \textit{not} \text{ read.}$$

Let $p = \text{Prob}\{a_j = 1\}$ independent of j. Now for the estimate of the spectral density take

$$f_N^*(\lambda) = \int_{-\pi}^{\pi} W_N(\lambda - s)I_N'(s)\,ds \tag{2.1}$$

where $I_N'(s)$, the modified normalized periodogram is now defined as

$$I_N'(s) = (2\pi N)^{-1}\{\textstyle\sum_{j=1}^N (a_j^2/p)x_j^2 + \sum_{k=1,k\neq j}^N \sum_{j=1}^N (a_ja_k/p^2)x_jx_ke^{-i(j-k)s}\}. \tag{2.2}$$

Now we wish that $Ef_N^*(\lambda)$ be asymptotically unbiased. Because of the interchangeability of summation and expectation, we have

$$EI_N'(s) = (2\pi N)^{-1}\{\textstyle\sum_{j=1}^N [E(a_j^2x_j^2)/p] + \sum_{k=1,k\neq j}^N \sum_{j=1}^N [E(a_ja_kx_jx_k)/p^2]e^{-i(j-k)s}\}. \tag{2.3}$$

Note that the a_j's are statistically independent of the x_j's, since the a_j's are due to a property of the device, rather than the process being observed. We thus have

$$E(a_j^2x_j^2) = E(a_j^2)E(x_j^2) = pr_0 \tag{2.4}$$

and

$$E(a_ja_kx_jx_k) = E(a_ja_k)E(x_jx_k) = p^2r_{j-k}, \qquad j \neq k. \tag{2.5}$$

Because the x_n process is considered to be real, one has that $r_{-k} = r_k$. Hence,

$$Ef_N^*(\lambda) = (2\pi N)^{-1}\textstyle\sum_{j,k=1}^N \int_{-\pi}^{\pi} W_N(\lambda - s)r_{j-k}e^{-i(j-k)s}\,ds. \tag{2.6}$$

It was shown in [1] that this expression for the expectation has the property that $\lim_{N\to\infty} Ef_N^*(\lambda) = f(\lambda)$, which makes the estimate asymptotically unbiased. This author also proved in [7] that under suitable regularity conditions on the spectral density, as well as on the weight functions, the bias $|b_N(\lambda)| =$

$|Ef_N^*(\lambda) - f(\lambda)|$ is bounded by $K_N(\log N)/N$, where K_N is a function of f and the sequence W_N. For a more accessible source see [6].

3. The variance of the estimate. Under the assumption that the observed process is Gaussian, and the assumptions in Section 1, we shall try to compute the asymptotic variance of the estimate (2.1). We know that

$$\text{(3.1)} \quad \operatorname{Var} f_N^*(\lambda) = \int \int_{-\pi}^{\pi} W_N(\lambda - s) W_N(\lambda - t) \operatorname{Cov}\{I_N'(s), I_N'(t)\}\, ds\, dt.$$

Again, let us confine our attention to the covariance of the modified periodogram

$$\text{(3.2)} \quad \begin{aligned} \operatorname{Cov}\{I_N'(s), I_N'(t)\} &= (4\pi^2 N^2)^{-1}\{p^{-2} \textstyle\sum_{k=1}^N \sum_{j=1}^N \operatorname{Cov}\{a_j^2 x_j^2, a_k^2 x_k^2\} \\ &+ p^{-3} \textstyle\sum_{m=1}^N \sum_{k=1,k\neq j}^N \sum_{j=1}^N \operatorname{Cov}\{a_m^2 x_m^2, a_j a_k x_j x_k\}\{e^{-i(j-k)t} + e^{-i(j-k)s}\} \\ &+ p^{-4} \textstyle\sum_{n=1,n\neq m}^N \sum_{m=1}^N \sum_{k=1,k\neq j}^N \sum_{j=1}^N \operatorname{Cov}\{a_j a_k x_j x_k, a_m a_n x_m x_n\} e^{-i[(j-k)s+(m-n)t]}\}. \end{aligned}$$

Now use the fact that for Gaussian processes

$$\text{(3.3)} \quad E(x_i x_j x_k x_m) = r_{i-k} r_{j-m} + r_{i-m} r_{j-k} + r_{i-j} r_{k-m}.$$

This implies that

$$\text{(3.4)} \quad \begin{aligned} &\textstyle\sum_{k=1}^N \sum_{j=1}^N \operatorname{Cov}\{a_j^2 x_j^2, a_k^2 x_k^2\} \\ &\qquad = 3N r_0^2 p(1-p) + 2p^2 \textstyle\sum_{k=-N}^N (N - |k|) r_k^2. \end{aligned}$$

Now consider the second term in (3.2). This term can be broken up into two sums. One sum has j or k equal to m, and the other sum has none of the indices equal to one another. Again, using (3.3) one gets that

$$\text{(3.5)} \quad \begin{aligned} &p^{-3} \textstyle\sum_{m=1}^N \sum_{k=1,(k\neq j)}^N \sum_{j=1}^N \operatorname{Cov}\{a_m^2 x_m^2, a_j a_k x_j x_k\}\{e^{-i(j-k)t} + e^{-i(j-k)s}\} \\ &\quad = 2(3p^{-1} - 1) \textstyle\sum_{k=1,(k\neq j)}^N \sum_{j=1}^N r_0 r_{k-j} [\cos(k-j)t + \cos(k-j)s] \\ &\qquad + 2 \textstyle\sum_{m=1}^N \sum_{k=1}^N \sum_{j=1,\{j\neq k\neq m\}}^N r_{m-j} r_{m-k}\{e^{-i(j-k)t} + e^{-i(j-k)s}\}. \end{aligned}$$

The last summation in (3.2) can be broken up into three different sums. One sum consists of those terms for which $j \neq k \neq m \neq n$. There are $N(N-1)\cdot(N-2)(N-3)$ terms in this sum. The next sum consisting of $2N(N-1)$ terms has $j = m$, $k = n$, or $j = n$, $k = m$. The third sum covers the situation when $j = m$, $k \neq n$; $j = n$, $k \neq m$; $k = m$, $j \neq n$; or $k = n$, $j \neq m$. There are $4N(N-1)(N-2)$ terms in this sum. Applying (3.3), one gets that

$$\text{(3.6)} \quad \begin{aligned} &p^{-4} \textstyle\sum_{n=1,(n\neq m)}^N \sum_{m=1}^N \sum_{k=1,(k\neq j)}^N \sum_{j=1}^N \operatorname{Cov}\{a_j a_k x_j x_k, a_m a_n x_m x_n\} e^{-i[(j-k)s+(m-n)t]} \\ &\quad = \textstyle\sum_{n=1}^N \sum_{m=1}^N \sum_{k=1}^N \sum_{j=1,\{j\neq k\neq m\neq n\}}^N \{r_{j-m} r_{k-n} + r_{j-n} r_{k-m}\} e^{-i[(j-k)s+(m-n)t]} \\ &\quad + p^{-2} \textstyle\sum_{k=1,(k\neq j)}^N \sum_{j=1}^N [r_0^2 + (2-p^2) r_{j-k}^2][e^{-i(j-k)(s+t)} + e^{-i(j-k)(s-t)}] \\ &\quad + 4p^{-1} \textstyle\sum_{m=1}^N \sum_{k=1}^N \sum_{j=1,\{j\neq k\neq m\}}^N \{r_0 r_{m-k} + (2-p) r_{k-j} r_{m-j}\} \\ &\qquad \cdot \cos(k-j)s \cos(m-j)t. \end{aligned}$$

This finishes the calculation of $4\pi^2N^2 \operatorname{Cov}\{I_N'(s),\ I_N'(t)\}$. Substitution in (3.1) will give us an expression for $4\pi^2N^2 \operatorname{Var}\{f_N^*(\lambda)\}$. Unfortunately, such an expression is of a form which conceals its salient features, and thus is not of much practical use. One is thus forced to resort to asymptotic considerations.

4. Asymptotic estimates of the variance. We shall now examine the terms calculated for $4\pi^2N^2 \operatorname{Cov}\{I_N'(s),\ I_N'(t)\}$ asymptotically. We shall be working in the frequency domain. Since the time series is assumed to be real, we have $f(x) = f(-x)$ where $f(x)$ enters into our calculations via (1.1).

For (3.4), one thus has that

$$\begin{aligned} 3Nr_0^2p(1-p) &+ 2p^2 \textstyle\sum_{k=-N}^{N} (N-|k|)r_k^2 \\ &= \textstyle\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x)f(y)\{3Np(1-p) \\ &\qquad + [2p^2 \sin^2 \tfrac{1}{2}N(x+y)/\sin^2 \tfrac{1}{2}(x+y)]\}\, dx\, dy. \end{aligned} \tag{4.1}$$

From the theory of the Fejér kernel, it is well-known that

$$\textstyle\int_{-\pi}^{\pi} f(x)[\sin^2 \tfrac{1}{2}N(x+y)/\sin^2 \tfrac{1}{2}(x+y)]\, dx \sim 2N\pi f(-y) = 2N\pi f(y). \tag{4.2}$$

Thus, asymptotically, after multiplying by $(4\pi^2N^2p^2)^{-1}$, (4.1) is of the order of

$$3(4\pi^2N)^{-1}(1-p^{-1})[\textstyle\int_{-\pi}^{\pi} f(x)\, dx]^2 + (\pi N)^{-1} \int_{-\pi}^{\pi} f^2(x)\, dx. \tag{4.3}$$

In order to deal with the terms (3.5) and (3.6), we first note that

$$\textstyle\sum_{n=1}^{N} e^{imt} = e^{it}[(e^{iNt}-1)/(e^{it}-1)] = e^{i(N+1)t/2}(\sin \tfrac{1}{2}Nt/\sin \tfrac{1}{2}t). \tag{4.4}$$

As in [6], this sum will be denoted by $D_N(t)$.

The actual calculations involving (3.5) and (3.6) are exceedingly lengthy and laborious. Instead of going through the gory details for each individual sum, we shall cover the details of one of the sums, and subsequently present the total asymptotic expressions for (3.5) and (3.6) in their entirety. Consider the last sum in (3.5). We have that by (1.1),

$$\begin{aligned} 2 \textstyle\sum_{m=1}^{N} \sum_{k=1}^{N} \sum_{j=1,\{j\neq k\neq m\}}^{N} r_{m-j}r_{m-k}\{e^{-i(j-k)t} + e^{-i(j-k)s}\} \\ = 2 \textstyle\int \int_{-\pi}^{\pi} f(x)f(y)\, dx\, dy \sum_{m=1}^{N} \sum_{k=1}^{N} \sum_{j=1,\{j\neq k\neq m\}}^{N} \\ \cdot\{e^{i[m(x+y)-j(x+t)-k(y-t)]} + e^{i[m(x+y)-j(x+s)-k(y-s)]}\}. \end{aligned} \tag{4.5}$$

Now let us devote our attention to the triple sum. We can rewrite the sum involving t as follows. (We get an analogous sum for the terms involving s.)

$$\begin{aligned} \textstyle\sum_{m=1}^{N} \sum_{k=1}^{N} \sum_{j=1,\{j\neq k\neq m\}}^{N} e^{i[m(x+y)-j(x+t)-k(y-t)]} \\ = \textstyle\sum_{m=1}^{N} \sum_{k=1}^{N} \sum_{j=1}^{N} e^{i[m(x+y)-j(x+t)-k(y-t)]} \\ - \textstyle\sum_{m=1}^{N} \sum_{j=1}^{N} e^{i(m-j)(x+y)} - \sum_{k=1}^{N} \sum_{m=1}^{N} e^{i(m-k)(y-t)} \\ - \textstyle\sum_{j=1}^{N} \sum_{k=1}^{N} e^{i(k-j)(x+t)} + 2N. \end{aligned} \tag{4.6}$$

Summing these terms and using (4.4), we get that (4.5) can be rewritten as

$$
\begin{aligned}
&2 \int \int_{-\pi}^{\pi} f(x)f(y)\{D_N(x+y)\overline{[D_N(x+t)D_N(y-t)}\\
&\qquad\qquad + \overline{D_N(x+s)D_N(y-s)}]\\
&- 2\,|D_N(x+y)|^2 - |D_N(y-t)|^2 - |D_N(x+t)|^2 - |D_N(x+s)|^2 + 4N\}\,dx\,dy.
\end{aligned}
\tag{4.7}
$$

Using (4.2) and the theory in [6], p. 172, one gets that asymptotically (4.7) is of the order of

$$
\begin{aligned}
8\pi^2 N[f^2(t) + f^2(s) - \pi^{-1} \textstyle\int_{-\pi}^{\pi} f(x)[f(x) + f(s) + f(t)]\,dx \\
+ \pi^{-2}\{\textstyle\int_{-\pi}^{\pi} f(x)\,dx\}^2].
\end{aligned}
\tag{4.8}
$$

A similar break-up and analysis is done for each term in (3.5) and (3.6). The result is that after division by $4\pi^2 N^2$ (3.5) is asymptotically of the order of

$$
\begin{aligned}
&(\pi N)^{-1}[3p^{-1} - 1] \textstyle\int_{-\pi}^{\pi} f(x)\,dx[f(t) + f(s) - \pi^{-1} \int_{-\pi}^{\pi} f(y)\,dy]\\
&\qquad +2N^{-1}[f^2(t) + f^2(s) - \pi^{-1} \textstyle\int_{-\pi}^{\pi} f(x)\,dx[f(x)\\
&\qquad\qquad + f(s) + f(t)] + \pi^{-2}\{\textstyle\int_{-\pi}^{\pi} f(x)\,dx\}^2]
\end{aligned}
\tag{4.9}
$$

and (3.6) is of the order of

$$
\begin{aligned}
&[f(s)f(t)]N^{-2}\{|D_N(s+t)|^2 + |D_N(s-t)|^2\} - 2N^{-1}\{f(t) + f(s)\}^2\\
&- (2\pi N^2)^{-1}\{|D_N(s+t)|^2 + |D_N(s-t)|^2\}[f(s) + f(t)] \textstyle\int_{-\pi}^{\pi} f(x)\,dx\\
&+ 4(\pi N)^{-1}[f(s) + f(t)] \textstyle\int_{-\pi}^{\pi} f(x)\,dx + (\pi N)^{-1} \int_{-\pi}^{\pi} f^2(x)\,dx\\
&+ (4\pi^2 N^2)^{-1}\{|D_N(s+t)|^2 + |D_N(s-t)|^2 - 12N\}[\textstyle\int_{-\pi}^{\pi} f(x)\,dx]^2\\
&+ (4\pi^2 N^2 p^2)^{-1}\{|D_N(s+t)|^2 + |D_N(s-t)|^2\\
&\quad - 2N(3 - p^2)\}[\textstyle\int_{-\pi}^{\pi} f(x)\,dx]^2\\
&+ (\pi N p^2)^{-1} \textstyle\int_{-\pi}^{\pi} f(x)[f(x+s+t) + f(x+s-t)]\,dx\\
&+ (p\pi N)^{-1}\{[f(s)N^{-1}[|D_N(s+t)|^2 + |D_N(s-t)|^2]\\
&\quad - 2(3-p)(f(s) + f(t))] \textstyle\int_{-\pi}^{\pi} f(x)\,dx\\
&+ 4\pi(2-p)f(s)f(t) - (2\pi N)^{-1}[\textstyle\int_{-\pi}^{\pi} f(x)\,dx]^2\\
&\quad \cdot[|D_N(s+t)|^2 + |D_N(s-t)|^2 - 2N(3-p)]\\
&- \tfrac{1}{2}(2-p) \textstyle\int_{-\pi}^{\pi} f(x)[f(x+s+t) + f(x-s+t)\\
&\quad + f(x+s-t) + f(x-s-t)]\,dx\}.
\end{aligned}
\tag{4.10}
$$

The asymptotic approximation of Cov $\{I_N'(s), I_N'(t)\}$ is thus embodied in the sum of the terms (4.3), (4.9) and (4.10). Our final approximation will involve the behavior of these expressions when acted upon by the weighting functions.

5. The second asymptotic approximation. In order to apply asymptotic approximations to (3.1), we note that under the assumptions of Section 2, we have the following auxiliary results. We use the techniques of [4] and [7] to obtain that

$$\begin{aligned} &\int\int_{-\pi}^{\pi} [|D_N(s+t)|^2 + |D_N(s-t)|^2] f(s)f(t)W_N(s-\lambda)W_N(t-\lambda)\, ds\, dt \\ (5.1)\qquad &\sim 2\pi N \int_{-\pi}^{\pi} f^2(t)[W_N^2(t-\lambda) + W_N(t-\lambda)W_N(-t-\lambda)]\, dt \\ &\sim 2\pi N f^2(\lambda) B_N^{-1} \int_{-\pi}^{\pi} W^2(u)\, du, \qquad \lambda \neq 0, \pi, \end{aligned}$$

and

$$\sim 4\pi N f^2(\lambda) B_N^{-1} \int_{-\pi}^{\pi} W^2(u)\, du, \qquad \lambda = 0, \pi.$$

In the light of the theory developed by Parzen [4], one is interested in considering $\lim_{N\to\infty} NB_N \operatorname{Var} f_N^*(\lambda)$. If one examines all the terms constituting $\operatorname{Var} f_N^*(\lambda)$, one notes that except for the terms involving $[|D_N(s+t)|^2 + |D_N(s-t)|^2]$, all other terms go to zero under the above limiting operation. Thus we are primarily concerned with the term

$$\begin{aligned} A(s,t) &= N^{-2}\{|D_N(s+t)|^2 + |D_N(s-t)|^2\} \\ (5.2)\qquad &\cdot [f(s)f(t) - (f(s)+f(t))(2\pi)^{-1} \int_{-\pi}^{\pi} f(x)\, dx \\ &+ f(s)(p\pi)^{-1} \int_{-\pi}^{\pi} f(x)\, dx + (4\pi^2)^{-1}(1-p^{-1})^2 \{\int_{-\pi}^{\pi} f(x)\, dx\}^2]. \end{aligned}$$

This term is now considered with the weighting functions as

$$(5.3)\qquad \lim_{N\to\infty} NB_N \int\int_{-\pi}^{\pi} A(s,t) W_N(s-\lambda) W_N(t-\lambda)\, ds\, dt.$$

Using (5.1), and the properties of the weighting functions listed in Section 1, one gets the final result that

$$\begin{aligned} \lim_{N\to\infty} NB_N \operatorname{Var} f_N^*(\lambda) &= 2\pi[f(\lambda) - (2\pi)^{-1}(1-p^{-1}) \\ (5.4)\qquad &\cdot \int_{-\pi}^{\pi} f(x)\, dx]^2 \int_{-\pi}^{\pi} W^2(u)\, du, \qquad \lambda \neq 0, \pm\pi, \\ &= 4\pi[f(\lambda) - (2\pi)^{-1}(1-p^{-1}) \\ &\cdot \int_{-\pi}^{\pi} f(x)\, dx]^2 \int_{-\pi}^{\pi} W^2(u)\, du, \qquad \lambda = 0, \pm\pi. \end{aligned}$$

It should be noted that the final result satisfies the essential criteria for such a situation, i.e.,

(a) the variance is positive;
(b) the variance is increased as p is decreased;
(c) as $p \to 1$ we obtain that

$$\begin{aligned} (5.5)\qquad \lim_{N\to\infty} NB_N \operatorname{Var} f_N^*(\lambda) &= 2\pi f^2(\lambda) \int_{-\pi}^{\pi} W^2(u)\, du, \qquad \lambda \neq 0, \pm\pi, \\ &= 4\pi f^2(\lambda) \int_{-\pi}^{\pi} W^2(u)\, du, \qquad \text{otherwise}, \end{aligned}$$

which is the classic result in spectral analysis of stationary time series.

6. Acknowledgment. I should hereby like to thank Professor E. Parzen for his reading of an earlier version, and for his invaluable suggestions.

REFERENCES

[1] GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[2] HERGLOTZ, G. (1911). Uber Potenzenreihen mit positivem reellen Teil im Einheitskreis. *Ber. Verh. Kgl. Sachs. Ges. Wiss. Leipzig. Math-Phys. Kl.* **63** 501.

[3] JONES, R. H. (1962). Spectral analysis with regularly missed observations. *Ann. Math. Statist.* **32** 455–461.

[4] PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28** 329–348.

[5] PARZEN, E. (1962). On spectral analysis with missing observations and amplitude modulation. Technical Report No. 46, Applied Mathematics and Statistical Laboratories, Stanford Univ.

[6] ROSENBLATT, M. (1962). *Random Processes*. Oxford Univ. Press.

[7] SCHEINOK, P. (1960). The error on using the asymptotic variance and bias of spectrograph estimates for finite observations. Unpublished thesis, Indiana University.