

## Periodogram and Other Direct Spectral Estimators

### 6.0 Introduction

Here we begin our study of nonparametric estimators of the spectral density function (SDF) for stationary processes with purely continuous spectra. The spectral properties for these processes are more readily seen from their SDFs than from their integrated spectra. Of course in this case the SDF and the integrated spectrum contain equivalent information, but the former is usually easier to interpret – just as probability density functions (rather than cumulative probability distribution functions) give a better visual indication of the distribution of probabilities for a random variable (RV). Our discussion will thus concentrate on estimation of the SDF.

We shall base our SDF estimates on the observed values of a time series. We consider a time series of sample size  $N$  as one realization of a contiguous portion  $X_0, X_1, \dots, X_{N-1}$  of a discrete parameter stationary process  $\{X_t\}$  with SDF  $S(\cdot)$  (see Section 4.5 for a discussion on sampling from a continuous parameter process). The motivation for nonparametric SDF estimators is the relationship between the SDF and the autocovariance sequence (ACVS)  $\{s_\tau\}$  given by Equation (113d), namely,

$$S(f) = \Delta_t \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f \tau \Delta_t}, \quad (163)$$

where, as before,  $\Delta_t$  is the sampling interval (the time that passes between when  $X_t$  and  $X_{t+1}$  are recorded). The key idea is to estimate  $\{s_\tau\}$  in some fashion and then apply Equation (163) to estimate the SDF. For a real-valued stationary process,

$$s_\tau = \text{cov}\{X_{t+\tau}, X_t\} = E\{(X_{t+\tau} - \mu)(X_t - \mu)\},$$

where  $\mu = E\{X_t\}$  is the mean of the process. In Section 6.1, we consider an obvious estimator of  $\mu$  (the sample mean), and then we look at two estimators of  $\{s_\tau\}$  in Section 6.2. Plugging one of these ACVS estimators into the right-hand side of Equation (163) leads to an SDF estimator known as the *periodogram*, which is the subject of Section 6.3. While easy to compute, the periodogram can have two statistical deficiencies. First, its expected value is not guaranteed to be in reasonable agreement with the true SDF (when this is true, the periodogram is said to be a *biased estimator* of  $S(\cdot)$ ). Second, it has a degree of variability that is unacceptable in many applications (statistically speaking, the periodogram is an *inconsistent*

estimator). Section 6.4 looks at one scheme (tapering) for alleviating bias in the periodogram, leading to *direct spectral estimators* of  $S(\cdot)$  (the periodogram is a special case of a direct spectral estimator). Section 6.5 considers a second bias-reduction scheme (prewhitening). The statistical properties of direct spectral estimators are the subject of Section 6.6. The remaining sections of this chapter give details about how to compute direct spectral estimates (Section 6.7), examples of direct spectral estimates for actual time series (6.8), comments on direct spectral estimators for complex-valued time series (6.9), a summary (6.10) and some exercises (6.11). (None of the estimators we consider in this chapter improves upon the inconsistency of the periodogram, but they form building blocks for nonparametric estimators that do address this statistical deficiency, as discussed in subsequent chapters.)

### 6.1 Estimation of the Mean

Let  $\{X_t\}$  be a discrete parameter (second-order) stationary process with mean  $\mu$  and ACVS  $\{s_\tau\}$ . We want to estimate  $\mu$  and  $\{s_\tau\}$  from a sample of  $N$  observations of our process, which we consider as observed values of the RVs  $X_0, X_1, \dots, X_{N-1}$ .

A natural estimator of  $\mu$  is just the sample mean of  $X_0, X_1, \dots, X_{N-1}$ , given by the RV

$$\bar{X} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{t=0}^{N-1} X_t.$$

Since

$$E\{\bar{X}\} = \frac{1}{N} \sum_{t=0}^{N-1} E\{X_t\} = \mu,$$

this estimator is unbiased. Moreover, under certain mild conditions (see section 6.1 of Fuller, 1996),  $\bar{X}$  is a consistent estimator of  $\mu$ , by which we mean that for every  $\delta > 0$ ,

$$\lim_{N \rightarrow \infty} \mathbf{P} [|\bar{X} - \mu| > \delta] = 0, \quad (164a)$$

where  $\mathbf{P}[\cdot]$  is the probability measure associated with the stationary process. Consistency follows directly from Chebyshev's inequality

$$\mathbf{P} [|\bar{X} - \mu| > \delta] \leq \frac{E\{(\bar{X} - \mu)^2\}}{\delta^2} = \frac{\text{var}\{\bar{X}\}}{\delta^2},$$

as soon as we can show that  $\text{var}\{\bar{X}\} \rightarrow 0$  as  $N \rightarrow \infty$ . For the important special case of an absolutely summable ACVS,

$$\text{var}\{\bar{X}\} \approx \frac{S(0)}{N \Delta_t} \quad (164b)$$

for large  $N$ , where  $S(0)$  is a finite number (if  $S(0) = 0$ , this approximation is not particularly useful). The above implies that the variance of the sample mean decreases to 0 as  $N$  goes to  $\infty$ . A proof of this fact is outlined in the first of the Comments and Extensions (C&Es) for this section. The effectiveness of the approximation for selected first-order moving average and autoregressive processes is explored in Exercises [6.1] and [9.2c].

### Comments and Extensions to Section 6.1

[1] We wish to show that approximation (164b) is valid for large  $N$  for a stationary process with an ACVS that is absolutely summable; i.e.,  $\sum_{\tau=-\infty}^{\infty} |s_{\tau}| < \infty$ .

▷ **Exercise [165]** Show that, for any stationary process (whether or not its ACVS is absolutely summable),

$$\text{var} \{\bar{X}\} = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_{\tau}. \quad (165a) \triangleleft$$

As  $N \rightarrow \infty$ , we can reduce this expression to a compact form by appealing to the Cesàro summability theorem stated in Equation (79). In the present context, this theorem says that, if  $\sum_{\tau=-(N-1)}^{N-1} s_{\tau}$  converges to a limit as  $N \rightarrow \infty$  (this must be true since the ACVS is assumed to be absolutely summable), then  $\sum_{\tau=-(N-1)}^{N-1} \left(1 - |\tau|/N\right) s_{\tau}$  converges to the same limit. We can thus conclude that

$$\lim_{N \rightarrow \infty} N \text{var} \{\bar{X}\} = \lim_{N \rightarrow \infty} \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_{\tau} = \lim_{N \rightarrow \infty} \sum_{\tau=-(N-1)}^{N-1} s_{\tau} = \sum_{\tau=-\infty}^{\infty} s_{\tau}. \quad (165b)$$

The assumption of absolute summability implies that  $\{X_t\}$  has a purely continuous spectrum with SDF given by

$$S(f) = \Delta_t \sum_{\tau=-\infty}^{\infty} s_{\tau} e^{-i2\pi f \tau \Delta_t} \text{ and hence } S(0) = \Delta_t \sum_{\tau=-\infty}^{\infty} s_{\tau}$$

(this result follows from theorems 15.10 and 15.2, Champeney, 1987). Equation (165b) now becomes

$$\lim_{N \rightarrow \infty} N \text{var} \{\bar{X}\} = \frac{S(0)}{\Delta_t}; \text{ i.e., } \text{var} \{\bar{X}\} \approx \frac{S(0)}{N \Delta_t} \quad (165c)$$

for large  $N$ , which completes the proof (theorem 18.2.1 of Ibragimov and Linnik, 1971, has a proof that assumes only the existence of an SDF continuous at  $f = 0$ ).

[2] There are stationary processes for which the sample mean is *not* a consistent estimator of the true mean of the process (see Exercise [4.9] for a simple example). Theorems that state conditions under which the sample mean converges in some sense to the true mean fall under the category of *ergodic theorems* (see section 16 of Yaglom, 1987a, for a nice discussion). We do not pay much attention to ergodicity in this book – our experience is that it is of limited use in practical applications (this point is expanded upon in section 1.1.6, Chilès and Delfiner, 2012).

[3] If we denote the covariance matrix of  $\mathbf{X} \stackrel{\text{def}}{=} [X_0, X_1, \dots, X_{N-1}]^T$  by  $\Sigma$  (see Section 2.4), and if we assume that  $\Sigma^{-1}$  exists, the best (in the sense of having minimum variance) linear unbiased estimator of  $\mu$  is given by

$$\hat{\mu} = \frac{\mathbf{O}^T \Sigma^{-1} \mathbf{X}}{\mathbf{O}^T \Sigma^{-1} \mathbf{O}}, \text{ where } \mathbf{O} \stackrel{\text{def}}{=} \underbrace{[1, 1, \dots, 1]^T}_{N \text{ of these}}.$$

Note that this estimator requires knowledge of the ACVS out to lag  $N - 1$ , which is rarely the case in situations where estimating the SDF is of interest. One way to quantify the gain in using  $\hat{\mu}$  rather than  $\bar{X}$  is via the ratio

$$e_N(\bar{X}, \hat{\mu}) \stackrel{\text{def}}{=} \frac{\text{var} \{\hat{\mu}\}}{\text{var} \{\bar{X}\}},$$

which is called the *relative efficiency* of  $\bar{X}$  with respect to  $\hat{\mu}$ . Necessarily  $e_N(\bar{X}, \hat{\mu}) \leq 1$ , but typically  $e_N(\bar{X}, \hat{\mu}) < 1$  (with white noise being a notable exception). A related concept is the *asymptotic relative efficiency*:

$$e(\bar{X}, \hat{\mu}) \stackrel{\text{def}}{=} \lim_{N \rightarrow \infty} e_N(\bar{X}, \hat{\mu}).$$

It can be argued that, if the SDF  $S(\cdot)$  is piecewise continuous with no discontinuity at  $f = 0$  and  $0 < S(f) < \infty$  for all  $f$ , then  $e(\bar{X}, \hat{\mu}) = 1$  (see chapter 7 of Grenander and Rosenblatt, 1984, and theorem 6.1.3 of Fuller, 1996). Because  $e(\bar{X}, \hat{\mu})$  is unity, we say that  $\bar{X}$  is an *asymptotically efficient* estimator of  $\mu$ . Samarov and Taqu (1988) discuss some interesting cases in which  $\bar{X}$  is not asymptotically efficient; i.e.,  $e(\bar{X}, \hat{\mu}) < 1$ . They consider the class of *fractionally differenced processes*, for which the SDF is given by

$$S(f) = C |\sin(\pi f \Delta_t)|^\alpha, \text{ where } C > 0 \text{ and } \alpha > -1.$$

Note that, for small  $|f|$ , we have  $S(f) \propto |f|^\alpha$  approximately, so  $S(f) \rightarrow 0$  as  $f \rightarrow 0$  if  $\alpha > 0$  while  $S(f) \rightarrow \infty$  as  $f \rightarrow 0$  if  $-1 < \alpha < 0$ . Samarov and Taqu show that  $0.98 < e(\bar{X}, \hat{\mu}) < 1$  when  $-1 < \alpha < 0$ ;  $e(\bar{X}, \hat{\mu})$  decreases from 1 to 0 as  $\alpha$  increases from 0 to 1; and  $e(\bar{X}, \hat{\mu}) = 0$  for all  $\alpha \geq 1$ . Hence the sample mean can be a poor estimator of  $\mu$  when  $S(\cdot)$  decreases to 0 rapidly as  $f \rightarrow 0$ .

## 6.2 Estimation of the Autocovariance Sequence

Because

$$s_\tau \stackrel{\text{def}}{=} E\{(X_{t+\tau} - \mu)(X_t - \mu)\},$$

a natural estimator for the ACVS is

$$\hat{s}_\tau^{(U)} \stackrel{\text{def}}{=} \frac{1}{N - |\tau|} \sum_{t=0}^{N-|\tau|-1} (X_{t+|\tau|} - \bar{X})(X_t - \bar{X}), \quad (166a)$$

$\tau = -(N-1), \dots, -1, 0, 1, \dots, N-1$ ; i.e.,  $|\tau| < N$ . Note that  $\hat{s}_{-\tau}^{(U)} = \hat{s}_\tau^{(U)}$ , so this estimator mimics the symmetry property of  $\{s_\tau\}$ . For a sample of  $N$  observations there are no observations more than  $N-1$  lag units apart, so we cannot estimate  $s_\tau$  for  $|\tau| \geq N$  by an estimator with a form dictated by the right-hand side of Equation (166a). If we momentarily replace  $\bar{X}$  by  $\mu$  in Equation (166a), we see that

$$E\{\hat{s}_\tau^{(U)}\} = \frac{1}{N - |\tau|} \sum_{t=0}^{N-|\tau|-1} E\{(X_{t+|\tau|} - \mu)(X_t - \mu)\} = \frac{1}{N - |\tau|} \sum_{t=0}^{N-|\tau|-1} s_\tau = s_\tau$$

for all  $|\tau| < N$ . Thus  $\hat{s}_\tau^{(U)}$  is an unbiased estimator of  $s_\tau$  when  $\mu$  is known. If  $\mu$  is estimated by  $\bar{X}$ , it is a straightforward (but messy) exercise to derive  $E\{\hat{s}_\tau^{(U)}\}$  exactly (Anderson, 1971, pp. 448–9). Such calculations can be used to argue that, when the process mean must be estimated from the sample mean,  $\hat{s}_\tau^{(U)}$  is typically a biased estimator of  $s_\tau$  (see Exercise [6.2]).

There is a second estimator of  $s_\tau$  that is usually preferred to  $\hat{s}_\tau^{(U)}$ , namely,

$$\hat{s}_\tau^{(P)} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} (X_{t+|\tau|} - \bar{X})(X_t - \bar{X}) \quad (166b)$$

(the rationale for the superscript on  $\hat{s}_\tau^{(P)}$  is given in the next section). The only difference between the two estimators is the multiplicative factor in front of the summation, i.e.,  $1/(N - |\tau|)$  for  $\hat{s}_\tau^{(U)}$ , and  $1/N$  for  $\hat{s}_\tau^{(P)}$ . If we again momentarily replace  $\bar{X}$  by  $\mu$ , it follows that

$$E\{\hat{s}_\tau^{(P)}\} = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} s_\tau = \left(1 - \frac{|\tau|}{N}\right) s_\tau, \quad (166c)$$

so that  $\hat{s}_\tau^{(P)}$  is a biased estimator, and the magnitude of its bias increases as  $|\tau|$  increases. In the literature  $\hat{s}_\tau^{(U)}$  and  $\hat{s}_\tau^{(P)}$  are frequently called, respectively, the “unbiased” and “biased”

estimators of  $s_\tau$  (even though, when  $\bar{X}$  is used, both are typically biased, and, in fact, the magnitude of the bias in  $\hat{s}_\tau^{(U)}$  can be *greater* than that in  $\hat{s}_\tau^{(P)}$  – see Exercise [6.2] and the example shown in the left-hand column of Figure 168).

Why should we prefer the biased estimator  $\hat{s}_\tau^{(P)}$  to the unbiased estimator  $\hat{s}_\tau^{(U)}$ ? Here are some thoughts:

- [1] For many stationary processes of practical interest, the mean square error (MSE) of  $\hat{s}_\tau^{(P)}$  is smaller than that of  $\hat{s}_\tau^{(U)}$ ; i.e.,

$$\text{MSE} \{ \hat{s}_\tau^{(P)} \} \stackrel{\text{def}}{=} E \{ (\hat{s}_\tau^{(P)} - s_\tau)^2 \} < E \{ (\hat{s}_\tau^{(U)} - s_\tau)^2 \} = \text{MSE} \{ \hat{s}_\tau^{(U)} \} \quad (167a)$$

(Exercise [6.3] concerns a special case for which the above holds). Here is the basis for this statement. Since

$$\text{MSE} = \text{variance} + (\text{bias})^2,$$

where the bias for an estimator  $\hat{\theta}$  of the parameter  $\theta$  is defined as the quantity  $E\{\hat{\theta}\} - \theta$ , inequality (167a) evidently says that the variability in  $\hat{s}_\tau^{(U)}$  is more harmful than the bias in  $\hat{s}_\tau^{(P)}$ . This effect is particularly marked for  $|\tau|$  close to  $N$ , where the variability of  $\hat{s}_\tau^{(U)}$  is inherently large. Consider, for example,  $\hat{s}_{N-1}^{(U)}$  and  $\hat{s}_{N-1}^{(P)}$  – these two quantities are, respectively,

$$(X_{N-1} - \bar{X})(X_0 - \bar{X}) \quad \text{and} \quad \frac{1}{N}(X_{N-1} - \bar{X})(X_0 - \bar{X}), \quad (167b)$$

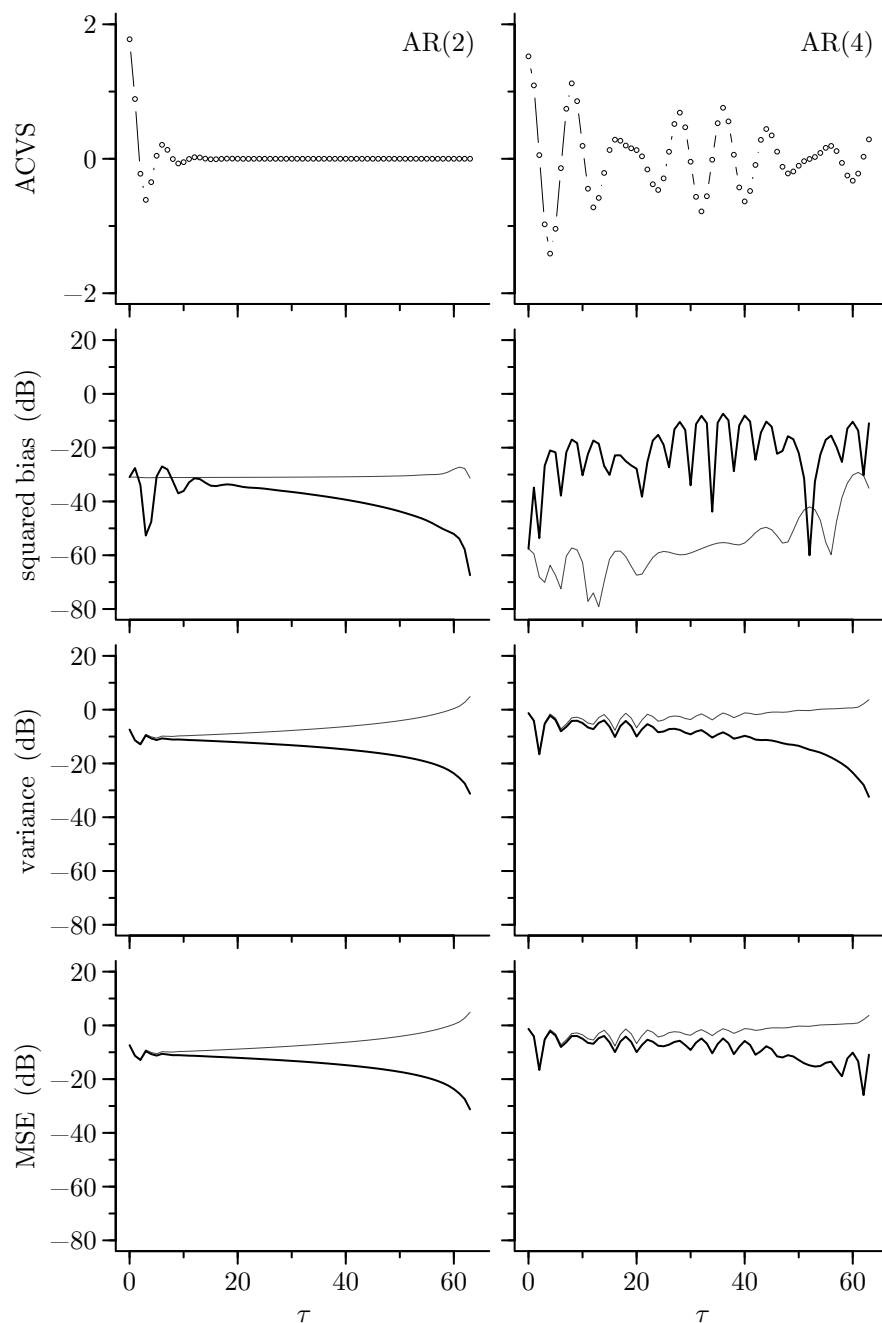
from which we can see that the variance of  $\hat{s}_{N-1}^{(U)}$  is  $N^2$  times the variance of  $\hat{s}_{N-1}^{(P)}$ . As concrete examples, Figure 168 shows the true ACVS for the second-order autoregressive (AR(2)) process defined by Equation (34) (upper left-hand plot) and for the AR(4) process defined by (35a) (upper right-hand plot) for lags 0 to 63 (see Exercise [9.6] for how to compute these ACVSs). Under the assumptions that these processes are Gaussian, that the sample mean is used to estimate the process mean and that the sample size  $N$  is 64, we can compute exactly the square bias, variance and MSE at various lags  $\tau$  for both  $\{\hat{s}_\tau^{(U)}\}$  and  $\{\hat{s}_\tau^{(P)}\}$ . These three quantities are shown beneath the ACVS plots for the two AR processes (light curves for  $\{\hat{s}_\tau^{(U)}\}$ , dark for  $\{\hat{s}_\tau^{(P)}\}$ ). Except for the biased estimator in the AR(4) case at large lags, the curves for the variance and the MSE are practically the same – this supports the assertion that the variance is the dominant contributor to the MSE. For both cases, the MSE of  $\hat{s}_\tau^{(P)}$  is strictly smaller than that of  $\hat{s}_\tau^{(U)}$  for all nonzero  $\tau$  (note also that, except at a few small lags, the square bias of  $\{\hat{s}_\tau^{(P)}\}$  is smaller than that of  $\{\hat{s}_\tau^{(U)}\}$  in the AR(2) case).

- [2] If  $\{X_t\}$  has a purely continuous spectrum, it follows from the Riemann–Lebesgue lemma (Titchmarsh, 1939, p. 403) that

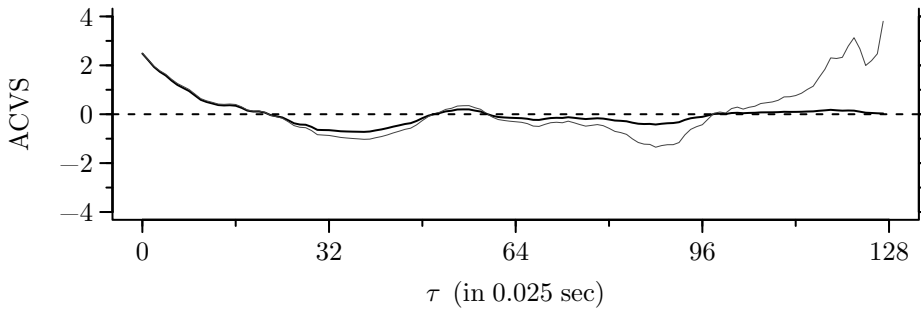
$$s_\tau \rightarrow 0 \quad \text{as} \quad |\tau| \rightarrow \infty$$

(the condition  $\sum_{\tau=-\infty}^{\infty} |s_\tau| < \infty$  also implies the above, as does  $\sum_{\tau=-\infty}^{\infty} s_\tau^2 < \infty$ ). It makes some sense – but not very much – to prefer an estimator of the sequence  $\{s_\tau\}$  that decreases to 0 nicely as  $|\tau|$  approaches  $N - 1$ . Consider, again, Equation (167b), which shows that  $\hat{s}_{N-1}^{(P)}$  is much closer to 0 than  $\hat{s}_{N-1}^{(U)}$  is.

- [3] We have noted that the ACVS of a stationary process must be positive semidefinite. Whereas the sequence  $\{\hat{s}_\tau^{(P)}\}$  is always positive semidefinite, the sequence  $\{\hat{s}_\tau^{(U)}\}$  need not be so (here we assume that both  $\hat{s}_\tau^{(P)}$  and  $\hat{s}_\tau^{(U)}$  are defined to be zero for all  $|\tau| \geq N$ ).



**Figure 168.** Comparison of square bias, variance and MSE of the unbiased and biased estimators of the ACVS for the AR(2) process of Equation (34) (left column) and the AR(4) process of Equation (35a) (right). The true ACVSs are shown (to lag 63) in the first row of plots. In the other three rows, the light and dark curves refer to, respectively, the unbiased and biased estimators. The sample size is 64; the process means are assumed unknown; both processes are Gaussian; and all curves are based on exact computations.



**Figure 169.** Comparison of unbiased and biased estimates of ACVS for wind speed time series shown in Figure 2(a). The light and dark curves show, respectively, the unbiased and biased estimates.

The following simple proof of the first claim is given in McLeod and Jiménez (1984, 1985). (The second claim is the subject of Exercise [6.6].) Let

$$d_k \stackrel{\text{def}}{=} x_k - \bar{x},$$

where the right-hand side is any observed value of  $X_k - \bar{X}$ . Let  $\{\epsilon_t\}$  be a zero mean white noise process with variance  $\sigma_\epsilon^2 = 1/N$ , and construct the following  $N$ th-order moving average process:

$$Y_t = \sum_{k=0}^{N-1} d_k \epsilon_{t-k}.$$

Then the *theoretical* (and hence necessarily positive semidefinite) ACVS of  $\{Y_t\}$ ,

$$s_{Y,\tau} = \sigma_\epsilon^2 \sum_{t=0}^{N-|\tau|-1} d_{t+|\tau|} d_t = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} (x_{t+|\tau|} - \bar{x})(x_t - \bar{x}),$$

equals the *estimated* ACVS of  $\{X_t\}$  using  $\{\hat{s}_\tau^{(P)}\}$  (cf. Equation (32b)). Because this result holds for all realizations, we can conclude that the estimated ACVS  $\{\hat{s}_\tau^{(P)}\}$  is positive semidefinite. One practical consequence of the possible failure of  $\{\hat{s}_\tau^{(U)}\}$  to be positive semidefinite is “curious behavior of the estimates of the spectrum” (Jenkins and Watts, 1968, p. 183); for example, use of  $\{\hat{s}_\tau^{(U)}\}$  can lead to estimates of the SDF – a nonnegative quantity, recall – that are negative for some frequencies.

Figure 169 illustrates the just-discussed points [2] and [3]. The light and dark curves show, respectively,  $\hat{s}_\tau^{(U)}$  and  $\hat{s}_\tau^{(P)}$  versus the lag  $\tau$ . Note, first, that  $\{\hat{s}_\tau^{(P)}\}$  damps to zero gracefully while  $\{\hat{s}_\tau^{(U)}\}$  does not and, second, that  $\hat{s}_\tau^{(U)} > \hat{s}_0^{(U)}$  for some large values of  $\tau$  – this cannot happen for a valid theoretical ACVS.

For these reasons, statisticians prefer  $\hat{s}_\tau^{(P)}$  over  $\hat{s}_\tau^{(U)}$  as the estimator of  $s_\tau$ . Hereinafter we shall use it as our standard “raw” estimator of the ACVS.

### Comments and Extensions to Section 6.2

[1] A comparison of Equations (166b) and (101h) shows that  $\hat{s}_\tau^{(P)}$  might correspond to an autocorrelation of the finite sequence  $X_0 - \bar{X}, X_1 - \bar{X}, \dots, X_{N-1} - \bar{X}$ , but with modifications needed for the beginning and end of the sequence. To within a multiplicative constant, this correspondence can be made exact by considering the autocorrelation of the zero padded sequence

$$X_0 - \bar{X}, X_1 - \bar{X}, \dots, X_{N-1} - \bar{X}, \underbrace{0, \dots, 0}_{N-1 \text{ zeros}}.$$

The fact that the discrete Fourier transform (DFT) of an autocorrelation is the square magnitude of the DFT of the sequence involved in the autocorrelation can be exploited to good use: we can compute  $\hat{s}_\tau^{(P)}$  for  $\tau = 0, \dots, N-1$  using substantially fewer numerical operations than would be needed if we were to use (166b) directly. The details are the subject of Exercise [6.7d].

[2] We have shown that the sequence  $\{\hat{s}_\tau^{(P)}\}$  is necessarily positive semidefinite. Since the  $\hat{s}_\tau^{(P)}$  terms are a sequence of RVs, this statement means that each realization of these RVs (a sequence of numbers) is positive semidefinite. Newton (1988, p. 165) gives a stronger result: a realization of the sequence  $\{\hat{s}_\tau^{(P)}\}$  as defined by Equation (166b) is positive definite if and only if the corresponding realizations of  $X_0, \dots, X_{N-1}$  are not all identical (see also McLeod and Jiménez, 1985).

### 6.3 A Naive Spectral Estimator – the Periodogram

Suppose the discrete parameter real-valued stationary process  $\{X_t\}$  with zero mean has a purely continuous spectrum with SDF  $S(\cdot)$ . For the remainder of this chapter, we assume that the relationship

$$S(f) = \Delta_t \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f \tau \Delta_t} \quad (170a)$$

holds and that  $S(\cdot)$  is in fact continuous for all  $f$ . If  $\sum |s_\tau| < \infty$ , theorem 15.10 of Champeney (1987) says that  $S(\cdot)$  must be continuous everywhere, so the assumption of continuity is restrictive but not overly so.

Given a time series that can be regarded as a realization of  $X_0, X_1, \dots, X_{N-1}$ , our problem is to estimate  $S(\cdot)$ . Now it is possible to estimate  $s_\tau$  for  $|\tau| = 0, 1, \dots, N-1$ , but not for  $|\tau| \geq N$ , by

$$\hat{s}_\tau^{(P)} = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} X_{t+|\tau|} X_t \quad (170b)$$

(this uses the assumption that the process mean is known to be 0). It thus seems natural to replace  $s_\tau$  in Equation (170a) by  $\hat{s}_\tau^{(P)}$  for  $|\tau| \leq N-1$  and to truncate the summation over  $\tau$  at the points  $\pm(N-1)$  – this amounts to *defining*  $\hat{s}_\tau^{(P)} = 0$  for  $|\tau| \geq N$ . These substitutions lead to an estimator known as the *periodogram* (even though it is a function of frequency and not period):

$$\hat{S}^{(P)}(f) \stackrel{\text{def}}{=} \Delta_t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(P)} e^{-i2\pi f \tau \Delta_t}. \quad (170c)$$

Note that the periodogram mimics an SDF in that it is an even function and is periodic with a period of  $2f_N$ , where  $f_N \stackrel{\text{def}}{=} 1/(2\Delta_t)$  is the Nyquist frequency (i.e.,  $\hat{S}^{(P)}(-f) = \hat{S}^{(P)}(f)$  and  $\hat{S}^{(P)}(f + 2nf_N) = \hat{S}^{(P)}(f)$  for all  $f$  and all  $n \in \mathbb{Z}$ ).

The following exercise points out a second way we can obtain the periodogram.

▷ **Exercise [170]** Show that

$$\hat{S}^{(P)}(f) = \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi f t \Delta_t} \right|^2. \quad (170d) \triangleleft$$

This equation makes it clear that, as is true for theoretical SDFs, the periodogram is non-negative. If we compare the summation above to Equation (74a), we can regard it as being proportional to the Fourier transform of the infinite sequence  $\dots, 0, X_0, X_1, \dots, X_{N-1}, 0, \dots$ , i.e., one for which  $X_t$  is defined to be zero for  $t < 0$  and  $t \geq N$ . Thus the periodogram can be taken to be either the Fourier transform of the biased estimator of the ACVS (with  $\hat{s}_\tau^{(P)}$



defined to be zero for  $|\tau| \geq N$ ) or, to within a multiplicative factor, the square modulus of the Fourier transform of  $X_0, \dots, X_{N-1}$  padded before and after with an infinite number of zeros.

Equation (170d) can be linked to the DFT of the finite sequence  $X_0, X_1, \dots, X_{N-1}$ , thus providing a convenient means of computing the periodogram. To see this connection, note that this DFT is given by

$$\mathcal{X}_k = \Delta_t \sum_{t=0}^{N-1} X_t e^{-i2\pi kt/N}, \quad k = 0, 1, \dots, N-1 \quad (171a)$$

(see Equation (91b)), noting that  $\mathcal{X}_k$  is associated with Fourier frequency  $f_k \stackrel{\text{def}}{=} k/(N \Delta_t)$ . Evaluation of (170d) at  $f = f_k$  yields

$$\hat{S}^{(P)}(f_k) = \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi f_k t \Delta_t} \right|^2 = \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi kt/N} \right|^2 = \frac{|\mathcal{X}_k|^2}{N \Delta_t}, \quad (171b)$$

which gives us the periodogram over the grid of Fourier frequencies. Because we are typically interested in plotting the periodogram over the frequency interval  $[0, f_N]$ , we need only consider  $f_k$  such that  $0 \leq f_k \leq f_N$ . These are the Fourier frequencies indexed by  $k = 0, 1, \dots, \lfloor N/2 \rfloor$ , where  $\lfloor N/2 \rfloor$  is the greatest integer less than or equal to  $N/2$ . (C&E [1] discusses how to compute the periodogram over a grid of frequencies finer than what is given by the Fourier frequencies.)

Just as the ACVS and the SDF form a Fourier transform pair, so do  $\{\hat{s}_\tau^{(P)}\}$  and  $\hat{S}^{(P)}(\cdot)$ :

$$\{\hat{s}_\tau^{(P)}\} \longleftrightarrow \hat{S}^{(P)}(\cdot)$$

(the above is the rationale for the superscript on  $\hat{s}_\tau^{(P)}$ ). The inverse Fourier transform corresponding to Equation (170c) says that

$$\hat{s}_\tau^{(P)} = \int_{-f_N}^{f_N} \hat{S}^{(P)}(f) e^{i2\pi f \tau \Delta_t} df, \quad \tau \in \mathbb{Z} \quad (171c)$$

(see Equation (75a), recalling that  $f_N = 1/(2 \Delta_t)$ ). In particular we have

$$\hat{s}_0^{(P)} = \int_{-f_N}^{f_N} \hat{S}^{(P)}(f) df, \quad \text{which mimics the relationship } s_0 = \int_{-f_N}^{f_N} S(f) df. \quad (171d)$$

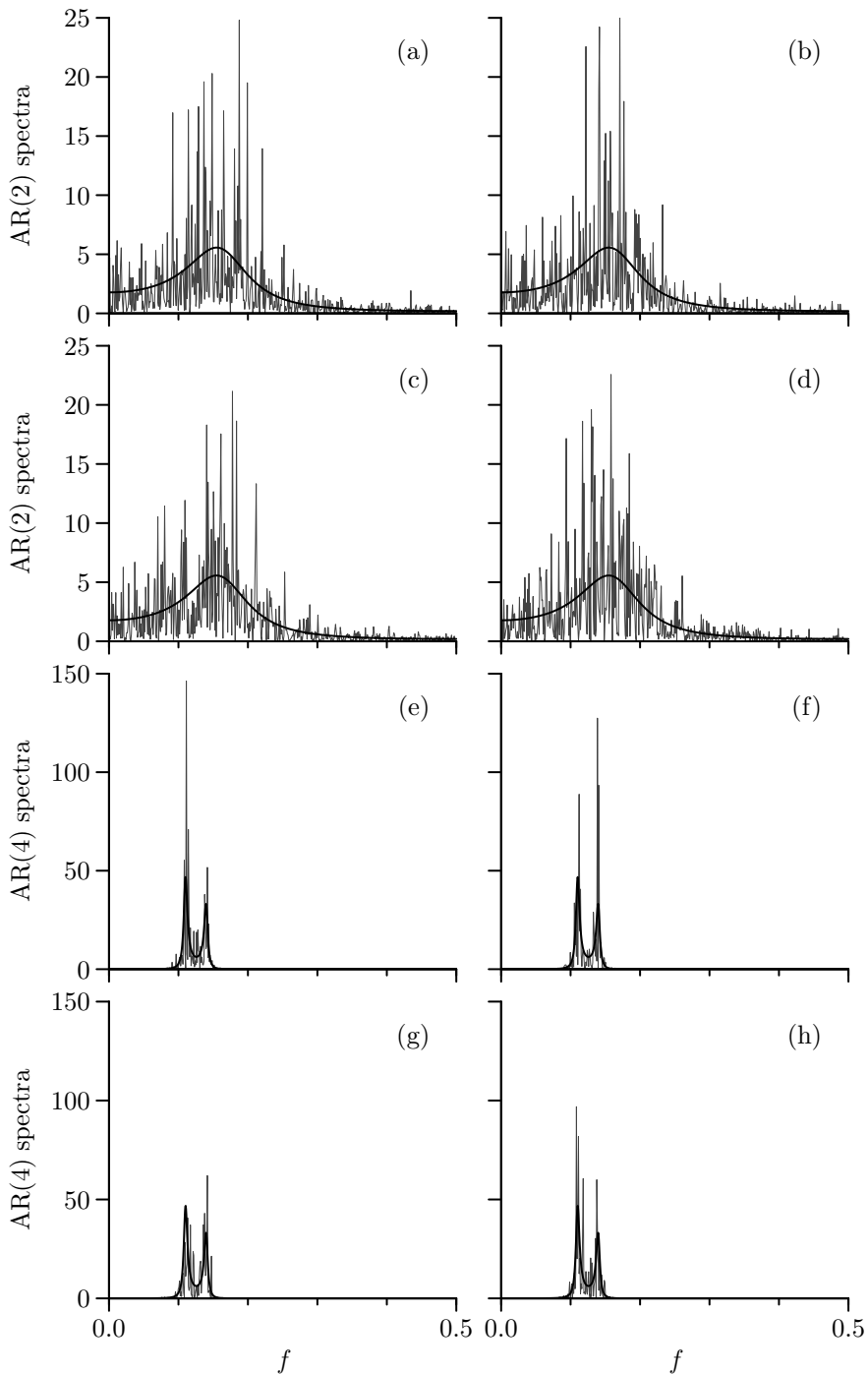
Note that the integral in Equation (171c) must evaluate to zero for all  $|\tau| \geq N$  since  $\hat{s}_\tau^{(P)}$  is by definition zero at those indices.

The relationship  $\{\hat{s}_\tau^{(P)}\} \longleftrightarrow \hat{S}^{(P)}(\cdot)$  assumes that  $\{\hat{s}_\tau^{(P)}\}$  is an infinite sequence with  $\hat{s}_\tau^{(P)} \stackrel{\text{def}}{=} 0$  for  $|\tau| \geq N$ . If we consider the finite sequence  $\{\hat{s}_\tau^{(P)} : \tau = -(N-1), \dots, N-1\}$  of  $2N-1$  variables, we can argue from Equations (100g) and (101b) that

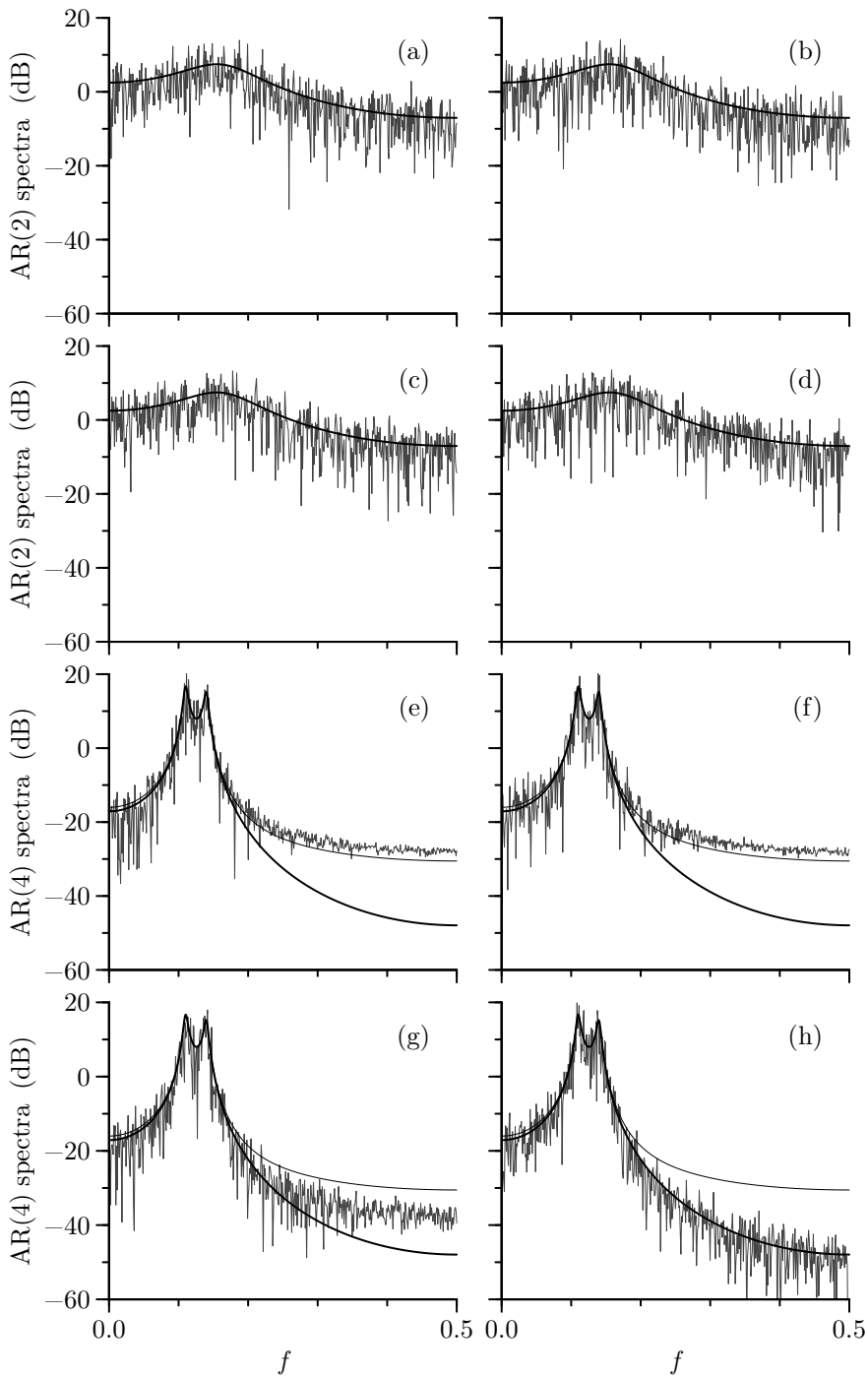
$$\{\hat{s}_\tau^{(P)} : \tau = -(N-1), \dots, N-1\} \longleftrightarrow \{\hat{S}^{(P)}(f'_k) : k = -(N-1), \dots, N-1\} \quad (171e)$$

with  $f'_k = k/[(2N-1) \Delta_t]$ . Hence, in addition to (171c), which makes use of continuous frequencies, the inverse DFT of Equation (101a) says that

$$\hat{s}_\tau^{(P)} = \frac{1}{(2N-1) \Delta_t} \sum_{k=-(N-1)}^{N-1} \hat{S}^{(P)}(f'_k) e^{i2\pi f'_k \tau \Delta_t}, \quad \tau = -(N-1), \dots, N-1, \quad (171f)$$



**Figure 172.** Periodograms (rough-looking light curves) for eight autoregressive time series of length  $N = 1024$  shown in Figure 34, along with true SDFs (smooth thick curves). The periodograms are computed over the grid of Fourier frequencies  $f_k = k/N$  using Equation (171b) with  $\Delta_t = 1$  and are plotted on a linear/linear scale.



**Figure 173.** As in Figure 172, but now plotted on a decibel/linear scale. The additional smooth thin curves in plots (e) to (h) show the expected value of the periodogram.

which uses discrete frequencies. Note that the right-hand side of Equation (171f) can be regarded as a Riemann sum approximation to the integral in Equation (171c) based upon an evenly spaced grid of frequencies with a spacing of  $1/[(2N - 1)\Delta_t]$ ; however, in this case the “approximation” is in fact exact!

Eight examples of periodograms are shown in Figure 172, one for each of the eight AR time series of length  $N = 1024$  shown in Figure 34. The periodograms (choppy-looking light curves) are evaluated over the grid of Fourier frequencies (as per Equation (171b)) and are compared with the true SDFs for the AR processes (smooth dark curves, as per Equations (145a), (33), (34) and (35a)). Figure 173 shows the same periodograms and true SDFs, but now on a decibel scale; i.e., we plot  $10 \log_{10}(\hat{S}^{(P)}(f))$  rather than  $\hat{S}^{(P)}(f)$  versus  $f$ . We will use these examples to illustrate certain undesirable properties of the periodogram, which led Broersen (2002) to refer to it as a quick and dirty estimator of the SDF, and caused Tukey (1980) to comment that “more lives have been lost looking at the raw periodogram than by any other action involving time series!”

Let us now investigate the sampling properties of  $\hat{S}^{(P)}(\cdot)$ . We will be chiefly concerned at first with

$$E\{\hat{S}^{(P)}(f)\}, \quad \text{var}\{\hat{S}^{(P)}(f)\} \quad \text{and} \quad \text{cov}\{\hat{S}^{(P)}(f'), \hat{S}^{(P)}(f)\}, \quad f' \neq f.$$

If  $\hat{S}^{(P)}(\cdot)$  were an ideal estimator of  $S(\cdot)$ , we would have

- [1]  $E\{\hat{S}^{(P)}(f)\} \approx S(f)$  for all  $f$ ,
- [2]  $\text{var}\{\hat{S}^{(P)}(f)\} \rightarrow 0$  as  $N \rightarrow \infty$  and
- [3]  $\text{cov}\{\hat{S}^{(P)}(f'), \hat{S}^{(P)}(f)\} \approx 0$  for  $f' \neq f$ .

We shall find, however, that, while [1] is a good approximation for some processes and some values of  $f$ , it can be very poor in other cases; [2] is blatantly false when  $S(f) > 0$ ; and [3] holds if  $f'$  and  $f$  are certain distinct frequencies, namely, the Fourier frequencies  $f_k$ , and if the bias in the periodogram is negligible.

We start by examining  $E\{\hat{S}^{(P)}(f)\}$  and determining how close it is to  $S(f)$ , the quantity it is supposed to estimate. From Equations (170d) and (166c) it follows that

$$E\{\hat{S}^{(P)}(f)\} = \Delta_t \sum_{\tau=-(N-1)}^{N-1} E\{\hat{s}_\tau^{(P)}\} e^{-i2\pi f \tau \Delta_t} = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_\tau e^{-i2\pi f \tau \Delta_t}. \quad (174a)$$

This is a convenient formula for computing  $E\{\hat{S}^{(P)}(f)\}$  for various stationary processes with known ACVS, but the following expression gives more insight.

▷ **Exercise [174]** Show that

$$E\{\hat{S}^{(P)}(f)\} = \int_{-f_N}^{f_N} \mathcal{F}(f - f') S(f') \, df', \quad (174b)$$

where

$$\mathcal{F}(f) \stackrel{\text{def}}{=} \Delta_t N \mathcal{D}_N^2(f \Delta_t); \quad (174c)$$

here  $\mathcal{D}_N(\cdot)$  is Dirichlet's kernel, which is defined in Equation (17c) and is discussed in Section 3.8. ◁

The function  $\mathcal{F}(\cdot)$  is called *Fejér's kernel*. We note that, for  $0 < |f| \leq f_N$ ,

$$\mathcal{F}(f) = \frac{\Delta_t \sin^2(N\pi f \Delta_t)}{N \sin^2(\pi f \Delta_t)}, \quad (174d)$$

which follows directly from Equations (174c) and (17c). The true SDF is thus convolved with Fejér's kernel to give  $E\{\hat{S}^{(P)}(\cdot)\}$ . Because it dictates how well we can view  $S(\cdot)$ , Fejér's kernel is referred to as the *spectral window* for  $\hat{S}^{(P)}(\cdot)$ .

To understand the implications of Equation (174b), we need to investigate the properties of Fejér's kernel.

- [1]  $\mathcal{F}(\cdot)$  is even, periodic with a period of  $2f_N$  and continuous (these follow directly from (174c) and the properties of Dirichlet's kernel);
- [2] it follows from Exercise [1.2c] that

$$N^2 \mathcal{D}_N^2(f \Delta_t) = \left| \sum_{t=0}^{N-1} e^{i2\pi f t \Delta_t} \right|^2, \text{ so we have } \mathcal{F}(f) = \frac{\Delta_t}{N} \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} e^{i2\pi f(t-u) \Delta_t}; \quad (175a)$$

a term-by-term integration of the right-hand side shows that

$$\int_{-f_N}^{f_N} \mathcal{F}(f) df = 1; \quad (175b)$$

- [3] for all sample sizes  $N$ ,  $\mathcal{F}(0) = N \Delta_t$  (this follows readily from Equations (174c) and (17c));
- [4] for  $N > 1$ ,  $f \in [-f_N, f_N]$  and  $f \neq 0$ ,  $\mathcal{F}(f) < \mathcal{F}(0)$  (a calculus problem);
- [5] for any fixed  $f \in [-f_N, f_N]$  except for  $f = 0$ ,  $\mathcal{F}(f) \rightarrow 0$  as  $N \rightarrow \infty$  (the  $N$  in the denominator of (174d) dominates); and
- [6] for any integer  $k \neq 0$  such that  $f_k = k/(N \Delta_t)$  satisfies  $|f_k| \leq f_N$ , we have  $\mathcal{F}(f_k) = 0$  (because  $\sin(k\pi) = 0$ ).

The quantities  $10 \log_{10}(\mathcal{F}(f))$  and  $\mathcal{F}(f)$  are plotted versus  $f$  in, respectively, the left- and right-hand columns of Figure 176 for  $\Delta_t = 1$  and  $N = 4, 16$  and  $64$ . We see that  $\mathcal{F}(\cdot)$  is indeed an even function and consists of a broad central peak (also called the *central lobe*) and  $N - 2$  sidelobes, which decrease in magnitude as  $|f|$  increases. The sidelobes just before and after the central lobe are only about 13 dB below that of the central lobe. This pattern holds for all  $N$ , but note that the location of these sidelobes gets closer and closer to zero as  $N$  increases (this is consistent with property [5]).

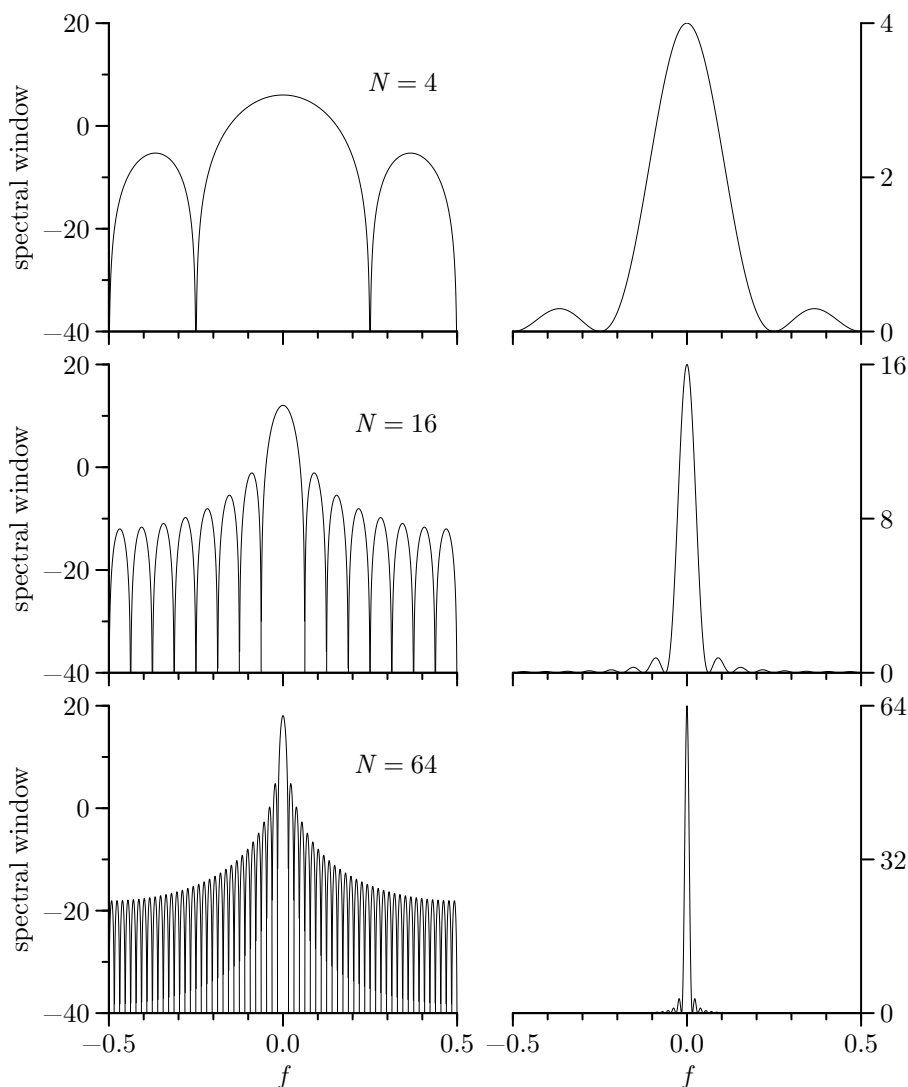
From properties [2], [3] and [5], it follows that, as  $N \rightarrow \infty$ , Fejér's kernel acts as a Dirac delta function with an infinite spike at  $f = 0$ . Since  $S(\cdot)$  is assumed to be continuous, we have from Equation (174b)

$$\lim_{N \rightarrow \infty} E\{\hat{S}^{(P)}(f)\} = S(f); \quad (175c)$$

i.e., for all  $f$ ,  $\hat{S}^{(P)}(f)$  is an asymptotically unbiased estimator of  $S(f)$ . In addition, the following theorem gives us a rate of decrease for the bias under a certain regularity condition (Brillinger, 1981a, p. 123): if  $\{X_t\}$  is a stationary process with ACVS  $\{s_\tau\}$  such that

$$\sum_{\tau=-\infty}^{\infty} |\tau s_\tau| < \infty, \text{ then } E\{\hat{S}^{(P)}(f)\} = S(f) + O\left(\frac{1}{N}\right).$$

The regularity condition essentially implies that the ACVS dies down to zero quickly; alternatively, it implies that  $S(\cdot)$  is a smooth function (for example, the condition holds if  $S(\cdot)$  has a continuous first derivative). When this is true, the bias in the periodogram decreases at the rate of  $1/N$  – but note carefully that nothing is said about the *absolute* magnitude of the bias.

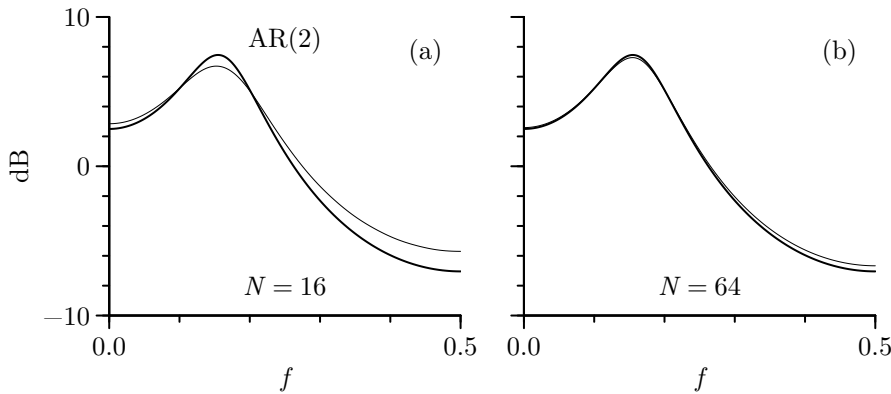


**Figure 176.** Fejér's kernel for sample sizes  $N = 4, 16$  and  $64$  on a decibel scale (left-hand plots) and a linear scale (right-hand). This kernel is the spectral window for the periodogram.

With a stronger regularity condition (say,  $\sum_{\tau=-\infty}^{\infty} |\tau^2 s_{\tau}| < \infty$ ), the rate of decrease is even faster (see Brillinger, 1981b, for details).

While there are some stationary processes for which the bias in the periodogram is negligible for small  $N$  (see, for example, the discussion surrounding Figure 177), unfortunately asymptotic unbiasedness does not imply that, for an arbitrary process, the bias in  $\hat{S}^{(P)}(f)$  is necessarily small for any particular  $N$ . A quote from Thomson (1982) is appropriate here:

... for processes with spectra typical of those encountered in engineering, the sample size must be extraordinarily large for the periodogram to be reasonably unbiased. While it is not clear what sample size, if any, gives reasonably valid results, in my experience periodogram estimates computed from 1.2 million data points on the WT4 waveguide project were too badly biased to be useful. The best that could be said for them is that they were so obviously incorrect as not to be dangerously misleading. In other applications



**Figure 177.** Bias properties of periodogram for a process with low dynamic range. The thick curve in each plot is the true SDF  $S(\cdot)$  for the AR(2) process described by Equation (34). The thin curves are  $E\{\hat{S}^{(P)}(\cdot)\}$  for sample sizes  $N = 16$  and  $64$  plotted versus frequency. The vertical axis is in decibels – either  $10 \log_{10}(S(f))$  or  $10 \log_{10}(\hat{S}^{(P)}(f))$ .

where less is known about the process, such errors may not be so obvious.

As examples of processes for which bias in the periodogram is – and is not – of concern for sample sizes in the range of  $N = 16$  to  $1024$ , let us consider three specific stationary processes (each with  $\Delta_t = 1$  so that  $f_N = 1/2$ ). The first process is white noise with a variance of 1; the second is the AR(2) process defined by Equation (34); and the third, the AR(4) process defined by Equation (35a). The SDFs for the AR(2) and AR(4) processes are shown by the thick curves in, respectively, Figures 177 and 178. A useful crude characterization of these three processes is in terms of their *dynamic range*, which we define by the ratio

$$10 \log_{10} \left( \frac{\max_f S(f)}{\min_f S(f)} \right). \quad (177a)$$

The dynamic range of a white noise process is 0 dB; the AR(2) and AR(4) processes have dynamic ranges of about 14 dB and 65 dB, respectively (the latter is typical of the spectra for many geophysical processes).

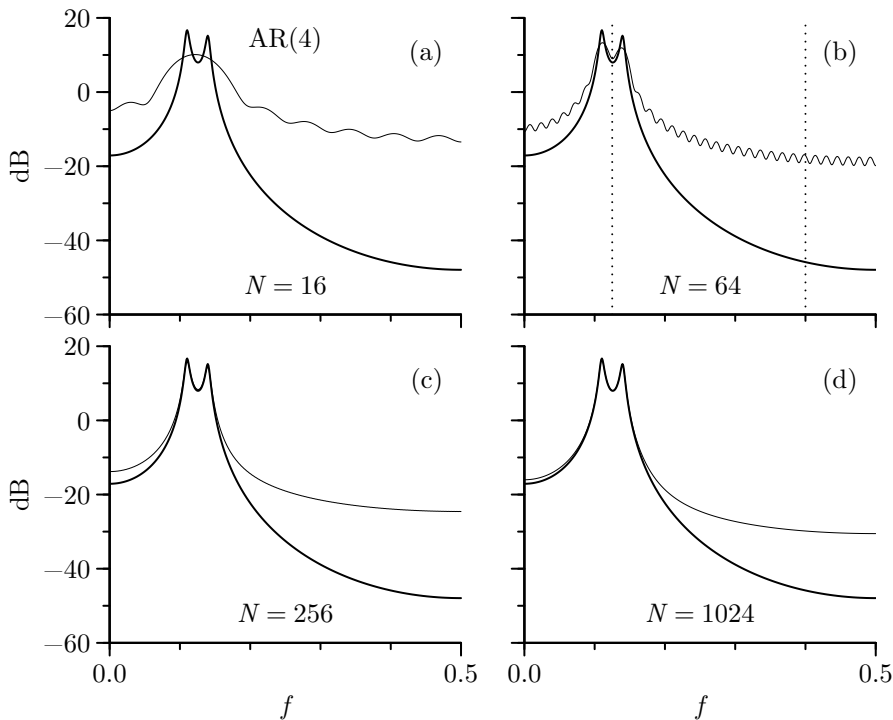
The SDF for the white noise process is just 1 for all  $f \in [-1/2, 1/2]$ . It follows from Equations (174b), (175b) and the fact that  $\mathcal{F}(\cdot)$  is periodic with a period of unity in this case that

$$E\{\hat{S}^{(P)}(f)\} = \int_{-1/2}^{1/2} \mathcal{F}(f - f') \, df' = 1, \quad (177b)$$

so the periodogram is unbiased in the case of white noise.

Figure 177 shows  $E\{\hat{S}^{(P)}(f)\}$  as a function of  $f$  for  $N = 16$  and  $64$  for the AR(2) process. For both plots in the figure the thin curve shows  $E\{\hat{S}^{(P)}(f)\}$ , while the thick curve is the true SDF. For  $N = 16$  we see that  $E\{\hat{S}^{(P)}(f)\}$  is within 2 dB of  $S(f)$  for all  $f$ , while the two functions nearly coincide for  $N = 64$ . Thus the periodogram is already largely unbiased by  $N = 64$  for this process with a rather small dynamic range. This unbiasedness is evident in the top two rows of Figures 172 and 173: although highly erratic, the periodograms for the four AR(2) time series of length  $N = 1024$  generally follow the true SDF.

The situation is quite different for the AR(4) process, which has the largest dynamic range of the three. For  $N = 64$  (Figure 178(a)) there are biases of more than 30 dB (i.e., three orders of magnitude) for  $f$  close to  $1/2$ . Even for  $N = 1024$  (Figure 178(d)),  $E\{\hat{S}^{(P)}(f)\}$



**Figure 178.** Bias properties of periodogram for a process with high dynamic range.

and  $S(f)$  differ by almost 20 dB (two orders of magnitude) for values of  $f$  for which  $S(f)$  is small. This bias is apparent in the bottom two rows of Figure 173, where we see that the actual degree of bias can vary considerably from one AR(4) time series to the next. While  $E\{\hat{S}^{(p)}(f)\}$  should reflect an average over many realizations, it need not be a good match for any particular realization.

The bias in the periodogram for processes with high dynamic range can be attributed to the sidelobes of Fejér's kernel. Equation (174b) says that  $E\{\hat{S}^{(p)}(f)\}$  is the integral of the product of that kernel centered at  $f$  with the true SDF. If Fejér's kernel consisted of just a central lobe, the product would involve just frequencies close to  $S(f)$ ; because of the sidelobes, the product also involves distant frequencies. For the AR(4) process,  $E\{\hat{S}^{(p)}(f)\}$  is badly biased at those frequencies for which  $S(f)$  is small compared with the rest of the SDF. This transfer of power from one region of the SDF to another via the kernel of the convolution is known as *leakage*.

As an illustration of how leakage arises in the AR(4) example, let's consider the case  $N = 64$  at the frequencies  $f = 1/8$  and  $f = 2/5$  (marked by dotted lines in Figure 178(b)). The first row of Figure 180 shows both  $S(f)$  (thick curves) and  $\mathcal{F}(\frac{1}{8} - f)$  (thin) versus  $f \in [-1/2, 1/2]$  with the frequency  $f = 1/8$  indicated by dotted lines (plots (a) and (b) depict the same content, but the vertical scale in (a) is expressed in decibels, whereas it is linear in (b) – the remaining rows of this figure have a similar layout). The second row shows the product  $S(f)\mathcal{F}(\frac{1}{8} - f)$ , which is concentrated about  $f = 1/8$ . Equation (174b) says that the integral of this product is equal to  $E\{\hat{S}^{(p)}(1/8)\}$ , and Figure 178(b) shows that  $E\{\hat{S}^{(p)}(1/8)\}$  agrees fairly well with  $S(1/8)$  (note that the agreement becomes even better as we increase the sample size to  $N = 256$  or  $1024$ ). This good agreement is due the fact that the frequency  $f = 1/8$  falls in the high-power “twin-peaks” portion of the SDF. By contrast, the bottom



two rows consider instead the frequency  $f = 2/5$ , which falls in the low-power portion of the AR(4) SDF. The alignment of  $\mathcal{F}(\frac{2}{5} - f)$  with  $S(f)$  is such that their product (shown in plots (g) and (h) of Figure 180) is not concentrated about  $f = 2/5$ , but rather is dominated by the interaction of the sidelobes of Fejér's kernel with the twin-peaks part of the SDF. The integral of  $S(f)\mathcal{F}(\frac{2}{5} - f)$  over  $f \in [-1/2, 1/2]$  is thus not a good approximation to  $S(2/5)$ , as is indicated by Figure 178(b).

The thin curve in Figure 178(d) showing  $10 \log_{10}(E\{\hat{S}^{(P)}(f)\})$  versus  $f$  is replicated in plots (e) to (h) of Figure 173 so that we can compare the periodograms for the four AR(4) time series with their expected values. We see that, over the portions of the SDF subject to leakage (i.e., where  $E\{\hat{S}^{(P)}(f)\}$  differs substantially from  $S(f)$ ), the actual periodogram can be systematically above or below its expected value (in fact the periodogram in Figure 173(h) is in better agreement with the true SDF than with its expected value). Thus different realizations of the AR(4) process can exhibit a degree of leakage that differs from what  $E\{\hat{S}^{(P)}(\cdot)\}$  predicts. C&E [2] offers an explanation as to why the four AR(4) time series have varying amounts of leakage based upon how well the beginning and end of each series match up.

There are two common techniques for lessening the bias in the periodogram – tapering and prewhitening. The former modifies the kernel in the convolution in Equation (174b); the latter preprocesses a time series to reduce the dynamic range of the SDF to be estimated. These techniques are discussed in the next two sections.

### Comments and Extensions to Section 6.3

[1] Equation (171b) says that we can compute the periodogram over the grid of Fourier frequencies by suitably scaling the square magnitude of the DFT of the time series. The periodogram, however, is defined for all  $f \in \mathbb{R}$ , and, as we shall see, there are good reasons for wanting to evaluate the periodogram over finer grids of frequencies. The key to doing so is zero padding, as the following exercise indicates.

▷ **Exercise [179]** For any integer  $N' > N$ , show that the DFT of the zero padded series

$$X_0, X_1, \dots, X_{N-1}, \underbrace{0, \dots, 0}_{N' - N \text{ zeros}}$$

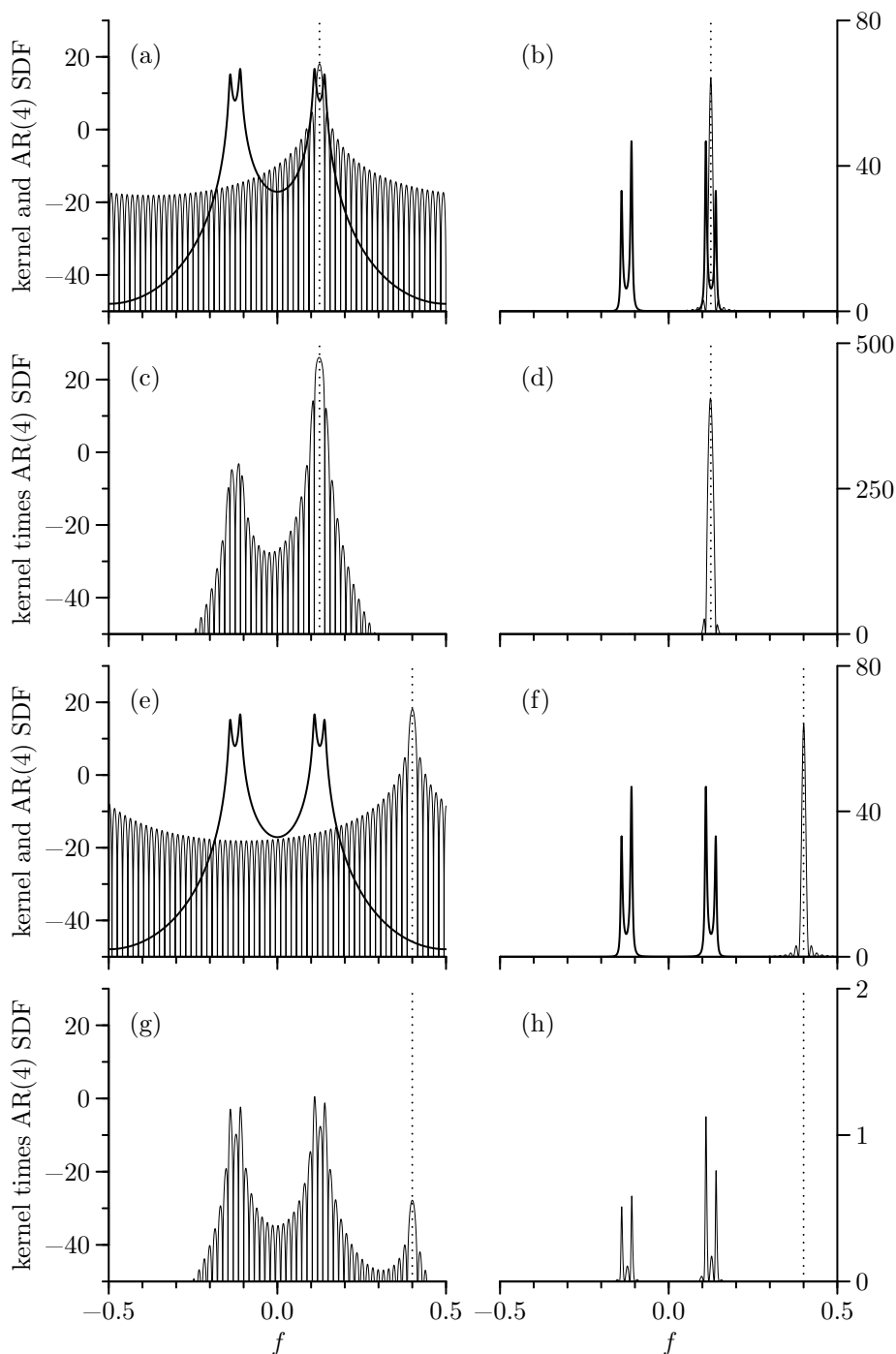
can be used to give the periodogram over a grid of frequencies dictated by  $f'_k \stackrel{\text{def}}{=} k/(N' \Delta_t)$ , which has a finer spacing than that given by the Fourier frequencies  $f_k$ . Show additionally that the placement of the zeros doesn't matter in the sense that, if  $N_1$  and  $N_2$  are any nonnegative integers such that  $N_1 + N_2 = N' - N$ , then the DFT of

$$\underbrace{0, \dots, 0}_{N_1 \text{ zeros}}, X_0, X_1, \dots, X_{N-1}, \underbrace{0, \dots, 0}_{N_2 \text{ zeros}}$$

can again be used to evaluate the periodogram over the frequencies  $f'_k$ . ◁

Thus, by padding a time series with enough zeros, we can use the DFT to compute its periodogram over an arbitrarily dense grid of frequencies.

Equation (171e) hints at one reason for wanting to evaluate the periodogram over a grid of frequencies other than the Fourier frequencies: if we know the periodogram over the grid defined by  $f'_k = k/[(2N - 1) \Delta_t]$ , then we can use an inverse DFT to compute the biased estimator of the ACVS (see Section 6.7 for more details). We can readily compute the periodogram over  $f'_k$  by zero padding the time series with  $N - 1$  zeros; however, since the advent of the fast Fourier transform (FFT) algorithm, sample sizes of digitally recorded time series are often chosen to be a power of two, say,  $N = 2^J$ , where  $J$  is a positive integer. If we pad such a series with  $N - 1$  zeros, the length  $N' = 2^{J+1} - 1$  of the padded series is not a power of two, and we cannot take advantage of some simple implementations of



**Figure 180.** Illustration of how leakage arises in the periodogram (left-hand plots have vertical scales in decibels; right-hand plots show the same content, but on linear scales). The first and third rows show the AR(4) SDF (thick curves) along with Fejér's kernel (thin) with its central lobe located at either  $f = 1/8$  (first row) or  $f = 2/5$  (third). The second and fourth rows show the product of the Fejér's kernel and the SDF in the rows just above.

the FFT algorithm. For this reason, it is convenient to add another zero so that  $N'$  becomes  $2^{J+1}$ , again a power of two. In effect we are evaluating the periodogram over the grid defined by  $\tilde{f}_k = k/(2N \Delta_t)$ , which has a spacing twice as fine as the one given by the Fourier frequencies  $f_k = k/(N \Delta_t)$ . Using an argument similar to the one leading to Equation (171e), we find that

$$\{\hat{s}_\tau^{(P)} : \tau = -(N-1), \dots, N\} \longleftrightarrow \{\hat{S}^{(P)}(\tilde{f}_k) : k = -(N-1), \dots, N\}, \quad (181)$$

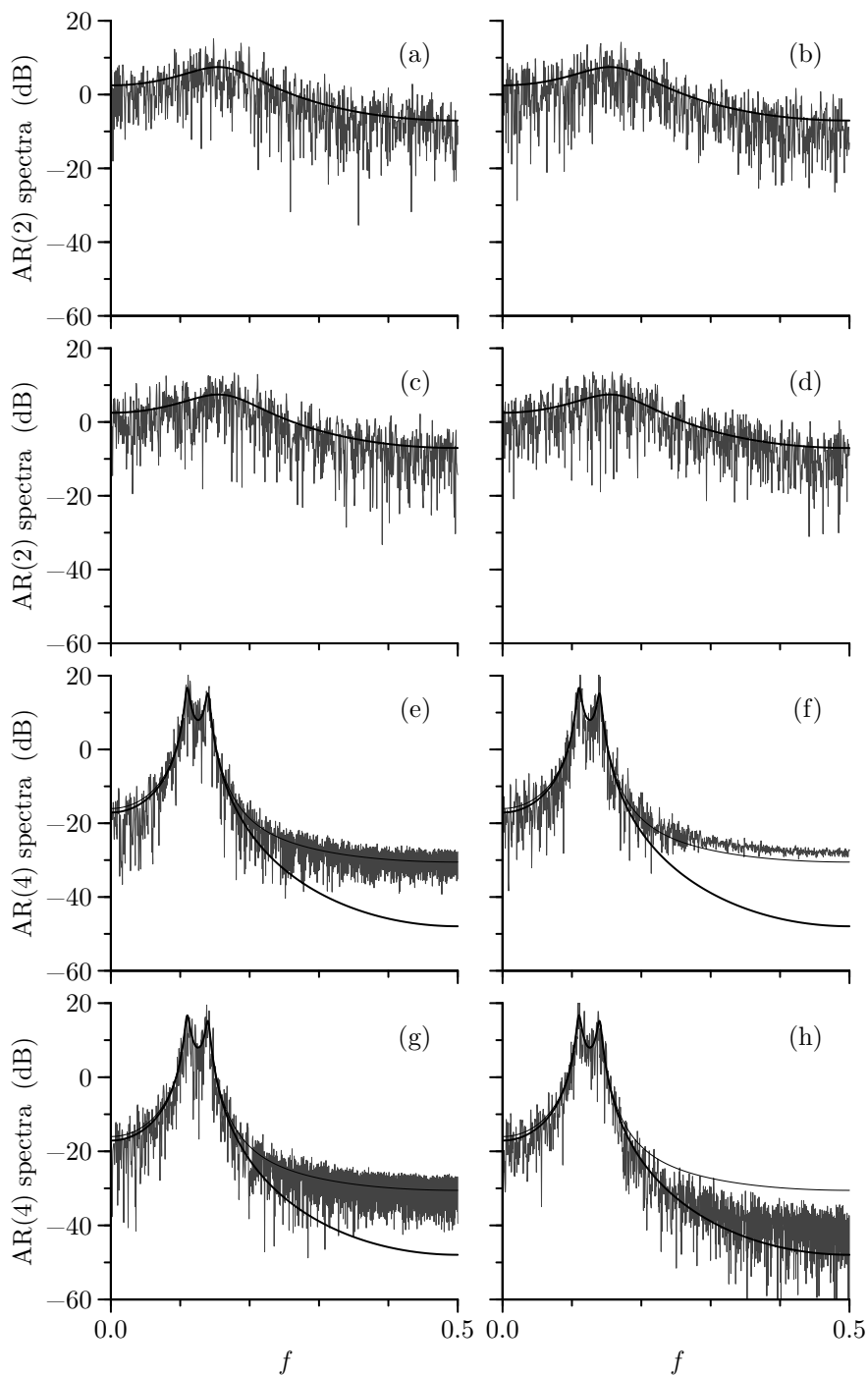
where we recall that  $\hat{s}_N^{(P)} = 0$  by definition. Hence grids defined by either  $k/[(2N-1)\Delta_t]$  or  $k/(2N\Delta_t)$  allow computation of  $\{\hat{s}_\tau^{(P)}\}$  via an inverse DFT. Exercise [6.7b] notes that if we use a grid at least as finely spaced as  $1/[(2N-1)\Delta_t]$ , then we can recover  $\{\hat{s}_\tau^{(P)}\}$  from  $\{\hat{S}^{(P)}(f'_k)\}$ ; on the other hand, we cannot fully recover  $\{\hat{s}_\tau^{(P)}\}$  if we use a grid with any coarser spacing – this includes the one defined by the Fourier frequencies  $f_k$ . (We note in C&E [4] for Section 7.1 that the common practice of evaluating the periodogram over just the Fourier frequencies can lead to a curiosity that disappears if the  $\tilde{f}_k$  grid is used instead.)

For comparison, Figure 182 shows the same eight periodograms as depicted in 173, but now evaluated over the “twice-as-fine” grid  $\tilde{f}_k$  rather than the Fourier frequencies. While most of the periodograms are qualitatively similar, the ones in plots (e), (g) and (h) look different in the high-frequency portion of the AR(4) SDF, which is subject to leakage. The amount of leakage is greater in 182(g) and (h) than in 173(g) and (h), but appears to be less in 182(e) than in 173(e). Thus examining the periodogram at just the Fourier frequencies might not give a good picture of how it varies over  $f \in [0, f_N]$ .

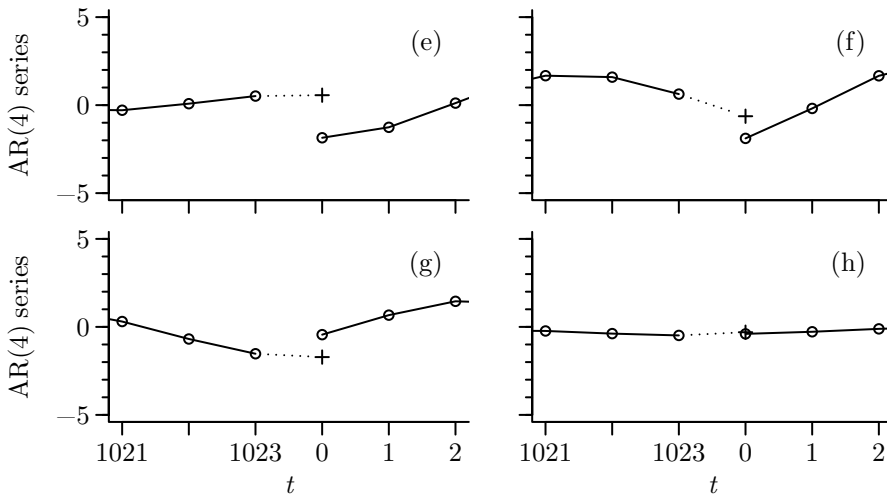
[2] Leakage in the periodogram has sometimes been attributed to a mismatch between the beginning and end of a time series (see, for example, Fougere, 1985b). This argument is based on the interpretation that the DFT treats a time series of length  $N$  as it were a periodic sequence with a period of  $N$  (see Section 3.11). Credence for this interpretation is supplied by Exercise [6.12], which notes that the periodograms for a time series and all of its  $N-1$  circular shifts are identical at the Fourier frequencies (Equation (45a) displays the circular shifts for a time series of length  $N=4$ ). The reciprocity relationships between the time and frequency domains suggest that a sharp feature in one domain manifests itself as a diffuse feature in the other. Because of the implied periodicity, a mismatch between the beginning and end of a time series is a sharp feature that manifests itself in the periodogram as leakage (a diffuse feature).

To explore this interpretation of leakage, the circles in Figure 183a show the ends  $X_{1021}, X_{1022}, X_{1023}$  and beginnings  $X_0, X_1, X_2$  of the four AR(4) time series plotted in Figure 34. We can use the definition for the AR(4) process in Equation (35a) to forecast where we would expect  $X_{1024}$  to be for each series, given the observed values  $X_{1020}, \dots, X_{1023}$  (this is done by letting  $t=1024$  and setting  $\epsilon_{1024}$  to its expected value of zero). These forecasts are indicated by pluses in the figure. Given the presumed periodicity, the magnitude of the difference between the forecast of  $X_{1024}$  and the observed  $X_0$  is a rough measure of the mismatch between the beginning and end of each series. The mismatch is largest for series (e) and smallest for series (h). A comparison with Figures 173 and 182 shows that leakage in the four periodograms does seem to increase as the mismatch increases. Figure 183b looks at this relationship further. Here we generated 100 additional AR(4) time series of length  $N=1024$ . For each series we computed its periodogram over the  $\tilde{f}_k = k/(2N)$  grid of frequencies and then formed the average of  $\hat{S}^{(P)}(\tilde{f}_k)$  over  $0.4 < \tilde{f}_k \leq 0.5$  (for this range of frequencies, the true SDF is relatively flat, and the periodogram is prone to leakage). The mean levels for the true SDF and the expected value of the periodogram over  $f \in [0.4, 0.5]$  are  $-47.2$  dB and  $-30.3$  dB (indicated by the dotted and dashed lines in the figure). The circles show the 100 averages from the periodograms versus the absolute values of the difference between  $X_0$  and the prediction of  $X_{1024}$ . The curve is a smoothing of the circles using a locally weighted regression (Cleveland, 1979). We see that, while the degree of leakage does tend to increase as the absolute prediction error increases, a small error does not necessarily imply the absence of leakage.

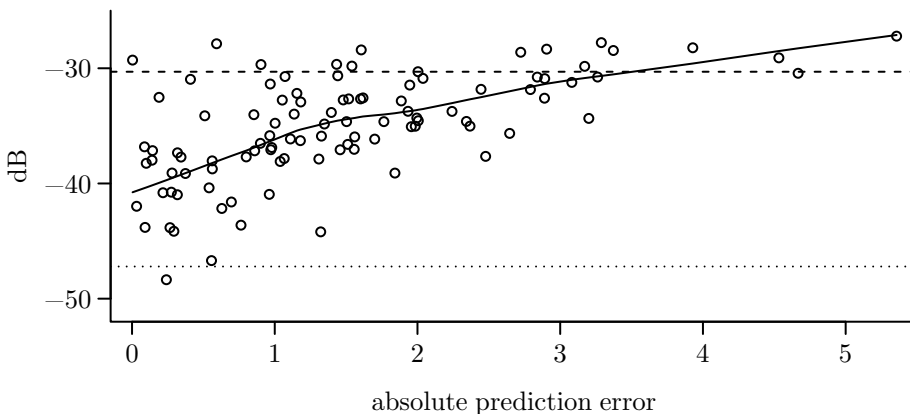
As we note in the next section, tapering essentially forces both the beginning and the end of a time series toward zero and hence can be thought of as one way of compensating for a potential end-point mismatch. Two other methods that achieve a similar effect are based on end-point matching and use of reflecting boundary conditions. End point matching consists of subtracting off a line prior to computing



**Figure 182.** As in Figure 173, but now with the periodograms evaluated over a grid of frequencies twice as fine as the Fourier frequencies, i.e.,  $\tilde{f}_k = k/(2N)$  rather than  $f_k = k/N$ .



**Figure 183a.** Detailed look at the beginning and end of each of the four AR(4) time series shown in the bottom of Figure 34 (circles), along with a forecast of  $X_{1024}$  (plotted as a plus at  $t = 0$ ). The distance between the plus and circle at  $t = 0$  is a measure of the mismatch between the extremes of the time series. This measure matches up well with the degree of leakage evident in the corresponding periodogram (see Figures 173 and 182).



**Figure 183b.** Demonstration that degree of observed leakage in AR(4) periodograms is influenced by the mismatch between the beginning and end of the time series.

the periodogram so that the ends of the adjusted series  $\{\tilde{X}_t\}$  are both zero:

$$\tilde{X}_t = X_t - X_0 - \frac{X_{N-1} - X_0}{N-1}t, \quad t = 0, \dots, N-1. \quad (183a)$$

While this method adequately corrects the bias in the periodogram for some time series (Fougere, 1985b), computer experiments indicate that it fails to do so for others, including rather spectacularly in our AR(4) example – see Exercise [6.14] (for these series, tapering accomplishes more than just end-point matching because it forces a smooth transition between the beginning and end of the series). The idea behind reflecting boundary conditions is to use either the series of length  $2N$

$$X_0, X_1, X_2, \dots, X_{N-2}, X_{N-1}, X_{N-1}, X_{N-2}, \dots, X_2, X_1, X_0 \quad (183b)$$

or the series of length  $2N - 2$

$$X_0, X_1, X_2, \dots, X_{N-2}, X_{N-1}, X_{N-2}, \dots, X_2, X_1 \quad (183c)$$

in place of the original time series. By construction, both these extended series have end points that match up well and hence are a better match for the periodic treatment imposed on a time series by the DFT. The use of reflecting boundary conditions is related to a method for estimating the SDF based upon the discrete cosine transform of type II (DCT-II) – see C&E [7] for Section 6.6 for further discussion.

[3] If  $E\{X_t\} = \mu \neq 0$  as we have assumed so far, we can replace  $X_t$  by either

$$X_t - \mu \text{ or } X_t - \bar{X} \quad (184)$$

accordingly as  $\mu$  is either known or unknown. Three interesting points about this centering of  $\{X_t\}$  are the subject of Exercise [6.15]. First, for *any* value  $C$  independent of  $t$ ,

$$\begin{aligned} \tilde{S}^{(P)}(f_k) &\stackrel{\text{def}}{=} \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} (X_t - C) e^{-i2\pi f_k t \Delta_t} \right|^2 \\ &= \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi f_k t \Delta_t} \right|^2 = \hat{S}^{(P)}(f_k), \end{aligned}$$

where  $f_k = k/(N \Delta_t)$  for integers  $k$  such that  $0 < f_k \leq f_N$ ; i.e., the frequency  $f_k$  is one of the nonzero Fourier frequencies. Second,  $\tilde{S}^{(P)}(0) = 0$  when  $C = \bar{X}$ , the sample mean. Since the periodogram is a continuous function of frequency, this latter fact implies that the periodogram of a centered time series can be a badly biased estimate of  $S(f)$  for  $f$  close to 0 (say,  $|f| < f_1 = 1/(N \Delta_t)$ ). Third, failure to center a time series that has a nonzero mean can lead to serious bias in the periodogram at non-Fourier frequencies, with the bias being dictated by Fejér's kernel.

[4] In older works on spectral analysis, the spectrum at frequency  $f$  of a segment  $X_0, X_1, \dots, X_{N-1}$  of a stationary process is sometimes *defined* – at least implicitly – to be  $E\{\hat{S}^{(P)}(f)\}$ . This matches our definition in the case of white noise, but the plots in Figure 178 shows that the two definitions can be quite different in general. In fact, if we translate the definition (116b) for the SDF discussed in Section 4.2 into the notation of this chapter, we have

$$S(f) = \lim_{N \rightarrow \infty} E\{\hat{S}^{(P)}(f)\}.$$

Thus an operational difficulty with using  $E\{\hat{S}^{(P)}(f)\}$  as the definition for the spectrum is that the spectrum can depend critically on the sample size. This is undesirable, if for no other reason than the difficulty it introduces in meaningfully comparing the spectra from data collected in different experiments concerning the same physical phenomenon.

[5] We have chosen to label our sample of size  $N$  as  $X_0, X_1, \dots, X_{N-1}$ ; a second common convention is  $X_1, X_2, \dots, X_N$ . The periodogram is the same for both conventions: if we let  $X'_t \stackrel{\text{def}}{=} X_{t-1}$  for  $t = 1, \dots, N$ , then

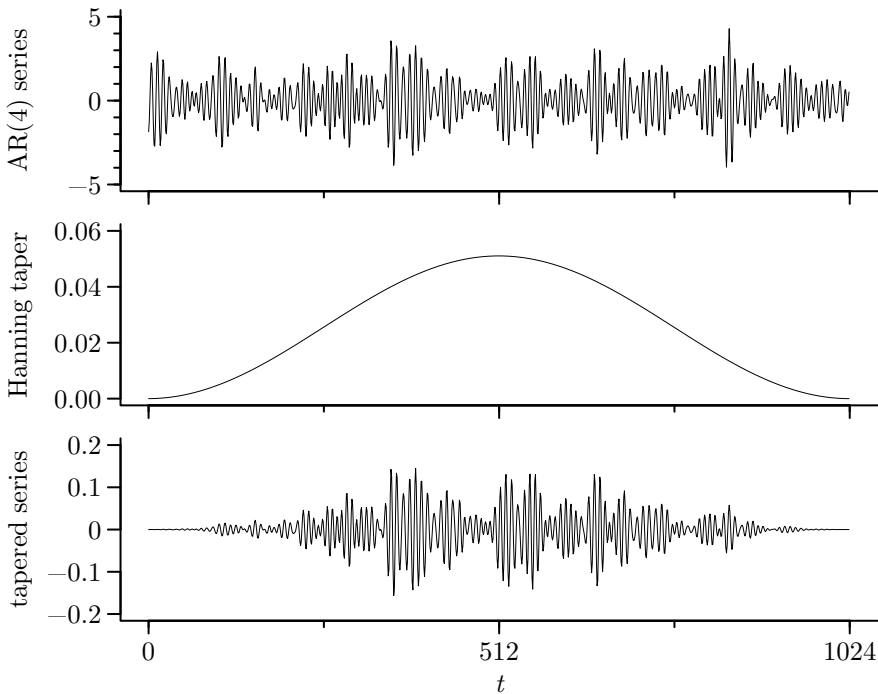
$$\left| \sum_{t=1}^N X'_t e^{-i2\pi f t \Delta_t} \right|^2 = \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi f t \Delta_t} \right|^2$$

even though in general

$$\sum_{t=1}^N X'_t e^{-i2\pi f t \Delta_t} \neq \sum_{t=0}^{N-1} X_t e^{-i2\pi f t \Delta_t}$$

due to differences in phase.

[6] Suppose we let  $\Sigma$  denote the covariance matrix of  $[X_0, X_1, \dots, X_{N-1}]^T$ , i.e., a vector whose elements are a segment of size  $N$  from a stationary process with SDF  $S(\cdot)$  (see Section 2.4). The dynamic range of  $S(\cdot)$  has an interesting relationship to an accepted measure of the ill-conditioning of



**Figure 185.** AR(4) time series  $\{X_t\}$  from Figure 34(e), a Hanning data taper  $\{h_t\}$  and their product of  $\{h_t X_t\}$  (the Hanning taper is given by Equation (189b)).

the matrix  $\Sigma$  (this is important to know if we wish to compute its inverse). Let  $\lambda_N$  and  $\nu_N$  represent the largest and smallest eigenvalues of  $\Sigma$ . This measure – called the condition number – is

$$d_N \stackrel{\text{def}}{=} \frac{\lambda_N}{\nu_N}.$$

Grenander and Szegő (1984) show that

$$\lim_{N \rightarrow \infty} d_N = \frac{\max_f S(f)}{\min_f S(f)}.$$

Hence the dynamic range of  $S(\cdot)$  is approximately equal to  $10 \log_{10}(d_N)$ , and, conversely, an approximation to the condition number  $d_N$  can be obtained from the dynamic range.

[7] In Section 6.2 we used an argument due to McLeod and Jiménez (1984, 1985) to show that any realization of  $\{\hat{s}_\tau^{(P)}\}$  must be positive semidefinite. In this section we have shown that  $\{\hat{s}_\tau^{(P)}\} \longleftrightarrow \hat{S}^{(P)}(\cdot)$ . Given that any realization of  $\hat{S}^{(P)}(\cdot)$  satisfies the properties of an SDF (a real-valued nonnegative function of frequency that is even and integrates to a finite value), it follows from Wold's theorem that the corresponding realization of  $\{\hat{s}_\tau^{(P)}\}$  is the ACVS for some stationary process and hence must be positive semidefinite (see the discussion at the end of Section 4.3).

### 6.4 Bias Reduction – Tapering

As noted in the previous section, much of the bias in the periodogram can be attributed to the sidelobes of Fejér's kernel  $\mathcal{F}(\cdot)$ . If these sidelobes were substantially smaller, there would be considerably less bias in  $\hat{S}^{(P)}(\cdot)$  due to the leakage of power from one portion of the SDF to another. Tapering is a technique that effectively reduces the sidelobes associated with Fejér's kernel (with some trade-offs). It was introduced in the context of spectral analysis by Blackman and Tukey (1958, p. 93).

Suppose that we have a time series that can be regarded as a realization of a portion  $X_0, X_1, \dots, X_{N-1}$  of a zero mean stationary process with SDF  $S(\cdot)$ . As illustrated in Figure 185, tapering a time series consists of forming the product  $h_t X_t$  for  $t = 0, \dots, N-1$ , where  $\{h_t\}$  is a suitable sequence of real-valued constants called a *data taper* (other names in the literature are *data window*, *temporal window*, *linear window*, *linear taper*, *fader* and *shading sequence*). Let

$$J(f) \stackrel{\text{def}}{=} \Delta_t^{1/2} \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi f t \Delta_t}, \quad (186a)$$

and let

$$\hat{S}^{(D)}(f) \stackrel{\text{def}}{=} |J(f)|^2 = \Delta_t \left| \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi f t \Delta_t} \right|^2. \quad (186b)$$

Spectral estimators of the above form are referred to as *direct spectral estimators*, hence the superscript on  $\hat{S}^{(D)}(\cdot)$ ; they are also called *modified periodograms*. Figure 187 shows eight examples of direct spectral estimates that correspond to the eight periodograms in Figure 182 (each of these estimates makes use of the data taper shown in Figure 185). In comparison to the periodograms, the estimates for the AR(2) series still show highly erratic fluctuations about the true SDF, with variations ranging approximately 20 dB (i.e., two orders of magnitude); however, the estimates for the four AR(4) series are in better agreement with their true SDF.

To understand why the direct spectral estimator seems to offer some improvement over the periodogram in the AR(4) case, we need to consider its first-moment properties, for which we need the following result.

- ▷ **Exercise [186]** Using the spectral representation for  $\{X_t\}$  given by Equation (113b) with  $\mu = 0$ , show that

$$J(f) = \frac{1}{\Delta_t^{1/2}} \int_{-f_N}^{f_N} H(f - f') dZ(f'), \quad (186c)$$

where  $\{h_t\} \longleftrightarrow H(\cdot)$  under the assumption that  $\{h_t\}$  is an infinite sequence with  $h_t = 0$  for  $t < 0$  and  $t \geq N$ ; i.e.,

$$H(f) = \Delta_t \sum_{t=0}^{N-1} h_t e^{-i2\pi f t \Delta_t}. \quad (186d) \triangleleft$$

Because the process  $\{Z(f)\}$  has orthogonal increments, we have (by an argument identical to that used in Exercise [4.5b])

$$E\{\hat{S}^{(D)}(f)\} = \int_{-f_N}^{f_N} \mathcal{H}(f - f') S(f') df', \quad (186e)$$

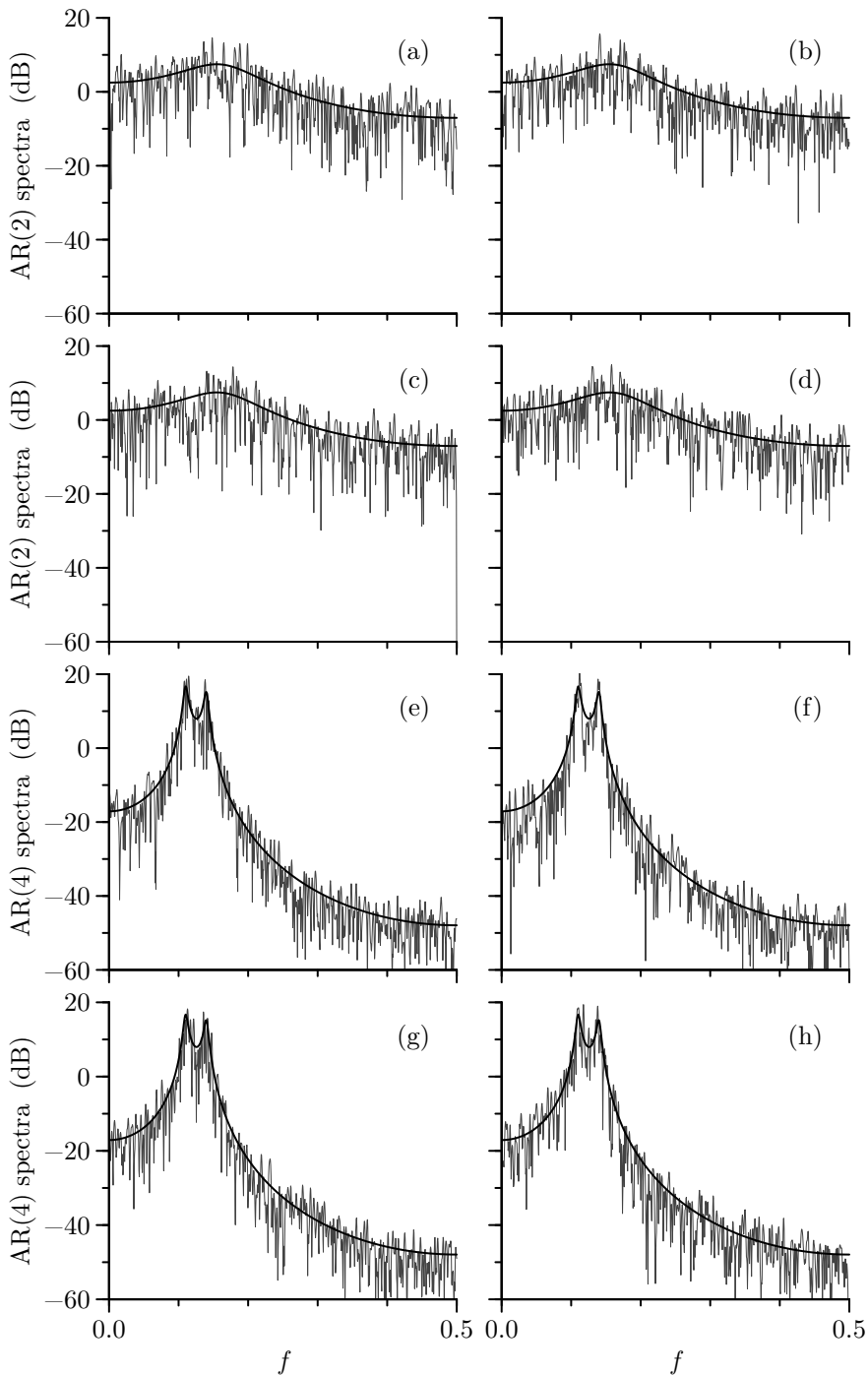
where

$$\mathcal{H}(f) \stackrel{\text{def}}{=} \frac{1}{\Delta_t} |H(f)|^2 = \Delta_t \left| \sum_{t=0}^{N-1} h_t e^{-i2\pi f t \Delta_t} \right|^2. \quad (186f)$$

For computational purposes, we note that

$$E\{\hat{S}^{(D)}(f)\} = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \left( \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t \right) s_\tau e^{-i2\pi f \tau \Delta_t} \quad (186g)$$





**Figure 187.** As in Figure 182, but now showing direct spectral estimates rather than periodograms (the direct spectral estimates use the Hanning data taper shown in Figure 185 and defined in Equation (189b)).

(see Exercise [6.16]). If  $h_t = 1/\sqrt{N}$  for  $0 \leq t \leq N-1$  (the so-called rectangular or default taper), we have

$$\hat{S}^{(D)}(f) = \hat{S}^{(P)}(f) \text{ and } \mathcal{H}(f) = \mathcal{F}(f).$$

That is,  $\hat{S}^{(D)}(\cdot)$  reverts to the periodogram;  $\mathcal{H}(\cdot)$  becomes Fejér's kernel; and Equation (186g) reduces to Equation (174a) because the inner sum in the former reduces to  $1 - |\tau|/N$  when  $h_t = 1/\sqrt{N}$ . Just as we referred to  $\mathcal{F}(\cdot)$  as the spectral window for the periodogram, so do we call  $\mathcal{H}(\cdot)$  the spectral window for the direct spectral estimator. We note that

$$\hat{S}^{(D)}(f) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{\tau}^{(D)} e^{-i2\pi f \tau \Delta_t}, \quad (188a)$$

where

$$\hat{s}_{\tau}^{(D)} \stackrel{\text{def}}{=} \begin{cases} \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} X_{t+|\tau|} h_t X_t, & |\tau| \leq N-1; \\ 0, & |\tau| \geq N \end{cases} \quad (188b)$$

(see Exercise [6.17]). Since  $\{\hat{s}_{\tau}^{(D)}\} \longleftrightarrow \hat{S}^{(D)}(\cdot)$ , we see that the sequence  $\{\hat{s}_{\tau}^{(D)}\}$  is the estimator of the ACVS corresponding to the SDF estimator  $\hat{S}^{(D)}(\cdot)$ .

The following exercise points to the need to properly normalize a data taper.

▷ **Exercise [188]** Show that

$$\int_{-f_N}^{f_N} E\{\hat{S}^{(D)}(f)\} df = s_0 \sum_{t=0}^{N-1} h_t^2. \quad \triangleleft$$

Thus, if we normalize  $\{h_t\}$  such that it has unit energy, i.e.,

$$\sum_{t=0}^{N-1} h_t^2 = 1, \quad (188c)$$

we have the desirable property

$$\int_{-f_N}^{f_N} E\{\hat{S}^{(D)}(f)\} df = s_0 = \int_{-f_N}^{f_N} S(f) df \quad (188d)$$

(note that (188c) holds for the rectangular data taper  $h_t = 1/\sqrt{N}$ ). Because  $\{h_t\} \longleftrightarrow H(\cdot)$  and since  $\mathcal{H}(\cdot)$  is proportional to the square modulus of  $H(\cdot)$ , it follows from Parseval's theorem (Equation (75b)) that

$$\sum_{t=0}^{N-1} h_t^2 = \int_{-f_N}^{f_N} \mathcal{H}(f) df.$$

The normalization (188c) is thus equivalent to

$$\int_{-f_N}^{f_N} \mathcal{H}(f) df = 1, \quad (188e)$$

which becomes Equation (175b) in the case of a rectangular data taper. For a white noise process with variance  $\sigma^2$  and SDF  $S(f) = \sigma^2 \Delta_t$ , it follows from (186e) that

$$E\{\hat{S}^{(D)}(f)\} = \sigma^2 \Delta_t \int_{-f_N}^{f_N} \mathcal{H}(f - f') \, df' = \sigma^2 \Delta_t \int_{-f_N}^{f_N} \mathcal{H}(f) \, df = \sigma^2 \Delta_t$$

(the second integral above is the same as the first since  $\mathcal{H}(\cdot)$  is periodic with period  $2f_N$ ). Thus direct spectral estimators that satisfy (188c) are unbiased in the case of white noise – a fact that we have already noted for the special case of a rectangular data taper.

The key idea behind tapering is to select  $\{h_t\}$  so that  $\mathcal{H}(\cdot)$  has much smaller sidelobes than  $\mathcal{F}(\cdot)$ . A discontinuity in a sequence of numbers means that there will be ripples in its Fourier transform (see Section 3.8). Since  $h_t = 0$  for  $t < 0$  and  $t \geq N$ , we can have small sidelobes in  $\mathcal{H}(\cdot)$  by making  $h_t$  decay to 0 in a smoother fashion than the abrupt change in the rectangular data taper. That this indeed can be done is the subject of Figure 190, which shows eight data tapers for sample size  $N = 64$ , and in Figure 191, which shows their associated spectral windows. For the sake of comparison, the first taper (Figure 190(a)) is just the rectangular data taper, so its spectral window is Fejér's kernel. The taper below it is the so-called 20% cosine taper. The  $p \times 100\%$  cosine taper is defined by

$$h_t = \begin{cases} C \left[ 1 - \cos \left( \frac{\pi(t+1)}{\lfloor pN/2 \rfloor + 1} \right) \right], & 0 \leq t \leq \lfloor \frac{pN}{2} \rfloor - 1; \\ 2C, & \lfloor \frac{pN}{2} \rfloor \leq t \leq N - \lfloor \frac{pN}{2} \rfloor - 1; \\ h_{N-t-1}, & N - \lfloor \frac{pN}{2} \rfloor \leq t \leq N - 1, \end{cases} \quad (189a)$$

where  $C$  is a normalizing constant that forces  $\sum_{t=0}^{N-1} h_t^2 = 1$  (when  $p = 0$ , the above reduces to  $h_t = 1/\sqrt{N}$ , i.e., the rectangular data taper). Thus the 20% cosine taper is like the rectangular taper, except the first 10% and last 10% of the points have been replaced by portions of a raised cosine. Figures 190(c) and (d) show the cases  $p = 0.5$  and  $p = 1$ , i.e., a 50% cosine taper and a 100% cosine taper, the latter being one formulation of the *Hanning data taper*:

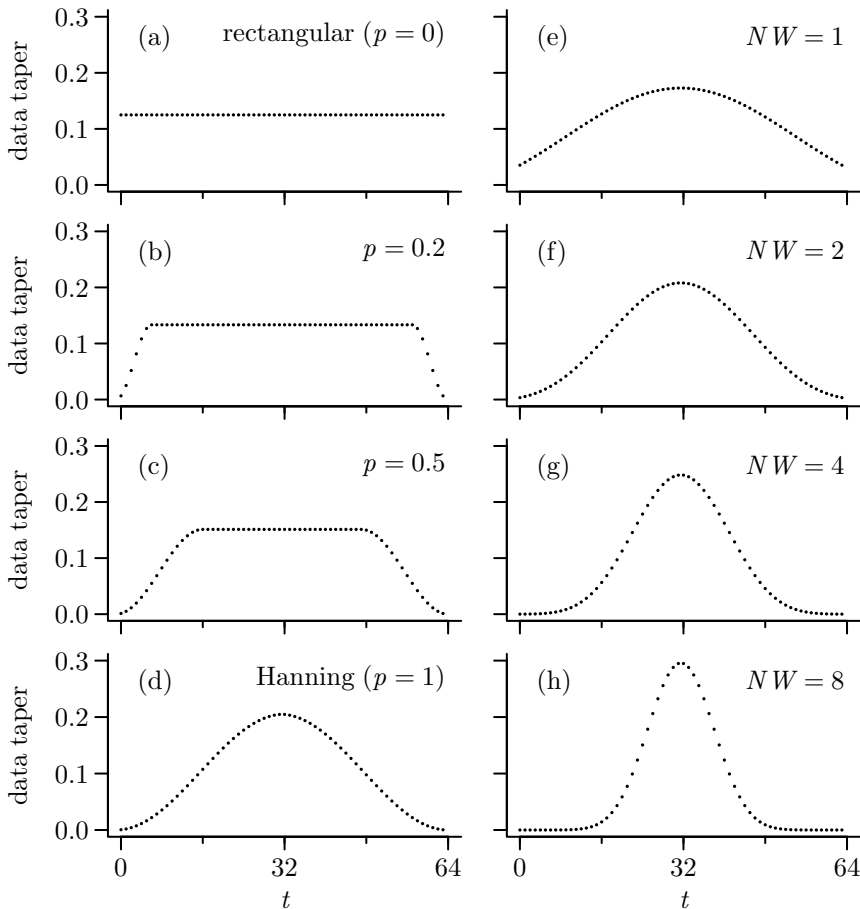
$$h_t = \begin{cases} C \left[ 1 - \cos \left( \frac{\pi(t+1)}{\lfloor N/2 \rfloor + 1} \right) \right], & 0 \leq t \leq \lfloor \frac{N+1}{2} \rfloor - 1; \\ h_{N-t-1}, & \lfloor \frac{N+1}{2} \rfloor \leq t \leq N - 1, \end{cases} \quad \text{and } C = \begin{cases} \frac{\sqrt{2}}{\sqrt{(3N-2)}}, & N \text{ even}; \\ \frac{\sqrt{2}}{\sqrt{(3N+3)}}, & N \text{ odd} \end{cases} \quad (189b)$$

(see Exercise [6.18a] for an alternative definition). The effect of the 20%, 50% and 100% cosine tapers is to “force” the tapered series  $\{h_t X_t\}$  smoothly toward zero (cf. Figure 185). The left-hand column of Figure 191 shows the spectral windows for these tapers. All have significantly smaller sidelobes than those of Fejér's kernel (upper left-hand plot); however, the widths of the central lobes are all larger than that of Fejér's kernel. We have suppressed the sidelobes at the expense of wider central lobes.

Another important family of data tapers is based upon the zeroth-order Slepian sequences (also called discrete prolate spheroidal sequences). These sequences were introduced in Section 3.10 (see the discussion following Equation (88b)) and also play a prominent role in Chapter 8. Recall that the zeroth-order Slepian sequence with parameters  $N$  and  $W$  is the sequence of length  $N$  that has the largest possible concentration of its energy in the frequency interval  $[-W, W]$  as measured by the ratio

$$\beta^2(W) \stackrel{\text{def}}{=} \int_{-W}^W |H(f)|^2 \, df / \int_{-1/2}^{1/2} |H(f)|^2 \, df$$

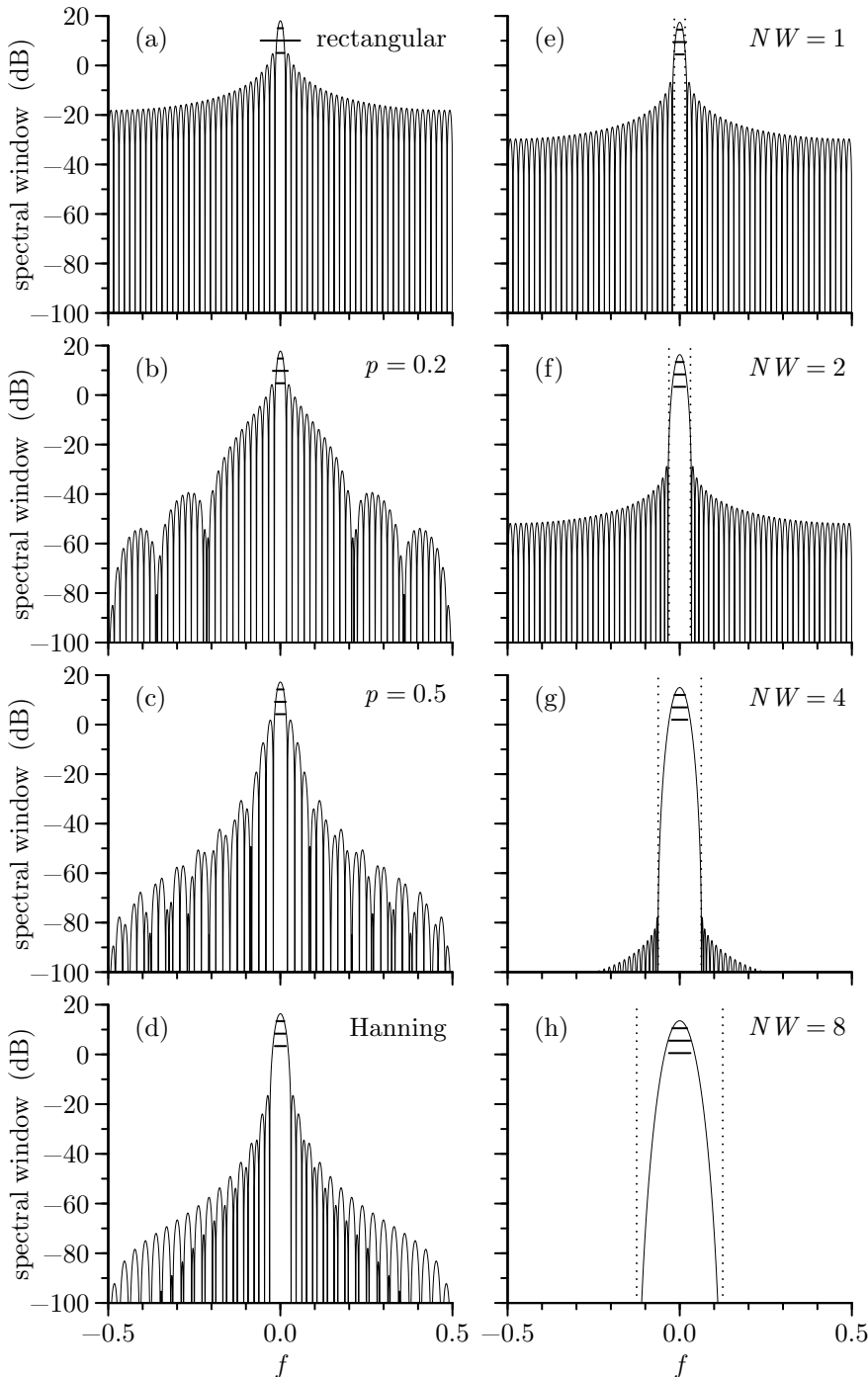
(cf. Equation (85b)). When  $\Delta_t = 1$  (as is assumed in the above equation), the spectral window  $\mathcal{H}(\cdot)$  is equal to  $|H(\cdot)|^2$  and hence is highly concentrated over  $f \in [-W, W]$ , which contains



**Figure 190.** Eight data tapers for use with a time series of length  $N = 64$ . The left-hand column shows four  $p \times 100\%$  cosine tapers ( $p = 0$ , which is the same as the default or rectangular taper,  $p = 0.2, 0.5$  and  $1.0$ , which is the Hanning taper). The right-hand column has approximations to four zeroth-order Slepian data tapers ( $NW = 1, 2, 4$  and  $8$ ).

the central lobe. The concentration of energy in the region of the central lobe implies that the sidelobes of  $\mathcal{H}(\cdot)$  over  $|f| \in [W, 1/2]$  should be small. There is a natural trade-off between the width  $2W$  of the region containing the central lobe and the energy in the sidelobes – as the former increases, the latter decreases. This point is illustrated in Figure 191. The right-hand column shows spectral windows for the four data tapers plotted in the corresponding panels of Figure 190. These are *approximations* to the zeroth-order Slepian data tapers for  $N = 64$  and  $NW = 1, 2, 4$  and  $8$  (Equation (196b) gives the approximation, which is discussed in C&E [5]). Note that the energy in the sidelobes of the spectral windows decreases markedly as  $W$  (and hence the width of the central lobe) increases. It is particularly interesting to compare the spectral window for the  $NW = 1$  Slepian data taper with that for the periodogram (Figure 191(a)). The central lobe widths of these two windows are nearly the same, yet the sidelobes for the window associated with the Slepian data taper are about 10 dB below those of Fejér's kernel.

Figure 193 illustrates how the trade-off between the central lobe width and the magnitude of the sidelobes affects the first moment of direct spectral estimators. The thin curve in each plot shows  $E\{\hat{S}^{(D)}(f)\}$  as a function of  $f$  for the AR(4) process with  $N = 64$ . The thick curves are the true SDF. The estimators in the eight plots use the data tapers depicted



**Figure 191.** Spectral windows  $\mathcal{H}(f)$  on a decibel scale (i.e.,  $10 \log_{10}(\mathcal{H}(f))$ ) versus  $f$  in cycles per unit time for the eight data tapers shown on Figure 190. The dotted lines in the right-hand column indicate the locations of  $f = \pm W$  for the Slepian tapers with  $W$  set via  $NW = 1, 2, 4$  and  $8$ . The three horizontal solid lines just below the central lobes are bandwidth measures based upon, from the top on down, the half-power, variance and autocorrelation widths.

in Figure 190. The expected value of the periodogram for this case is shown in the upper left-hand plot. As we go down each column, the degree of tapering increasing, resulting in less and less bias due to leakage at frequencies well away from the twin-peaks region. We can attribute this reduction in leakage to smaller sidelobes in the spectral windows; however, smaller sidelobes come at the expense of a wider central lobe, which acts to smooth out the SDF in the twin-peaks region. The deleterious effect of this smoothing is most apparent in the estimators using the  $NW = 4$  and 8 Slepian tapers, where the twin peaks are smoothed out into a single peak. This example shows that there is an inherent trade-off between reducing bias due to leakage (i.e., due to the sidelobes) and introducing bias due to the width of the central lobe. We refer to the bias due to the central lobe as *spectral window bias*.

We need some way of measuring the width of the central lobe of  $\mathcal{H}(\cdot)$  if we are to compare the potential effect of spectral window bias for different direct spectral estimators. A standard engineering practice is to measure the width as the distance between half-power points, i.e., those points  $f_0$  and  $-f_0$  such that  $\mathcal{H}(-f_0) = \mathcal{H}(f_0) = \mathcal{H}(0)/2$ , and hence the half-power width is  $\text{width}_{\text{hp}}\{\mathcal{H}(\cdot)\} \stackrel{\text{def}}{=} 2f_0$ . Note that  $10 \log_{10}(1/2) \approx -3$  so that the half-power points are about 3 dB down from the top of the central lobe. The half-power width for each spectral window in Figure 191 is depicted by the top-most of the three horizontal lines, drawn 3 dB below the central lobe.

While  $\text{width}_{\text{hp}}\{\mathcal{H}(\cdot)\}$  is intuitively appealing, this measure is basically a graphical procedure that in general is inconvenient to compute. As potential alternatives, let us consider measures based upon the variance width and the autocorrelation width, both of which were introduced in Chapter 3. The following exercise points out a connection between the variance width and the bias in  $\hat{S}^{(\text{D})}(f)$ .

▷ **Exercise [192]** Assume that  $S(\cdot)$  can be expanded in a Taylor series about  $f$  such that

$$S(f + \phi) \approx S(f) + \phi S'(f) + \frac{\phi^2}{2} S''(f),$$

where  $S'(\cdot)$  and  $S''(\cdot)$  are the first and second derivatives of  $S(\cdot)$ . Show that the bias  $b(f) \stackrel{\text{def}}{=} E\{\hat{S}^{(\text{D})}(f)\} - S(f)$  can be approximated by

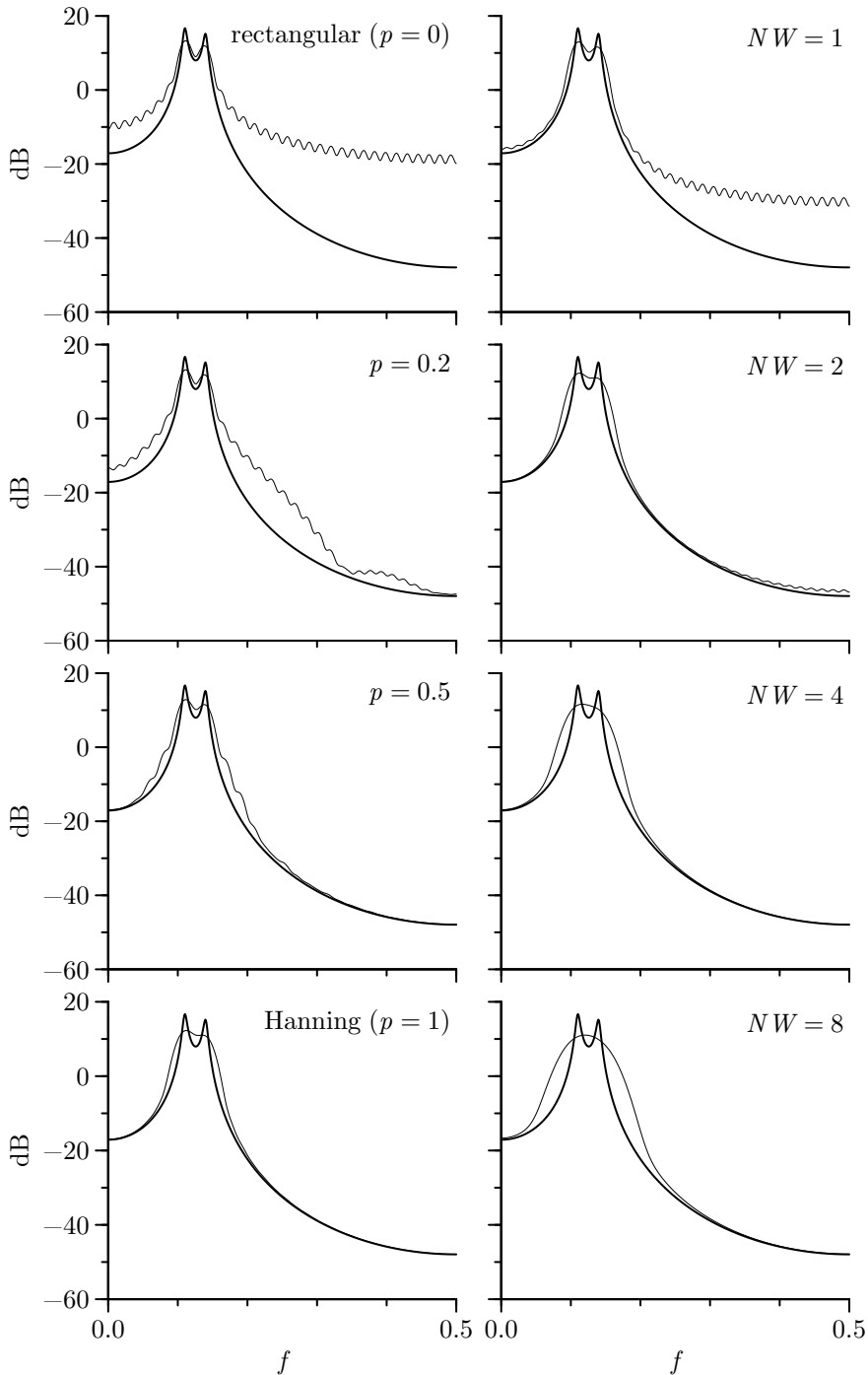
$$b(f) \approx \frac{S''(f)}{2} \int_{-f_N}^{f_N} \phi^2 \mathcal{H}(\phi) d\phi. \quad (192a) \triangleleft$$

The bias in  $\hat{S}^{(\text{D})}(\cdot)$  at frequency  $f$  is thus related to two quantities: first, the curvature (second derivative) of  $S(\cdot)$  at  $f$  – a large curvature due to a peak or a valley in  $S(\cdot)$  implies a large bias; and second, an integral that depends on  $\mathcal{H}(\cdot)$  alone – the larger this integral, the greater the bias. Because  $\mathcal{H}(\cdot)$  is necessarily nonnegative due to Equation (186f) and because its integral over  $[-f_N, f_N]$  is unity (Equation (188e)), the integral in Equation (192a) can be interpreted as the variance of an RV that takes on values over  $[-f_N, f_N]$  with a probability density function (PDF) given by  $\mathcal{H}(\cdot)$ . If we adapt the variance width defined in Equation (60c) to this PDF, we can link the bias  $b(f)$  to

$$\text{width}_v\{\mathcal{H}(\cdot)\} = \left( 12 \int_{-f_N}^{f_N} \phi^2 \mathcal{H}(\phi) d\phi \right)^{1/2} = \frac{1}{\Delta_t} \left( 1 + \frac{12}{\pi^2 \Delta_t} \sum_{\tau=1}^{N-1} \frac{(-1)^\tau}{\tau^2} h \star h_\tau \right)^{1/2}, \quad (192b)$$

where

$$h \star h_\tau = \Delta_t \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t \quad (192c)$$



**Figure 193.** Comparison of true SDF  $S(\cdot)$  and  $E\{\hat{S}^{(D)}(\cdot)\}$ . The thick curves are the true SDF of the AR(4) process of Equation (35a). The thin curves show  $E\{\hat{S}^{(D)}(\cdot)\}$  for eight direct spectral estimators using the tapers of length  $N = 64$  in Figure 190. Both  $S(\cdot)$  and  $E\{\hat{S}^{(D)}(\cdot)\}$  are displayed in decibels (i.e.,  $10 \log_{10}(S(f))$ ) versus  $f$ , where  $f$  is in cycles/ $\Delta_t$  with  $\Delta_t = 1$ .

is the autocorrelation of  $\{h_t\}$  as per Equation (99f) (the second equality in Equation (192b) is the subject of Exercise [6.19]). The variance width is depicted in Figure 191 by the middle of the three horizontal lines drawn below each central lobe. For the cases considered here,  $\text{width}_v \{\mathcal{H}(\cdot)\}$  is larger than  $\text{width}_{hp} \{\mathcal{H}(\cdot)\}$ , but the two measures are in better agreement as the degree of tapering increases, i.e., as  $p$  or  $NW$  increases. The measures are markedly different for Fejér's kernel (the rectangular taper), whose prominent sidelobes act to inflate the variance measure above what would be dictated by the central lobe alone. With the exception of Fejér's kernel, the approximation to the bias given by Equation (192a) is a reflection of spectral window bias (due to the central lobe) and has little to do with leakage (bias due to sidelobes).

A third measure of the central lobe of  $\mathcal{H}(\cdot)$  is its autocorrelation width, as defined in Equation (100f):

$$B_{\mathcal{H}} \stackrel{\text{def}}{=} \text{width}_a \{\mathcal{H}(\cdot)\} = \frac{\left( \int_{-f_N}^{f_N} \mathcal{H}(f) \, df \right)^2}{\int_{-f_N}^{f_N} \mathcal{H}^2(f) \, df} = \frac{\Delta_t}{\sum_{\tau=-(N-1)}^{N-1} (h \star h_{\tau})^2}, \quad (194)$$

where the second equality holds because  $\mathcal{H}(\cdot)$  integrates to unity and by applying Parseval's theorem to the relationship  $\{h \star h_{\tau}\} \longleftrightarrow \mathcal{H}(\cdot) \Delta_t$ . This measure of width is depicted in Figure 191 by the bottom-most horizontal lines and is seen to fall between the other two measures. The autocorrelation width is in better agreement with the half-power width than the variance width is, particularly in the case of Fejér's kernel. Because  $B_{\mathcal{H}}$  is the easiest of the three measures to compute and because of its good agreement with the intuitively appealing half-power width, we henceforth adopt it as our standard measure of the *effective bandwidth of the spectral estimator*  $\hat{S}^{(D)}(\cdot)$ . (An additional interpretation for  $B_{\mathcal{H}}$  is that, if  $0 < f < f' < f_N$  and  $|f - f'| \geq B_{\mathcal{H}}$ , then  $\hat{S}^{(D)}(f)$  and  $\hat{S}^{(D)}(f')$  are approximately uncorrelated – for details, see Walden et al., 1995, and C&E [4] for Section 6.6.)

In summary, tapering is a useful method for reducing the bias due to leakage in direct spectral estimators. This is particularly important for spectra with large dynamic range (and of less importance for those with small dynamic range). When spectral analysis is being used as a tool for an exploratory data analysis (i.e., little is known about the true SDF of the process generating the data), it is important to use spectral estimators with good bias characteristics to make sure that it is the data and *not* the spectral window that is determining the shape of the estimated SDF. Note that so far we have considered the effects of tapering as reflected only by the first-moment properties of direct spectral estimators. Other advantages and disadvantages of tapering are presented in Section 6.6, where we discuss second-moment properties.

#### Comments and Extensions to Section 6.4

[1] As we shall see in Chapter 8, tapering plays a fundamental role in modern nonparametric spectral estimation. Some researchers, however, regard it as an undesirable (and suspicious) operation. For example, Yuen (1979) states the following (edited slightly to follow the notation and nomenclature of this chapter):

... the idea [of tapering] is statistically quite unsound. ... the most serious criticism is this: [tapering] effectively “throws data away” because some of the values of  $X_t$  are multiplied into small weighting factors. ... Why is this statistically bad? In the theory of estimation it is the basic idiom of the subject that the weight given to any individual value should decrease with its variance. That is, the more uncertainty there is in the data point, the less weight it should be given. Now,  $X_t$  is a time-invariance [*sic*] random process. Its variance is constant. Thus all its values should receive equal weight.

In advocating maximum entropy spectral analysis (see Chapter 9) in preference to nonparametric spectral analysis, Fougere (1985a) stated



Also there is never any tapering – which really ought to be called tampering – of the data with resulting loss of information!

Evidently these researchers regard tapering as an operation that effectively modifies a time series. From our viewpoint, tapering is an operation that replaces Fejér's kernel in Equation (174b) with one having better sidelobe properties. Its actions should thus be judged in the frequency domain rather than the time domain. There are some caveats and trade-offs in using data tapers, but, if these are observed, tapering can be a useful way to control the bias due to leakage in direct spectral estimators for an SDF with a large dynamic range. Moreover, the multitaper approach discussed in Chapter 8 largely overcomes Yuen's objection (see also Brillinger, 1981b, for a good overall discussion of tapering in the context of a response to Yuen, 1979). Also, we note in Chapter 9 that, in one of its formulations (the Yule–Walker method), maximum entropy spectral analysis can actually be improved when used in combination with data tapering.

It should be clear from our AR(2) example that there are certain processes for which tapering with a single data taper accomplishes nothing useful (the same statement does not apply to the multitaper approach, which can be used to obtain estimates of the variability of spectral estimators even in situations where bias is not an issue – see Chapter 8). The most extreme case is a white noise process. Sloane (1969) shows that tapering such a process does little more than reduce the effective sample size; i.e., it amounts to throwing a certain portion of the data away.

[2] We have used the normalization  $\sum_{t=0}^{N-1} h_t^2 = 1$  (Equation (188c)) to ensure that  $\hat{S}^{(D)}(\cdot)$  mimics the true SDF in that the expected value of the integral of  $\hat{S}^{(D)}(\cdot)$  equals the variance of the process. An alternative normalization is to require that the sum of the squares of  $h_t X_t$  be equal to that of  $X_t/\sqrt{N}$ , i.e., to require that  $h_t$  be such that

$$\sum_{t=0}^{N-1} (h_t X_t)^2 = \frac{1}{N} \sum_{t=0}^{N-1} X_t^2 = \hat{s}_0^{(P)}. \quad (195a)$$

This makes the normalization of  $h_t$  data dependent. By taking the expected value of both sides above, we find that

$$\sum_{t=0}^{N-1} E\{h_t^2 X_t^2\} = s_0.$$

Hence the expected sum of squares of  $h_t X_t$  is an unbiased estimator of the variance  $s_0$  of the process (this assumes that  $X_t$  is known to have a mean of zero). If  $\hat{s}_0^{(P)}$  is a consistent estimator of  $s_0$ , this variance normalization (sometimes called *restoration of power* in the literature) is equivalent to that of (188c) for large  $N$ . Note that any (reasonable) normalization of  $h_t$  affects only the level of  $\hat{S}^{(D)}(\cdot)$  and not its shape.

[3] If  $E\{X_t\} = \mu \neq 0$ , we need to modify  $\hat{S}^{(D)}(\cdot)$  somehow. The obvious changes are

$$\Delta_t \left| \sum_{t=0}^{N-1} h_t (X_t - \mu) e^{-i2\pi f t \Delta_t} \right|^2 \quad \text{or} \quad \Delta_t \left| \sum_{t=0}^{N-1} h_t (X_t - \bar{X}) e^{-i2\pi f t \Delta_t} \right|^2$$

accordingly as  $\mu$  is either known or unknown. If both centering and tapering of a time series are needed, centering must be done before tapering (see Exercise [6.20]). There is, however, an alternative estimator to  $\bar{X}$  for  $\mu$  that is sometimes used in conjunction with tapering. Exercise [6.15b] shows that the periodogram  $\hat{S}^{(P)}(\cdot)$  of a centered time series is zero at zero frequency. This is arguably appropriate – since zero frequency corresponds to a constant term, the power of a centered series at zero frequency should be zero. If we now taper a centered series with a nonrectangular data taper, the resulting direct estimator  $\hat{S}^{(D)}(\cdot)$  need not be zero at zero frequency. Suppose, however, that we consider the following estimator of  $\mu$ :

$$\tilde{\mu} \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} h_t X_t / \sum_{t=0}^{N-1} h_t \quad (195b)$$

(this assumes that  $\sum h_t \neq 0$ , which is true for all the specific tapers we have considered so far, but is not true in general). Note that  $\tilde{\mu}$  is an unbiased estimator of  $\mu$  and that it reduces to  $\bar{X}$  when  $\{h_t\}$  is the rectangular data taper. If we center our time series with  $\tilde{\mu}$  instead of  $\bar{X}$  and then taper it, our direct spectral estimator becomes

$$\Delta_t \left| \sum_{t=0}^{N-1} h_t (X_t - \tilde{\mu}) e^{-i2\pi f t \Delta_t} \right|^2. \quad (196a)$$

When  $f = 0$  the quantity between the absolute value signs becomes

$$\sum_{t=0}^{N-1} h_t (X_t - \tilde{\mu}) = \sum_{t=0}^{N-1} h_t X_t - \sum_{t=0}^{N-1} h_t \left( \frac{\sum_{t=0}^{N-1} h_t X_t}{\sum_{t=0}^{N-1} h_t} \right) = 0,$$

as a result of which the direct spectral estimator is again identically zero at zero frequency. (We reconsider this alternative centering scheme in Section 10.10.)

**[4]** Since the formulae we have developed in this section show that the effect of tapering depends upon both the data taper and the true SDF, it might seem that, without knowledge of the SDF, there is no way of determining whether a data taper is required in practical applications. For fine details this is correct, but the need for tapering for a particular time series can usually be established by comparing its periodogram with a direct spectral estimate that uses a data taper with good sidelobe characteristics. Consider the bottom two rows of Figures 182 and 187, which show, respectively, periodograms and direct spectral estimates using a Hanning data taper for the four simulated AR(4) time series. If we compare the two estimates for each series, the main indicator that leakage is a concern is the difference in their levels in the high frequency region; moreover, as we would expect in the presence of substantial bias, the level of the periodogram is the higher of the two.

A secondary potential indication of the presence of leakage in a periodogram is a marked decrease in the amount of local variability in a low power region – compared with the high power region – when plotted on a logarithmic scale. This possibility is demonstrated in Figure 182(f). As we shall see in Section 6.4, nonparametric spectral estimators should have the same local variability at all frequencies when plotted on a logarithmic scale. The decrease in local variability in the high frequency region of Figure 182(f) is an indication that something is not right here – evidently this region is dominated by leakage of power from other portions of the SDF. However, it is important to note several things about this simple example. First, a decrease in local variability can also be due to other causes, a common one being the presence of outliers (discordant values) in a time series – see Martin and Thomson (1982) and Chave et al. (1987). Second, a *lack* of decrease of local variability in regions of low power does *not* automatically imply that tapering is unnecessary – consider, for example, Figure 182(g). Third, the need for tapering sometimes manifests itself as a ringing (oscillations) in a low power portion of an SDF (we illustrate this later on in Figures 226 and 227). Finally, leakage is not limited to the high frequency region of an SDF – it can potentially occur in any portion with relatively low power. The best advice we can offer is to try different degrees of tapering and to carefully check the corresponding SDF estimates for evidence of leakage – or lack thereof.

**[5]** Walden (1989) shows that, for a given  $N$  and  $W$ , the zeroth-order Slepian data taper can be approximated by

$$h_t = C I_0 \left( \pi W (N-1) \left( 1 - \frac{(2t+1-N)^2}{N^2} \right)^{1/2} \right), \quad (196b)$$

where  $C$  is a normalizing constant forcing  $\sum h_t^2 = 1$ , and  $I_0(\cdot)$  is the modified Bessel function of the first kind and zeroth order (this is an improvement on an approximation originally due to Kaiser, 1966, 1974). The advantage of Equation (196b) is that functions exist for calculating  $I_0(\cdot)$  in many high-level languages.

**[6]** The two families of data tapers we have focused on ( $p \times 100\%$  cosine and zeroth-order Slepian) are by no means the only ones that have been used extensively in spectral analysis – in fact, a bewildering number of tapers has been proposed and discussed in the literature. Harris (1978) compares 23 different

classes of tapers (see also Nuttall, 1981, and Geçkinli and Yavuz, 1978 and 1983). Some of these tapers arise as solutions to different optimality criteria. For example, Adams (1991) proposes an “optimal” taper that provides a quantified “best” trade-off between peak sidelobe level and total sidelobe energy. Researchers have also developed tapers with particular time series in mind. Hurvich and Chen (2000) provide an example in their development of a taper that is efficient for potentially overdifferentenced time series arising from a fractionally differenced or related process. Despite this plethora of choices, we have yet to encounter a practical application that provides a compelling case for using a taper outside of the cosine or Slepian families.

### 6.5 Bias Reduction – Prewhitening

In addition to tapering, another common method of controlling the bias in direct spectral estimators is *prewhitening* (Press and Tukey, 1956). This technique is based upon linear filtering theory. Suppose that we filter a portion  $X_0, \dots, X_{N-1}$  of a stationary process  $\{X_t\}$  with SDF  $S_X(\cdot)$  to create

$$Y_t \stackrel{\text{def}}{=} \sum_{l=0}^{L-1} g_l X_{t+L-1-l}, \quad 0 \leq t \leq M-1, \quad (197a)$$

where  $L \geq 2$  is the width of the filter  $\{g_l\}$ , and  $M \stackrel{\text{def}}{=} N - L + 1$ . This yields a portion of length  $M$  of a stationary process  $\{Y_t\}$  with SDF  $S_Y(\cdot)$ . The two SDFs are related by

$$S_Y(f) = \left| \sum_{l=0}^{L-1} g_l e^{-i2\pi f l \Delta_t} \right|^2 S_X(f), \quad (197b)$$

where the multiplier of  $S_X(f)$  defines the square gain function for the filter.

Now, if  $S_X(\cdot)$  has a wide dynamic range, it might be possible to choose  $\{g_l\}$  such that the dynamic range of  $S_Y(\cdot)$  is much less. This means we might be able to construct a direct spectral estimator of  $S_X(\cdot)$  with low bias by using the following procedure. First, we use the filtered data to produce

$$\hat{S}_Y^{(D)}(f) \stackrel{\text{def}}{=} \Delta_t \left| \sum_{t=0}^{M-1} h_t Y_t e^{-i2\pi f t \Delta_t} \right|^2,$$

where  $\{h_t\}$  is a data taper with a fairly narrow central lobe (such as the rectangular taper or the Slepian data taper with  $NW = 1$ ). If  $S_Y(\cdot)$  has a small dynamic range, we should have  $E\{\hat{S}_Y^{(D)}(f)\} \approx S_Y(f)$  to within a few decibels. This implies that

$$E \left\{ \frac{\hat{S}_Y^{(D)}(f)}{\left| \sum_{l=0}^{L-1} g_l e^{-i2\pi f l \Delta_t} \right|^2} \right\} \approx \frac{S_Y(f)}{\left| \sum_{l=0}^{L-1} g_l e^{-i2\pi f l \Delta_t} \right|^2} = S_X(f)$$

also to within a few decibels. Since we have effectively reduced the dynamic range of the SDF to be estimated and hence can now use a data taper with a narrow central lobe, this procedure could prevent the introduction of spectral window bias caused by data tapers with wide central lobes (see Figure 193). In the ideal case, the filter  $\{g_l\}$  would reduce  $\{X_t\}$  to white noise. It is thus called a *prewhitening filter*.

There are, however, some potential trade-offs and problems with prewhitening. There is an inevitable decrease in the sample size of the time series used in the direct spectral estimator (from  $N$  to  $M = N - L + 1$ ). This decrease can adversely affect the first-moment and other

statistical properties of the spectral estimator if  $L$  is large relative to  $N$ . This can happen in practice – an SDF with a large amount of structure can require a prewhitening filter of a large length to effectively reduce its dynamic range. Just such a situation occurs in exploration seismology where the power spectrum of seismic time series decays very rapidly with decreasing frequency because of the effect of analog recording filters.

There is also a “cart and horse” problem with prewhitening. Design of a prewhitening filter requires at least some knowledge of the shape of the underlying SDF, the very thing we are trying to estimate! Sometimes it is possible to make a reasonable guess about the general shape of the SDF from previous experiments or physical theory. This is obviously not a viable general solution, particularly since spectral analysis is often used as an exploratory data analysis tool. An alternative way of obtaining a prewhitening filter is to *estimate* it from the time series itself. One popular method for doing so is to fit an autoregressive model to the data and then use the fitted model as a prewhitening filter. The details of this approach are given in Section 9.10.

As examples, let’s consider four prewhitening filters for the AR(4) time series of Figure 34(f), for which  $\Delta_t = 1$ . The periodogram for this particular series has pronounced leakage (see Figures 173(f) and 182(f)). Letting  $\phi_{4,1}, \dots, \phi_{4,4}$  represent the coefficients of  $X_{t-1}, \dots, X_{t-4}$  in Equation (35a), we can write

$$Y_t = X_{t+4} - \sum_{l=1}^4 \phi_{4,l} X_{t+4-l} = \epsilon_{t+4},$$

where  $\{\epsilon_t\}$  is a white noise process. Hence perfect prewhitening is provided by  $\{1, -\phi_{4,1}, \dots, -\phi_{4,4}\}$ , which is a *prediction error filter* of length  $L = 5$ : we can regard  $\sum_l \phi_{4,l} X_{t+4-l}$  as a predictor of  $X_{t+4}$  based upon  $X_{t+3}, \dots, X_t$ , and the difference between  $X_{t+4}$  and its prediction is the prediction error. The top row of plots in Figure 199 shows the periodogram  $\hat{S}_Y^{(P)}(\cdot)$  for  $Y_0, Y_1, \dots, Y_{1019}$  (left-hand column) along with the theoretical SDF for  $\{Y_t\}$ , which is flat because it corresponds to a white noise process. Note here, with  $N = 1024$ , we have  $M = N - L + 1 = 1020$ . The right-hand plot shows the “postcolored” estimate of the SDF for  $\{X_t\}$ , namely,

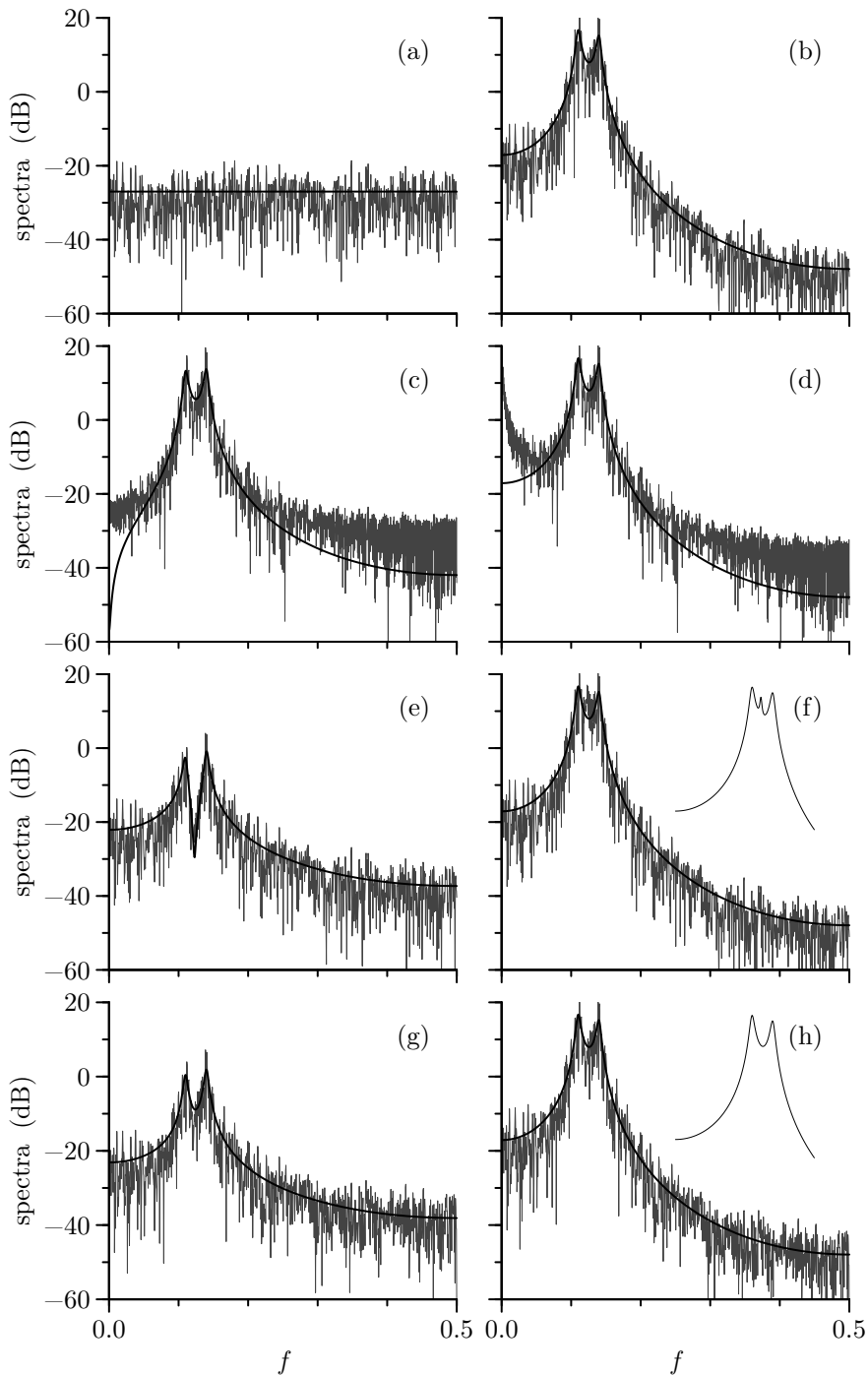
$$\hat{S}_X^{(PC)}(f) = \frac{\hat{S}_Y^{(P)}(f)}{\left| 1 - \sum_{l=1}^4 \phi_{4,l} e^{-i2\pi fl} \right|^2}, \quad (198)$$

along with the true AR(4) SDF  $S_X(\cdot)$ . Leakage has been eliminated at the expense of trivially shortening the time series by four values. (The fact that such a short filter prewhitens perfectly is obviously due to the AR(4) assumption and is an exception – rather than the rule – in practical applications.)

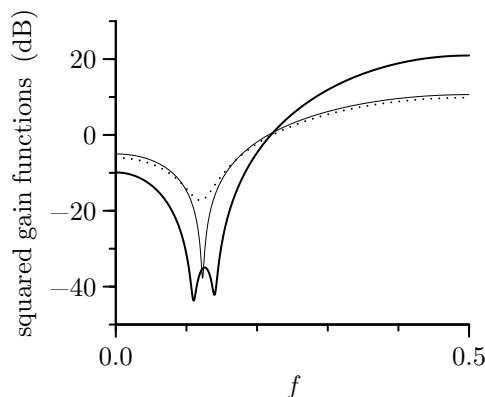
The bottom three rows of Figure 199 illustrate what can happen if we use a prewhitening filter that is less than perfect. The first of these is

$$Y_t = X_{t+1} - 0.99X_t,$$

which is a simple omnibus prewhitening filter that has proven to be effective in a number of applications (Tukey, 1967). This filter is close to the first difference filter  $\{g_0 = 1, g_1 = -1\}$  described in Exercise [5.4a], which is problematic as a prewhitening filter because postcoloring involves division by  $4 \sin^2(\pi f)$ , and this is zero when  $f = 0$ . Redefining  $g_1$  to be  $-0.99$  (or other values close to  $-1$  such as  $-0.98$ ) arguably improves the postcolored SDF estimate at frequencies close to zero. Figure 199(c) shows the true SDF  $S_Y(\cdot)$  for the prewhitened process (smooth curve) along with the periodogram based upon  $Y_0, Y_1, \dots, Y_{1022}$  (jagged curve).



**Figure 199.** Periodograms computed after prewhitening the AR(4) series of Figure 34(f) along with true SDFs for the prewhitened AR(4) process (left-hand column), and corresponding postcolored periodograms and SDF for the AR(4) process (right-hand). Each row uses a different prewhitening filter (see main text for details). The extra curves in (f) and (h) are the expected values of the postcolored periodograms at low frequencies (plotted versus  $f + 0.25$ ).



**Figure 200.** Square gain functions for the perfect prewhitening filter for the AR(4) process of Equation (35a) (thick solid curve), the filter used in Equation (200) (thin solid curve) and the filter used in Equation (201a) (dotted curve).

While this filter does reduce the dynamic range of the AR(4) SDF somewhat (from 65 down to 56 dB) if we consider just the frequencies between where the twin peaks are and 0.5, overall it pushes the dynamic range up to 71 dB due to the distortion it causes at low frequencies. In comparison to Figure 182(f), the postcolored estimate in Figure 199(d) displays somewhat less leakage at high frequencies, but considerably more at low frequencies. This “off-the-shelf” prewhitening filter is evidently best suited for SDFs with high power at low frequencies, which is not true for our AR(4) process. (Exercise [6.22] invites the reader to explore other filters of the form  $\{1, g_1\}$ .)

Motivated by the fact that the perfect prewhitening filter is a prediction error filter of width  $L = 5$ , let’s consider the following form as a candidate for prewhitening:

$$Y_t = X_{t+2} - \sum_{l=1}^2 g_l X_{t+2-l}.$$

We can again interpret the above as a prediction error filter, but the predictor of  $X_{t+2}$  is based upon just  $X_{t+1}$  and  $X_t$ . We can use the properties of the AR(4) process to set  $g_1$  and  $g_2$  such that, amongst all linear predictors of  $X_{t+2}$  that use just  $X_{t+1}$  and  $X_t$ , the prediction errors have a variance that is as small as possible (see Section 9.4 for details). The resulting prewhitening filter yields

$$Y_t \doteq X_{t+2} - 1.4198X_{t+1} + 0.9816X_t. \quad (200)$$

As shown in Figure 199(e), use of this filter yields an SDF  $S_Y(\cdot)$  whose dynamic range has been reduced to 36 dB. The periodogram based upon  $Y_0, Y_1, \dots, Y_{1021}$  overcomes the problem of leakage at high frequencies, but there is a fly in the ointment. A careful examination of the postcolored periodogram in Figure 199(f) reveals an elevation of power in the region between the twin peaks. This is not due just to sampling variability – the additional smooth curve in this plot shows the expected value of the postcolored periodogram at low frequencies  $f$ , but plotted versus  $f + 0.25$ . The new peak that emerges is caused by the fact that the prewhitening filter has created a region of low power between the twin peaks, and this region is prone to leakage. Thus, while we have successfully eliminated leakage at the high frequencies, we have *induced* leakage elsewhere! (If we use a direct spectral estimator with an  $NW = 1$  Slepian taper rather than the periodogram, the new peak disappears almost completely.)

An important clue as to why the prewhitening filter of Equation (200) gives mixed results is offered by a plot of its square gain function in Figure 200 (thin solid curve). The corresponding square gain function for the perfect prewhitening filter is shown by the thick solid curve. We see that the square gain function for the filter of Equation (200) has a pronounced dip in the midst of the range of frequencies trapping the twin peaks. As Figure 199(e) shows, this dip produces a deep valley in the prewhitened SDF  $S_Y(\cdot)$ , which is susceptible to leakage. A filter that has roughly the same square gain function but with a less pronounced dip is given by

$$Y_t = X_{t+2} - 1.3X_{t+1} + 0.8X_t. \quad (201a)$$

The square gain function for this prewhitening filter is shown as the dotted curve in Figure 200. The bottom row of Figure 199 shows that this filter works much better overall – there is no visual difference between the expected value of the postcolored periodogram and the true AR(4) SDF. This improvement comes about even though the dynamic range for  $S_Y(\cdot)$  has increased to 40 dB over the value of 36 dB obtained using the prewhitening filter of Equation (200).

These examples demonstrate the efficacy of prewhitening as a way of compensating for bias in the periodogram due to leakage, but they also show that care must be taken in choosing the prewhitening filter – an improper choice can actually increase the bias! On the other hand, a less than perfect prewhitening filter can work just as well as a perfect filter in eliminating leakage. The key requirement is that the filter needs to compensate for the general shape of the SDF for the time series under study. (Further examples and discussion of prewhitening are given in Section 9.10.)

### Comments and Extensions to Section 6.5

[1] There are some time series  $\{X_t\}$  for which we are interested in estimating the SDF  $S_X(\cdot)$  only over a selected range of frequencies, say,  $[f_L, f_H]$ . In this case we can subject  $\{X_t\}$  to a band-pass filter that attenuates all spectral components with frequencies  $|f| \notin [f_L, f_H]$  – this is sometimes known as *rejection filtration* (Blackman and Tukey, 1958, pp. 42–3). If  $S_Y(\cdot)$  is the SDF of the filtered time series and if the gain of the band-pass filter is approximately unity over the pass-band, then we should have  $S_Y(f) \approx S_X(f)$  for  $f_L \leq |f| \leq f_H$ . Because of the filtering operation, estimation of  $S_Y(f)$  over  $|f| \in [f_L, f_H]$  should be free of leakage from frequencies outside that range. This can be particularly valuable if there is a large contribution to the variance of  $\{X_t\}$  due to frequencies  $|f| \notin [f_L, f_H]$ . Alavi and Jenkins (1965) and Tsakiroglou and Walden (2002) give examples of rejection filtration, the former in the context of measurements of the error of a radar tracking a moving aircraft, and the latter in the context of the AR(4) process.

As with prewhitening, filtering of a time series to suppress certain frequencies decreases the number of data points available to estimate  $S_Y(\cdot)$ . Since we might have to make the filter width large to get a decent approximation to a band-pass filter, this procedure might not be worthwhile if leakage from frequencies  $|f| \notin [f_L, f_H]$  is small or can be controlled effectively with a data taper having a fairly narrow central lobe.

## 6.6 Statistical Properties of Direct Spectral Estimators

We investigated the first-moment (bias) properties of direct spectral estimators in Section 6.4. We now want to investigate some of their other statistical properties. We begin by considering the very special – but important – case of a Gaussian white noise process  $\{G_t\}$  with zero mean and variance  $\sigma^2$  (the letter  $G$  is used here to emphasize the Gaussian assumption).

Let

$$A(f) = \Delta_t^{1/2} \sum_{t=0}^{N-1} h_t G_t \cos(2\pi f t \Delta_t) \quad \text{and} \quad B(f) = \Delta_t^{1/2} \sum_{t=0}^{N-1} h_t G_t \sin(2\pi f t \Delta_t) \quad (201b)$$

be the real and imaginary parts of the complex conjugate of

$$J(f) \stackrel{\text{def}}{=} \Delta_t^{1/2} \sum_{t=0}^{N-1} h_t G_t e^{-i2\pi f t \Delta_t},$$

where  $\{h_t\}$  is a data taper. Since  $\{G_t\}$  has zero mean, it follows that

$$E\{A(f)\} = E\{B(f)\} = 0 \text{ for all } f.$$

From Exercise [2.1e] and the properties of a white noise process, we have

$$\begin{aligned} \text{cov}\{A(f), A(f')\} &= \sigma^2 \Delta_t \sum_{t=0}^{N-1} h_t^2 \cos(2\pi f t \Delta_t) \cos(2\pi f' t \Delta_t), \\ \text{cov}\{B(f), B(f')\} &= \sigma^2 \Delta_t \sum_{t=0}^{N-1} h_t^2 \sin(2\pi f t \Delta_t) \sin(2\pi f' t \Delta_t); \end{aligned}$$

and

$$\text{cov}\{A(f), B(f')\} = \sigma^2 \Delta_t \sum_{t=0}^{N-1} h_t^2 \cos(2\pi f t \Delta_t) \sin(2\pi f' t \Delta_t).$$

Letting  $f' = f$  in the first two equations above yields

$$\text{var}\{A(f)\} = \sigma^2 \Delta_t \sum_{t=0}^{N-1} h_t^2 \cos^2(2\pi f t \Delta_t), \quad \text{var}\{B(f)\} = \sigma^2 \Delta_t \sum_{t=0}^{N-1} h_t^2 \sin^2(2\pi f t \Delta_t).$$

The above equations simplify drastically if  $h_t = 1/\sqrt{N}$  for all  $t$  (i.e., the direct spectral estimator is in fact just the periodogram) and if  $f$  and  $f'$  are Fourier (or standard) frequencies  $f_k = k/(N \Delta_t)$  satisfying  $0 \leq f_k \leq f_N$ , where  $k$  is an integer. In this case we have

$$\text{cov}\{A(f_j), A(f_k)\} = \text{cov}\{B(f_j), B(f_k)\} = 0 \text{ for all } f_j \neq f_k; \quad (202a)$$

$$\text{cov}\{A(f_j), B(f_k)\} = 0 \text{ for all } f_j \text{ and } f_k; \quad (202b)$$

$$\text{var}\{A(f_k)\} = \begin{cases} \sigma^2 \Delta_t / 2, & f_k \neq 0 \text{ or } f_N; \\ \sigma^2 \Delta_t, & f_k = 0 \text{ or } f_N; \end{cases} \quad (202c)$$

and

$$\text{var}\{B(f_k)\} = \begin{cases} \sigma^2 \Delta_t / 2, & f_k \neq 0 \text{ or } f_N; \\ 0, & f_k = 0 \text{ or } f_N, \end{cases} \quad (202d)$$

all of which follow from the orthogonality relationships of Exercise [1.3].

Since  $A(f_k)$  and  $B(f_k)$  are linear combinations of Gaussian RVs, they in turn are Gaussian RVs. Because uncorrelatedness implies independence for Gaussian RVs, the  $A(f_k)$  and  $B(f_k)$  terms are zero mean independent Gaussian RVs with variances given by Equations (202c) and (202d). Recall that, if  $Y_0, Y_1, \dots, Y_{\nu-1}$  are independent Gaussian RVs with zero means and unit variances, then the RV

$$\chi_\nu^2 = Y_0^2 + Y_1^2 + \dots + Y_{\nu-1}^2$$

has a chi-square distribution with  $\nu$  degrees of freedom. When  $\nu = 2$ , this distribution is the same as an exponential distribution with a mean of 2 (see C&E [1] at the end of this section). Since

$$A^2(f_k) + B^2(f_k) = \hat{S}_G^{(\text{P})}(f_k),$$



i.e., the periodogram, and since for  $f_k \neq 0$  or  $f_N$

$$\left(\frac{2}{\sigma^2 \Delta_t}\right)^{1/2} A(f_k) \quad \text{and} \quad \left(\frac{2}{\sigma^2 \Delta_t}\right)^{1/2} B(f_k)$$

are independent Gaussian RVs with zero mean and unit variance, it follows that

$$\frac{2}{\sigma^2 \Delta_t} \hat{S}_G^{(P)}(f_k) \stackrel{d}{=} \chi_2^2; \quad \text{i.e.,} \quad \hat{S}_G^{(P)}(f_k) \stackrel{d}{=} \frac{\sigma^2 \Delta_t}{2} \chi_2^2 \quad \text{for } f_k \neq 0 \text{ or } f_N. \quad (203a)$$

Here  $\stackrel{d}{=}$  means “equal in distribution,” so the statement  $Y \stackrel{d}{=} c\chi_\nu^2$  means that the RV  $Y$  has the same distribution as a chi-square RV (with  $\nu$  degrees of freedom) after multiplication by a constant  $c$ . The corresponding result for  $f_k = 0$  or  $f_N$  is

$$\hat{S}_G^{(P)}(f_k) \stackrel{d}{=} \sigma^2 \Delta_t \chi_1^2.$$

Since  $S_G(f) = \sigma^2 \Delta_t$  is the SDF for  $\{G_t\}$ , we can rewrite the above as

$$\hat{S}_G^{(P)}(f_k) \stackrel{d}{=} \begin{cases} S_G(f_k) \chi_2^2 / 2, & f_k \neq 0 \text{ or } f_N; \\ S_G(f_k) \chi_1^2, & f_k = 0 \text{ or } f_N. \end{cases} \quad (203b)$$

Because  $E\{\chi_\nu^2\} = \nu$  and  $\text{var}\{\chi_\nu^2\} = 2\nu$ , it follows from the above that

$$E\{\hat{S}_G^{(P)}(f_k)\} = \sigma^2 \Delta_t = S_G(f_k) \quad \text{for all } f_k \quad (203c)$$

and that

$$\text{var}\{\hat{S}_G^{(P)}(f_k)\} = \begin{cases} \sigma^4 \Delta_t^2 = S_G^2(f_k), & f_k \neq 0 \text{ or } f_N; \\ 2\sigma^4 \Delta_t^2 = 2S_G^2(f_k), & f_k = 0 \text{ or } f_N \end{cases} \quad (203d)$$

(see Exercise [6.23] for an extension of the above to  $0 < f < f_N$ ). Moreover, all the RVs in the set  $\{\hat{S}_G^{(P)}(f_k) : 0 \leq k \leq \lfloor N/2 \rfloor\}$  are independent. Thus, the sampling properties of  $\hat{S}_G^{(P)}(f_k)$  on the grid of Fourier frequencies are completely known and have a simple form for a zero mean Gaussian white noise process  $\{G_t\}$ .

For a stationary process  $\{X_t\}$  that is not necessarily Gaussian with an SDF  $S(\cdot)$  that is not necessarily constant, it can be shown – subject to certain technical assumptions – that

$$\hat{S}^{(P)}(f) \stackrel{d}{=} \begin{cases} S(f) \chi_2^2 / 2, & 0 < |f| < f_N; \\ S(f) \chi_1^2, & |f| = 0 \text{ or } f_N, \end{cases} \quad (203e)$$

asymptotically as  $N \rightarrow \infty$  (see, for example, Brillinger, 1981a, section 5.2; Brockwell and Davis, 1991, section 10.3; Fuller, 1996, section 7.1; and Shumway and Stoffer, 2017, appendix C.2). Furthermore, for  $0 \leq |f'| < |f| \leq f_N$ ,  $\hat{S}^{(P)}(f)$  and  $\hat{S}^{(P)}(f')$  are asymptotically independent. We thus have, in addition to  $E\{\hat{S}^{(P)}(f)\} = S(f)$ , that

$$\text{var}\{\hat{S}^{(P)}(f)\} = \begin{cases} S^2(f), & 0 < |f| < f_N; \\ 2S^2(f), & f = 0 \text{ or } |f| = f_N; \end{cases} \quad (203f)$$

$$\text{cov}\{\hat{S}^{(P)}(f'), \hat{S}^{(P)}(f)\} = 0, \quad 0 \leq |f'| < |f| \leq f_N, \quad (203g)$$

asymptotically as  $N \rightarrow \infty$  (for finite sample sizes  $N$ , these asymptotic results are useful approximations if certain restrictions are observed – see C&E [3] for this section and Exercise [6.27f]). If we restrict ourselves to the nonnegative Fourier frequencies  $f_k$  and place certain constraints on  $\{X_t\}$  (see C&E [5]), we also find that the  $\lfloor N/2 \rfloor + 1$  RVs

$$\hat{S}^{(P)}(f_0), \hat{S}^{(P)}(f_1), \dots, \hat{S}^{(P)}(f_{\lfloor N/2 \rfloor})$$

are all approximately pairwise uncorrelated for  $N$  large enough; i.e.,

$$\text{cov} \{ \hat{S}^{(P)}(f_j), \hat{S}^{(P)}(f_k) \} \approx 0, \quad j \neq k \text{ and } 0 \leq j, k \leq \lfloor N/2 \rfloor. \quad (204a)$$

As we shall see, this last result plays an important role in the derivation of the sampling properties of spectral estimators based upon the periodogram (even though it is a large-sample result, it is often regarded as a reasonable approximation for finite  $N$  – see, for example, Bolt and Brillinger, 1979, and Brillinger, 1987).

Let us now turn our attention to the large sample theory for direct spectral estimators  $\hat{S}^{(D)}(\cdot)$  other than the periodogram (for details, see C&E [2]). For a particular sample size  $N$ , the variance of  $\hat{S}^{(D)}(f)$  can actually be larger or smaller than that of  $\hat{S}^{(P)}(f)$ , depending upon the particular data taper used and the characteristics of the true underlying SDF; however, for processes whose ACVS satisfies a summability condition and for a taper  $\{h_t\}$  that is reasonable in form, Brillinger (1981a, p. 127) concludes that the large-sample univariate distributional properties of  $\hat{S}^{(D)}(f)$  are the *same* as those of  $\hat{S}^{(P)}(f)$  – this implies that

$$\hat{S}^{(D)}(f) \stackrel{d}{=} \begin{cases} S(f)\chi_2^2/2, & 0 < |f| < f_N; \\ S(f)\chi_1^2, & |f| = 0 \text{ or } f_N, \end{cases} \quad (204b)$$

and

$$\text{var} \{ \hat{S}^{(D)}(f) \} = \begin{cases} S^2(f), & 0 < |f| < f_N; \\ 2S^2(f), & |f| = 0 \text{ or } f_N \end{cases} \quad (204c)$$

asymptotically as  $N \rightarrow \infty$  (see C&E [3] for caveats on applying these asymptotic results).

Equation (204b) allows us to construct an asymptotically correct confidence interval (CI) for  $S(f)$  at a fixed  $f$ . For  $0 < |f| < f_N$ , the large-sample distribution of  $\hat{S}^{(D)}(f)$  is related to the distribution of a chi-square RV  $\chi_2^2$  with two degrees of freedom, which has a PDF given by

$$f(u) = \begin{cases} \frac{1}{2}e^{-u/2}, & u \geq 0; \\ 0, & u < 0. \end{cases} \quad (204d)$$

Let  $Q_2(p)$  represent the  $p \times 100\%$  percentage point of this distribution; i.e.,

$$\mathbf{P} [\chi_2^2 \leq Q_2(p)] = \int_0^{Q_2(p)} f(u) du = 1 - e^{-Q_2(p)/2} = p \text{ so that } Q_2(p) = -2 \log(1 - p),$$

where  $\mathbf{P}[A]$  is the probability of the event  $A$  occurring. We thus have, for  $0 \leq p \leq 1/2$ ,

$$\mathbf{P} [Q_2(p) \leq \chi_2^2 \leq Q_2(1 - p)] = 1 - 2p.$$

When  $S(f) > 0$ , Equation (204b) says that  $2\hat{S}^{(D)}(f)/S(f)$  is a  $\chi_2^2$  RV. Hence

$$\mathbf{P} \left[ Q_2(p) \leq \frac{2\hat{S}^{(D)}(f)}{S(f)} \leq Q_2(1 - p) \right] = 1 - 2p.$$

The event in the brackets above occurs if and only if the events  $Q_2(p) \leq 2\hat{S}^{(D)}(f)/S(f)$  and  $2\hat{S}^{(D)}(f)/S(f) \leq Q_2(1 - p)$  both occur. These events in turn are equivalent to the single event in the brackets in the next equation, implying that

$$\mathbf{P} \left[ \frac{2\hat{S}^{(D)}(f)}{Q_2(1 - p)} \leq S(f) \leq \frac{2\hat{S}^{(D)}(f)}{Q_2(p)} \right] = 1 - 2p.$$

Hence

$$\left[ \frac{2\hat{S}^{(D)}(f)}{Q_2(1-p)}, \frac{2\hat{S}^{(D)}(f)}{Q_2(p)} \right] = \left[ -\frac{\hat{S}^{(D)}(f)}{\log(p)}, -\frac{\hat{S}^{(D)}(f)}{\log(1-p)} \right]$$

is a  $100(1-2p)\%$  CI for  $S(f)$ . Since  $0 < x < y$  implies that  $10 \log_{10}(x) < 10 \log_{10}(y)$ , we also have

$$\mathbf{P} \left[ 10 \log_{10} \left( \frac{2\hat{S}^{(D)}(f)}{Q_2(1-p)} \right) \leq 10 \log_{10}(S(f)) \leq 10 \log_{10} \left( \frac{2\hat{S}^{(D)}(f)}{Q_2(p)} \right) \right] = 1 - 2p,$$

and hence

$$\left[ 10 \log_{10}(\hat{S}^{(D)}(f)) + 10 \log_{10} \left( \frac{2}{Q_2(1-p)} \right), 10 \log_{10}(\hat{S}^{(D)}(f)) + 10 \log_{10} \left( \frac{2}{Q_2(p)} \right) \right]$$

is a  $100(1-2p)\%$  CI for  $10 \log_{10}(S(f))$ . The length of this CI is

$$10 \log_{10} \left( \frac{Q_2(1-p)}{Q_2(p)} \right) = 10 \log_{10} \left( \frac{\log(p)}{\log(1-p)} \right), \quad (205a)$$

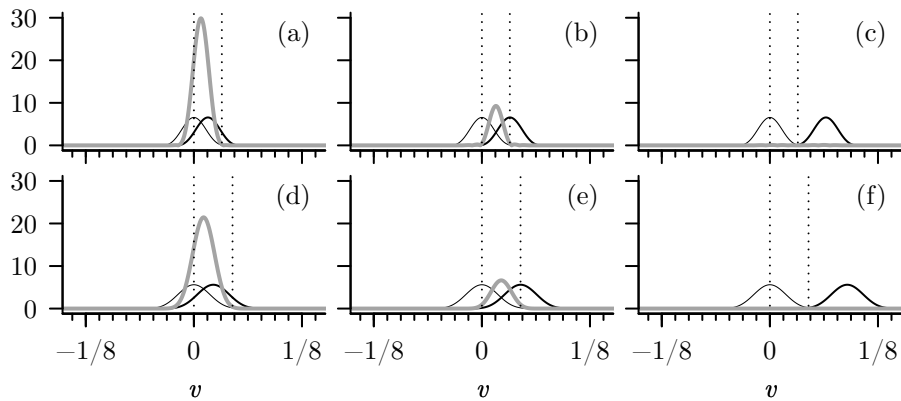
which is independent of the value of  $\hat{S}^{(D)}(f)$ , a fact we will put to good use shortly. For future reference we note that, if we set  $p = 0.025$  to obtain a 95% CI, the interval becomes

$$\left[ 10 \log_{10}(\hat{S}^{(D)}(f)) - 5.67, 10 \log_{10}(\hat{S}^{(D)}(f)) + 15.97 \right]. \quad (205b)$$

Its width is 21.63 dB, a result that holds for all direct spectral estimators, the periodogram included. (Note that the CIs presented here apply only to the SDF or its log at one particular frequency, and do not provide a simultaneous confidence band across all  $|f| \in (0, f_N)$  for either  $S(\cdot)$  or its log.)

While  $\hat{S}^{(P)}(f)$  and  $\hat{S}^{(D)}(f)$  have the same large-sample univariate distribution, the relationship between  $\hat{S}^{(D)}(f)$  and  $\hat{S}^{(D)}(f')$  at distinct frequencies  $f$  and  $f'$  is in general different from that between  $\hat{S}^{(P)}(f)$  and  $\hat{S}^{(P)}(f')$ . In particular, the grid of frequencies over which  $\text{cov}\{\hat{S}^{(D)}(f_j), \hat{S}^{(D)}(f_k)\} \approx 0$  is coarser than the one defined by the Fourier frequencies; thus, at adjacent Fourier frequencies  $f_k$  and  $f_{k+1}$ , the RVs  $\hat{S}^{(D)}(f_k)$  and  $\hat{S}^{(D)}(f_{k+1})$  can be significantly correlated. As a concrete example, the burden of Exercise [6.18b] is to show that  $\text{corr}\{\hat{S}_G^{(D)}(f_k), \hat{S}_G^{(D)}(f_{k+1})\} = 4/9$  for Gaussian white noise with  $\hat{S}_G^{(D)}(\cdot)$  based on a Hanning-like data taper (this result presumes that  $f_k$  is not too close to either zero or  $f_N$ ). More generally, in the case of a Gaussian process with SDF  $S_G(\cdot)$ , it follows from Equation (211c) in the C&Es along with an assumption that  $S_G(\cdot)$  is continuous on the interval  $[-f_N, f_N]$  that

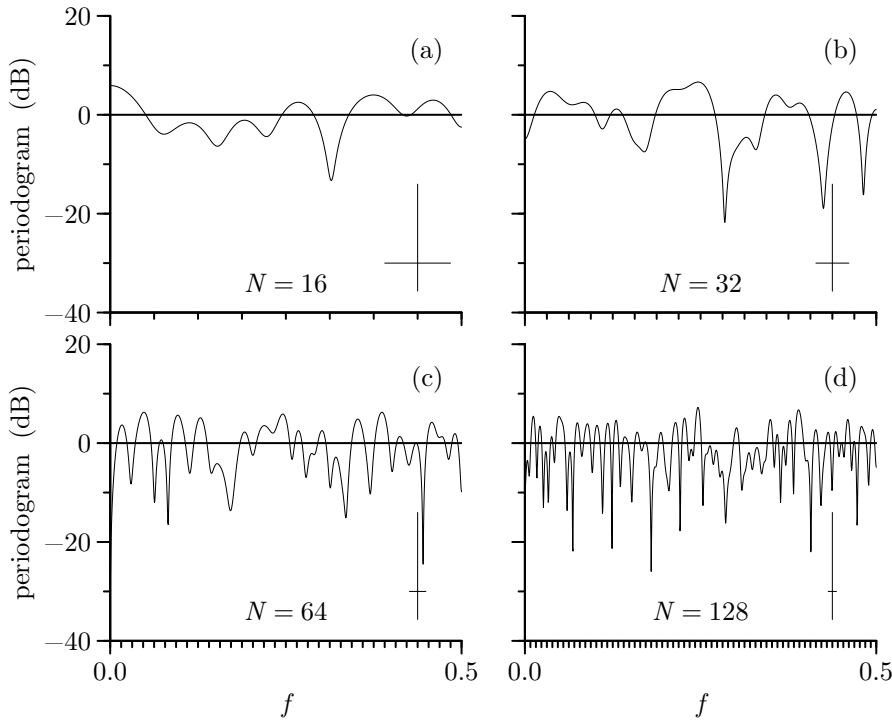
$$\begin{aligned} \text{cov}\{\hat{S}_G^{(D)}(f), \hat{S}_G^{(D)}(f')\} &\approx \frac{1}{\Delta_t^2} \left| \int_{-f_N}^{f_N} H^*(f-u)H(f'-u)S_G(u) du \right|^2 \\ &\leq \frac{S_{\max}^2}{\Delta_t^2} \left| \int_{-f_N}^{f_N} |H(f-u)| \cdot |H(f'-u)| du \right|^2 \\ &= \frac{S_{\max}^2}{\Delta_t^2} \left| \int_{-f_N}^{f_N} |H(f-f'+v)| \cdot |H(v)| dv \right|^2, \quad (205c) \end{aligned}$$



**Figure 206.** Examples of the integrand (thick gray curves) dictating the bound in Equation (205c) for the covariance between  $\hat{S}_G^{(D)}(f)$  and  $\hat{S}_G^{(D)}(f')$ . The integrand is the product of  $|H(\cdot)|$  and  $|H(f - f' + \cdot)|$  (the thin and thick dark curves), where  $H(\cdot)$  is the Fourier transform of a Slepian data taper for  $N = 64$  with  $W = 2/N$  (top row) and  $4/N$  (bottom). The distance between the dotted lines indicates the standard bandwidth measure  $B_{\mathcal{H}}$  of Equation (194) ( $B_{\mathcal{H}} \doteq 0.0323$  in the top row and  $0.0447$  in the bottom). In plots (a), (b) and (c) (and similarly in (d), (e) and (f)),  $f' - f$  is set to, respectively,  $B_{\mathcal{H}}/2$ ,  $B_{\mathcal{H}}$  and  $2B_{\mathcal{H}}$ . The spacing between the tick marks on the  $v$  axis is  $1/64$  (the spacing between the Fourier frequencies  $f_j = j/N$ ).

where  $v = f' - u$  and  $S_{\max} \stackrel{\text{def}}{=} \max_f S_G(f)$ . Suppose, for example, that  $H(\cdot)$  corresponds to a zeroth-order Slepian data taper. The above integrand will then be small whenever  $f$  and  $f'$  are far enough apart so that the central lobes of  $|H(f - f' + \cdot)|$  and  $|H(\cdot)|$  do not overlap significantly. Figure 206 illustrates the case  $N = 64$  with  $\Delta_t = 1$  and  $W$  set so that  $NW = 2$  (top row of plots) and  $NW = 4$  (bottom). The thin dark curve in the plots shows  $|H(v)|$  versus  $v$  while the thick dark curves show  $|H(f - f' + v)|$  versus  $v$  with  $f' - f$  set to, from left to right in each row,  $B_{\mathcal{H}}/2$ ,  $B_{\mathcal{H}}$  and  $2B_{\mathcal{H}}$ , where  $B_{\mathcal{H}}$  is our standard bandwidth measure for a direct spectral estimator (see Equation (194)). The distance between the vertical dotted lines in each plot is  $B_{\mathcal{H}}$ , while the minor tick marks depict the grid of Fourier frequencies. The thick gray curves show the integrands  $|H(f - f' + v)| \cdot |H(v)|$ . As  $|f - f'|$  increases, the area under the gray curve decreases, implying a lower upper bound for the covariance between  $\hat{S}_G^{(D)}(f)$  and  $\hat{S}_G^{(D)}(f')$  and hence a decrease in their correlation. If we specialize to white noise, we can interpret the square of the area under the gray curve as an approximation to  $\text{corr}\{\hat{S}_G^{(D)}(f), \hat{S}_G^{(D)}(f')\}$ . For the six cases shown in Figure 206, these approximations are (a) 0.43, (b) 0.03 and (c) 0.00 for the top row and, for the bottom, (d) 0.44, (e) 0.04 and (f) 0.00. Hence the correlation is quite small when  $|f - f'| \geq B_{\mathcal{H}}$ .

As discussed in C&Es [2] and [4], for a direct spectral estimator with an associated spectral window  $\mathcal{H}(\cdot)$ , the bandwidth measure  $B_{\mathcal{H}}$  quantifies the spacing  $|f - f'|$  required for  $\hat{S}^{(D)}(f)$  and  $\hat{S}^{(D)}(f')$  to be approximately uncorrelated. In particular, if  $0 \leq f < f' \leq f_N$  and  $|f - f'| \geq B_{\mathcal{H}}$ , then  $\text{corr}\{\hat{S}^{(D)}(f), \hat{S}^{(D)}(f')\} \approx 0$ , in agreement with the conclusion we drew from Figure 206. A grid of frequencies over which the direct spectral estimator  $\hat{S}^{(D)}(\cdot)$  is approximately pairwise uncorrelated is thus given by  $0, B_{\mathcal{H}}, 2B_{\mathcal{H}}, \dots, KB_{\mathcal{H}}$ , where  $K$  is the largest integer such that  $KB_{\mathcal{H}} \leq f_N$ . Alternatively, this grid can be expressed by  $kc\Delta_f = cf_k$ ,  $k = 0, \dots, K$ , where  $\Delta_f \stackrel{\text{def}}{=} 1/(N\Delta_t)$  is the distance between adjacent Fourier frequencies, and  $c = B_{\mathcal{H}}/\Delta_f$ . For the Hanning,  $NW = 2$  Slepian and  $NW = 4$  Slepian tapers, approximate values for  $c$  are, respectively, 2, 2 and 3 – these yield a grid size twice or three times as large as the Fourier frequencies (see the top row of Table 214, which is discussed in C&E [4]). As Brillinger (1981b) points out, although the grid size is now larger,



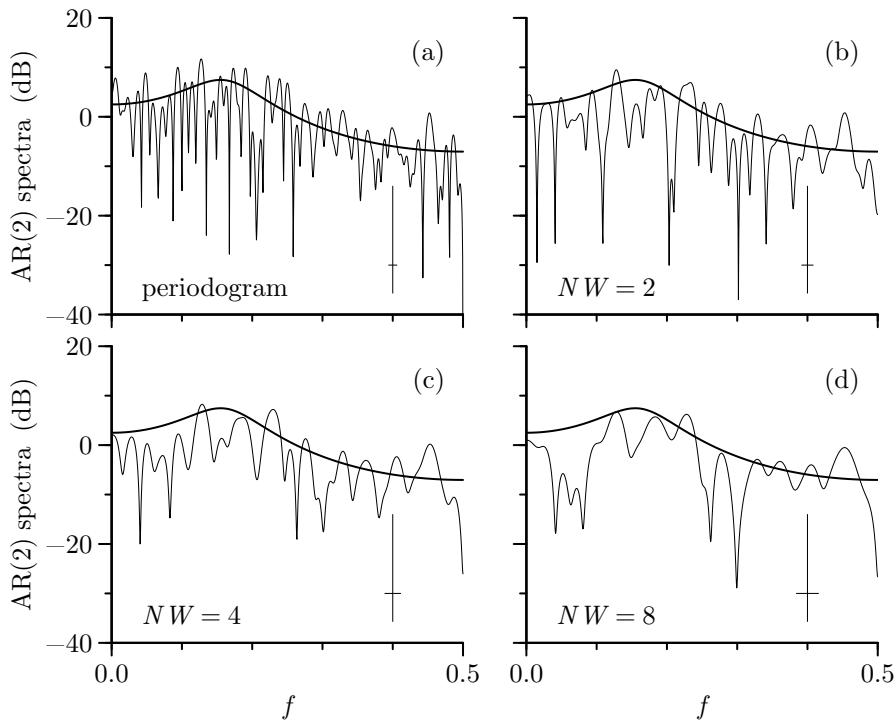
**Figure 207.** Inconsistency of the periodogram. The plots show the periodogram (on a decibel scale) versus frequency for samples of Gaussian white noise with  $N = 16, 32, 64$  and  $128$ . We take  $\Delta_t$  to be unity so the Nyquist frequency is  $f_N = 1/2$ . Each periodogram is evaluated over the grid of frequencies  $k/2048$ ,  $k = 0, 1, \dots, 1024$ . Tick marks on the horizontal axes indicate the locations of the Fourier frequencies. The horizontal lines at 0 dB show the true SDF. The widths of the crisscrosses depict the bandwidth measure  $B_{\mathcal{H}}$  of Equation (194), while their heights are the lengths of a 95% CI for  $10 \log_{10}(S(f))$  based upon  $10 \log_{10}(\hat{S}^{(P)}(f))$ .

the pairwise covariances of the  $K + 1$  RVs

$$\hat{S}^{(D)}(0), \hat{S}^{(D)}(c/(N \Delta_t)), \dots, \hat{S}^{(D)}(Kc/(N \Delta_t)) \quad (207)$$

are zero to a better approximation for finite  $N$  than the corresponding large-sample result stated in (204a) for the periodogram. This better approximation is useful for constructing certain statistical tests and is another benefit of tapering.

Since  $S(f) > 0$  typically, Equations (203f) and (204c) show that the variances of  $\hat{S}^{(P)}(f)$  and  $\hat{S}^{(D)}(f)$  do *not* decrease to 0 as  $N \rightarrow \infty$ . Thus the probability that  $\hat{S}^{(P)}(f)$  or  $\hat{S}^{(D)}(f)$  becomes arbitrarily close to its asymptotic expected value of  $S(f)$  is zero – in statistical terminology, neither  $\hat{S}^{(P)}(f)$  nor  $\hat{S}^{(D)}(f)$  is a consistent estimator of  $S(f)$ . Figure 207 illustrates the inconsistency of the periodogram for a Gaussian white noise process with unit variance and unit sampling interval  $\Delta_t$ : as the sample size increases from 16 to 128, the periodogram shows no sign of converging to the SDF for the process, namely,  $S(f) = 1$  (indicated by the horizontal lines at 0 dB). As  $N$  increases, there is an increase in the visual roughness of the periodogram, which is due to the decrease in covariance between  $\hat{S}^{(P)}(f)$  and  $\hat{S}^{(P)}(f')$  separated by a fixed distance  $|f - f'|$ . Inconsistency and the decrease in covariance are also reflected by the crisscrosses in the lower right-hand corners. In keeping with Equation (205b), the height of each crisscross at  $f = \frac{7}{16}$  would have depicted a 95% CI for  $S(\frac{7}{16})$  if  $10 \log_{10}(\hat{S}^{(P)}(\frac{7}{16}))$  had been equal to  $-30$  dB (never the case for the four periodograms in the figure). Mentally moving the center of a crisscross to an actual value for  $10 \log_{10}(\hat{S}^{(P)}(f))$  with  $0 < f < 0.5$



**Figure 208.** Variability of direct spectral estimators. The bumpy curve in plot (a) is the periodogram for a realization of size  $N = 128$  from the AR(2) process of Equation (34) (the realization itself is the first 128 points in Figure 34(a)). The corresponding curves in the remaining plots show direct spectral estimates based upon a Slepian data taper with  $NW$  set to (b) 2, (c) 4 and (d) 8. The thick curve in each plot is the true SDF (cf. Figure 177). All spectra are plotted on a decibel scale versus frequencies  $k/2048$ ,  $k = 0, 1, \dots, 1024$ . The crisscrosses depict the bandwidth measure  $B_{\mathcal{H}}$  (widths) and the length of a 95% CI for  $10 \log_{10}(S(f))$  (heights).

gives a 95% CI for  $S(f)$  because, as indicated by Equation (205b), the length of the CI does not depend on  $10 \log_{10}(\hat{S}^{(P)}(f))$  itself. The fact that the height of the crisscross remains the same as  $N$  increases is a manifestation of the inconsistency of the periodogram. On the other hand, the width of each crisscross is equal to the bandwidth measure  $B_{\mathcal{H}}$  of Equation (194). As  $N$  increases, this measure decreases, which indicates that the separation  $|f - f'|$  needed for  $\hat{S}^{(P)}(f)$  and  $\hat{S}^{(P)}(f')$  to be approximately uncorrelated is also decreasing.

Figure 208 shows four direct spectral estimates for the AR(2) process  $\{X_t\}$  defined in Equation (34). These estimates are based on the first  $N = 128$  points of the time series shown in Figure 34(a). The bumpy curve in plot (a) is the periodogram. Similar curves in the remaining plots are direct spectral estimates based upon zeroth-order Slepian data tapers with (b)  $NW = 2$ , (c)  $NW = 4$  and (d)  $NW = 8$ . The true SDF for  $\{X_t\}$  is shown by the smooth curve in each plot. Because the dynamic range of this SDF is small, the bias in the periodogram is already negligible for a sample size as small as 64 (see Figure 177). There is thus no need for a data taper here. The smoother appearance of the direct spectral estimates with the Slepian data tapers is due to the increased distance between uncorrelated spectral estimates. As before, we can quantify this distance using  $B_{\mathcal{H}}$ , which is shown by the width of the crisscross in each plot. A complementary interpretation is that tapering in this example has effectively reduced the size of our time series without any gain in bias reduction (since there really is no bias to be reduced!). Based upon the grid size of approximately uncorrelated spectral estimates, we can argue that the statistical properties of the Slepian-based direct

spectral estimators are approximately equivalent to those of a periodogram with  $128/(NW)$  data points, yielding 64, 32 and 16 values for cases (b), (c) and (d). The smoother appearance as  $NW$  increases can be attributed to this effective decrease in sample size. This effect is also evident in Figure 207 – as the sample size decreases, the periodogram has a smoother appearance. Indeed the degrees of smoothness in  $\hat{S}^{(p)}(\cdot)$  in plots (a) to (d) of Figure 207 are a fairly good visual match to the  $\hat{S}^{(D)}(\cdot)$  in plots (d) to (a) of Figure 208, indicating that the covariance structure across frequencies in the direct spectral estimates is controlled by the effective sample size  $128/(NW)$ . Finally we note that the local variability of all four direct spectral estimators in Figure 208 is about the same. This supports the claim that all direct spectral estimators of  $S(f)$  have approximately the same variance (see Equations (203f) and (204c)) and reflects the fact that, as indicated by the height of the crisscrosses and by Equation (205b), the lengths of the 95% CIs for  $10 \log_{10}(S(f))$  are the same for all four spectral estimates.

### Comments and Extensions to Section 6.6

[1] Equation (204b) says that, except at  $f = 0$  and  $f_N$ , a direct spectral estimator  $\hat{S}^{(D)}(f)$  approximately has the distribution of a scaled chi-square RV with two degrees of freedom, which is a special case of an exponential RV. Letting  $U \stackrel{\text{def}}{=} \hat{S}^{(D)}(f)$ , the PDF for this distribution is given by

$$f_U(u; S(f)) = \begin{cases} \frac{1}{S(f)} e^{-u/S(f)}, & u \geq 0; \\ 0, & u < 0. \end{cases} \quad (209a)$$

This function is shown in Figure 210a(a) for the case  $S(f) = 2$ , for which  $U$  becomes an ordinary (i.e., unscaled) chi-square RV (see Equation (204d)). Since this PDF has an upper tail that decays at a slower rate than that of a Gaussian PDF, a random sample of variates with this distribution will typically contain “upshoots,” i.e., a few values that appear to be unusually large compared with the bulk of the variates (see Figure 210b(a); cf. Figure 172); however, when direct spectral estimates are plotted on a decibel (or logarithmic) scale, the visual appearance is quite different – we typically see prominent “downshoots” instead (see Figure 210b(b); cf. Figures 173, 182 and 187). The explanation for this “paradox” is that we are actually examining variates with the PDF of  $V \stackrel{\text{def}}{=} 10 \log_{10}(U)$ . This PDF is given by

$$f_V(v; S(f)) = \frac{\log(10)}{10 S(f)} 10^{v/10} e^{-(10^{v/10})/S(f)}, \quad -\infty < v < \infty, \quad (209b)$$

and is shown Figure 210a(b), again for the case  $S(f) = 2$ . Note that, in contrast to  $f_U(\cdot)$ , this PDF has a prominent *lower* tail. Note also that

$$f_V(v; S(f)) = f_V(v - 10 \log_{10}(S(f)/2); 2).$$

Thus if  $S(f)$  is something other than 2, the PDF depicted in 210a(b) merely shifts along the  $v$ -axis. This result implies that the variance (and other central moments) of  $10 \log_{10}(\hat{S}^{(D)}(f))$  does not depend on  $S(f)$ . Another way to establish this result is to note that the assumption  $\hat{S}^{(D)}(f) \stackrel{d}{=} S(f)\chi_2^2/2$  implies that

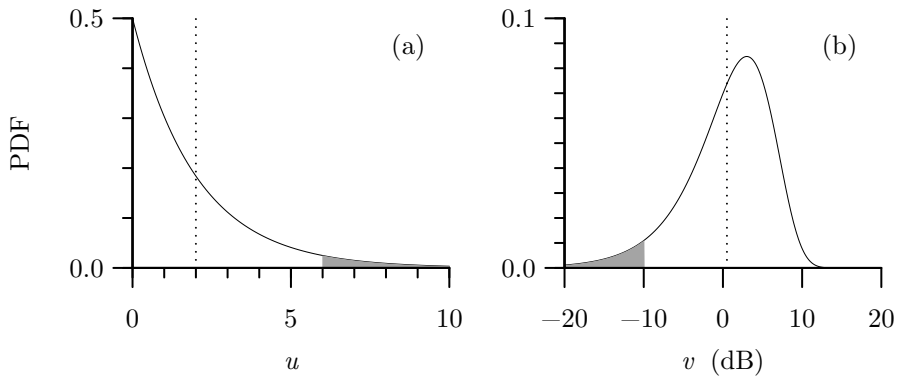
$$10 \log_{10}(\hat{S}^{(D)}(f)) \stackrel{d}{=} 10 \log_{10}(S(f)/2) + 10 \log_{10}(\chi_2^2)$$

and hence that, whereas  $\text{var}\{\hat{S}^{(D)}(f)\} = S^2(f)$ ,

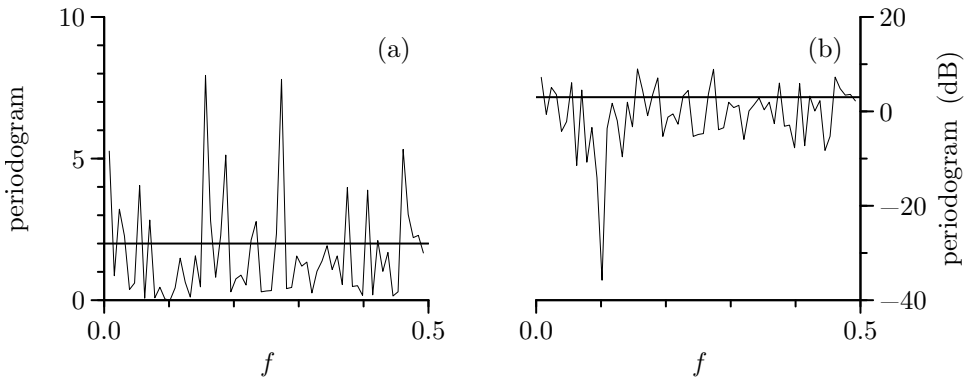
$$\text{var}\{10 \log_{10}(\hat{S}^{(D)}(f))\} = \text{var}\{10 \log_{10}(\chi_2^2)\}.$$

Taking the log of  $\hat{S}^{(D)}(f)$  acts as a variance-stabilizing transform in that the dependence of the variance on  $S(f)$  is removed.

[2] For use in Section 6.3, we develop here a more accurate expression for the covariance of direct spectral estimators than given by the qualitative “grid size” argument surrounding Equation (207) (Jones,



**Figure 210a.** Probability density functions for (a) a chi-square RV with two degrees of freedom and (b)  $10 \log_{10}$  of the same RV. The shaded areas depict the upper 5% tail area in (a) and the lower 5% tail area in (b). The vertical dotted lines indicate the expected values of RVs with these PDFs, i.e., 2 in (a) and  $10 \log_{10}(2/e^\gamma)$  in (b), where  $\gamma \doteq 0.5772$  is Euler's constant (this can be obtained from the digamma function  $\psi(\cdot)$  since  $\psi(1) = -\gamma$ ).



**Figure 210b.** Periodogram  $\hat{S}_G^{(P)}(\cdot)$  at Fourier frequencies  $f_k = k/128, k = 1, 2, \dots, 63$ , for a time series of length  $N = 128$  drawn from a Gaussian white noise process with mean  $\mu = 0$  and variance  $\sigma^2 = 2$  and plotted on (a) a linear scale and (b) a decibel scale. The variates  $\hat{S}_G^{(P)}(f_k)$  in (a) can be regarded as a random sample of 63 independent and identically distributed RVs from a chi-square distribution with two degrees of freedom. The PDF for this distribution is shown in Figure 210a(a). The variates  $10 \log_{10}(\hat{S}_G^{(P)}(f_k))$  in (b) constitute a random sample from a distribution whose PDF is shown in Figure 210a(b). The horizontal lines in the plots depict the SDF  $S_G(f) = 2$  for the white noise process in (a) and  $10 \log_{10}(S_G(f)) = 10 \log_{10}(2) \doteq 3.01$  in (b).

1971; Koopmans, 1974, Chapter 8; and Walden, 1990a). Suppose that  $\{G_t\}$  is a stationary Gaussian process – not necessarily white noise – with zero mean and SDF  $S_G(\cdot)$  and that, as before,

$$J(f) \stackrel{\text{def}}{=} \Delta_t^{1/2} \sum_{t=0}^{N-1} h_t G_t e^{-i2\pi f t \Delta_t} \quad \text{with } \{h_t\} \longleftrightarrow H(\cdot).$$

Now,

$$\text{cov}\{|J(f')|^2, |J(f)|^2\} = \text{cov}\{J(f')J^*(f'), J(f)J^*(f)\}.$$

Since  $\{G_t\}$  is a Gaussian process, the distribution of  $J(f)$  is complex-valued Gaussian for all  $f$  (for non-Gaussian processes, the distribution of  $J(f)$  can still be reasonably approximated as such, but this depends somewhat on the effect of the taper). We can now use the Isserlis theorem (Equation (30)) to



show that

$$\begin{aligned} \text{cov} \{ |J(f')|^2, |J(f)|^2 \} &= \text{cov} \{ J(f'), J(f) \} \text{cov} \{ J^*(f'), J^*(f) \} \\ &\quad + \text{cov} \{ J(f'), J^*(f) \} \text{cov} \{ J^*(f'), J(f) \} \\ &= E \{ J(f') J^*(f) \} E \{ J^*(f') J(f) \} \\ &\quad + E \{ J(f') J(f) \} E \{ J^*(f') J^*(f) \} \\ &= |E \{ J(f') J^*(f) \}|^2 + |E \{ J(f') J(f) \}|^2. \end{aligned} \quad (211a)$$

From the fact that

$$J(f) = \frac{1}{\Delta_t^{1/2}} \int_{-f_N}^{f_N} H(f - u) dZ(u)$$

(this is Equation (186c)) and the properties of  $dZ(\cdot)$ , we have

$$E \{ J(f') J^*(f) \} = \frac{1}{\Delta_t} \int_{-f_N}^{f_N} H(f' - u) H^*(f - u) S_G(u) du.$$

Since  $dZ(-u) = dZ^*(u)$  (see Equation (109b)), it is also true that

$$J(f) = -\frac{1}{\Delta_t^{1/2}} \int_{-f_N}^{f_N} H(f + u) dZ^*(u),$$

and hence

$$E \{ J(f') J(f) \} = -\frac{1}{\Delta_t} \int_{-f_N}^{f_N} H(f' + u) H(f - u) S_G(u) du.$$

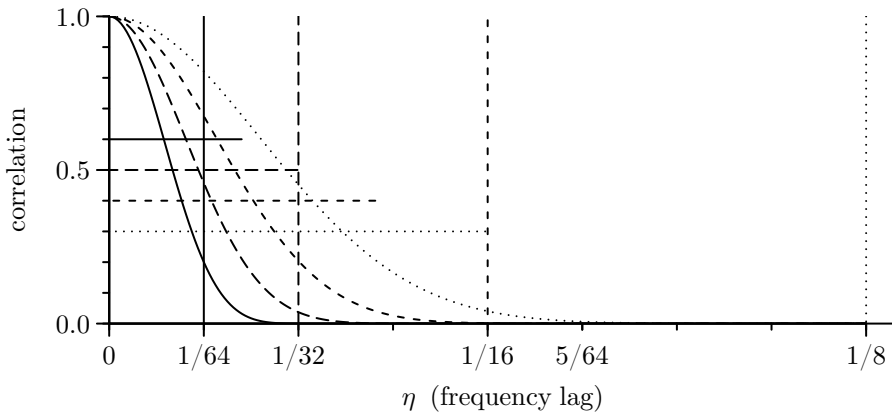
With these expressions and by substituting  $f + \eta$  for  $f'$ , we can now write

$$\begin{aligned} \text{cov} \{ \hat{S}_G^{(D)}(f + \eta), \hat{S}_G^{(D)}(f) \} &= \text{cov} \{ |J(f + \eta)|^2, |J(f)|^2 \} \\ &= \frac{1}{\Delta_t^2} \left| \int_{-f_N}^{f_N} H^*(f + \eta - u) H(f - u) S_G(u) du \right|^2 \\ &\quad + \frac{1}{\Delta_t^2} \left| \int_{-f_N}^{f_N} H(f + \eta + u) H(f - u) S_G(u) du \right|^2. \end{aligned} \quad (211b)$$

Although this result is exact for Gaussian processes (and, as is shown in Exercise [6.28a], can be computed using the inverse Fourier transforms of  $H(\cdot)$  and  $S_G(\cdot)$ ), it is unwieldy in practice, so we seek an approximation to it. Thomson (1977) points out that the second integral in Equation (211b) is large only for  $f$  near 0 and  $f_N$  – we exclude these cases here – so that the first integral is usually dominant. Suppose the central lobe of  $H(\cdot)$  goes from  $-W$  to  $W$  approximately. If the frequency separation  $\eta$  is less than  $2W$  so that the central lobes of the shifted versions of  $H(\cdot)$  in the first term overlap (cf. Figure 206(a) and (b)), then the covariance of Equation (211b) depends primarily on the SDF in the domain  $[f + \eta - W, f + W]$ . If we take the SDF to be locally constant about  $f$ , we obtain

$$\begin{aligned} \text{cov} \{ \hat{S}_G^{(D)}(f + \eta), \hat{S}_G^{(D)}(f) \} &\approx \frac{1}{\Delta_t^2} \left| \int_{-f_N}^{f_N} H^*(f + \eta - u) H(f - u) S_G(u) du \right|^2 \\ &\approx \frac{S_G^2(f)}{\Delta_t^2} \left| \int_{-f_N}^{f_N} H(u) H(\eta - u) du \right|^2 = \frac{S_G^2(f)}{\Delta_t^2} |H * H(\eta)|^2, \end{aligned} \quad (211c)$$

where  $*$  denotes convolution as defined by Equation (99b) – here we have made use of the facts that  $H(\cdot)$  is periodic with period  $2f_N = 1/\Delta_t$  and that  $H^*(-u) = H(u)$ .



**Figure 212.** Correlation of direct spectral estimators. The frequency domain correlation  $R(\eta)$  is plotted versus the frequency lag  $\eta$  for the four Slepian data tapers shown in the right-hand column of Figure 190 (recall that  $N = 64$  and  $\Delta_t = 1$ ). The cases  $NW = 1, 2, 4$  and  $8$  are shown, respectively, by the solid, long dashed, short dashed and dotted curves; the corresponding vertical lines indicate the location of  $W$  for these four cases ( $1/64, 1/32, 1/16$  and  $1/8$  cycles per unit time); and the horizontal lines depict the bandwidth measure width  $\{R(\cdot)\}$  defined by Equation (214a).

▷ **Exercise [212]** Show that the above covariance can be written in terms of the (real-valued) data taper as

$$\text{cov} \{ \hat{S}_G^{(D)}(f + \eta), \hat{S}_G^{(D)}(f) \} \approx S_G^2(f) \left| \sum_{t=0}^{N-1} h_t^2 e^{-i2\pi\eta t \Delta_t} \right|^2 \stackrel{\text{def}}{=} R(\eta, f). \quad \triangleleft$$

If we let  $\eta = 0$ , we obtain  $R(0, f) = S_G^2(f)$  because of the normalization  $\sum_{t=0}^{N-1} h_t^2 = 1$ . This implies  $\text{var} \{ \hat{S}_G^{(D)}(f) \} \approx R(0, f) = S_G^2(f)$ , in agreement with our earlier quoted result for  $|f| \neq 0$  or  $\pm f_N$  (see Equation (204c)). Under our working assumption that  $S_G(f + \eta) \approx S_G(f)$ , we have  $\text{var} \{ \hat{S}_G^{(D)}(f + \eta) \} \approx \text{var} \{ \hat{S}_G^{(D)}(f) \} \approx R(0, f)$ . It follows that the correlation between  $\hat{S}_G^{(D)}(f + \eta)$  and  $\hat{S}_G^{(D)}(f)$  is given approximately by

$$R(\eta) \stackrel{\text{def}}{=} \frac{R(\eta, f)}{R(0, f)} = \left| \sum_{t=0}^{N-1} h_t^2 e^{-i2\pi\eta t \Delta_t} \right|^2, \quad (212a)$$

which depends only on the data taper and not upon either  $f$  or  $S_G(\cdot)$  (Thomson, 1982).

The value  $R(\eta)$  thus gives us an approximation to the correlation between  $\hat{S}_G^{(D)}(f + \eta)$  and  $\hat{S}_G^{(D)}(f)$  under the assumption that  $S_G(\cdot)$  is constant in the interval  $[f + \eta - W, f + \eta + W]$ , where  $W$  is the half-width of the central lobe of the spectral window  $\mathcal{H}(\cdot)$  associated with  $\hat{S}_G^{(D)}(\cdot)$ . Figure 212 shows  $R(\cdot)$  for the four Slepian data tapers shown in the right-hand column of Figure 190 (here  $N = 64$  and  $\Delta_t = 1$ ). The grid of the Fourier frequencies  $f_k = k/N$  for this example has a spacing (or grid size) of  $1/N = 1/64$ . The tick marks on the  $\eta$ -axis show the grid over  $k = 0, \dots, 8$ . For a white noise process – a case where there is no bias in the periodogram so tapering is not needed – values of the periodogram at Fourier frequencies  $f_k = k/N$  and  $f_{k+1} = (k+1)/N$  are uncorrelated, whereas the corresponding values for a direct spectral estimator using a Slepian data taper with  $NW = K$  are positively correlated, with the correlation increasing as  $K$  increases (this assumes  $f_k$  and  $f_{k+1}$  are not too close to either 0 or  $f_N = 1/2$ ).

Since  $R(\cdot)$  is a  $2f_N$  periodic function (see Equation (212a)), its inverse Fourier transform is a sequence, say,  $\{r_\tau\}$ :

$$\{r_\tau\} \longleftrightarrow R(\cdot). \quad (212b)$$

To obtain an expression for  $r_\tau$  in terms of  $\{h_t\}$ , recall that  $\{\hat{s}_\tau^{(D)}\} \longleftrightarrow \hat{S}^{(D)}(\cdot)$  and note that, if we replace  $X_t$  by  $h_t/\Delta_t^{1/2}$  in Equation (186b) for  $\hat{S}^{(D)}(\cdot)$ , the latter reduces to  $R(\cdot)$ . With this substitution,

Equation (188b) tells us that

$$r_\tau = \begin{cases} (1/\Delta_t) \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|}^2 h_t^2, & |\tau| \leq N-1; \\ 0, & |\tau| \geq N. \end{cases} \quad (213a)$$

We note three facts: (1) the analog of Equation (188a) yields

$$R(\eta) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} r_\tau e^{-i2\pi\eta\tau\Delta_t}; \quad (213b)$$

(2) since  $\{r_\tau\}$  is a “rescaled ACVS” for  $\{h_t^2\}$ , it can be calculated efficiently using techniques discussed in Section 6.7; and (3) since Equation (212a) tells us that  $R(0) = 1$ , we have

$$\Delta_t \sum_{\tau=-(N-1)}^{N-1} r_\tau = 1.$$

In addition, under our working assumption that  $S_G(\cdot)$  is locally constant,  $\hat{S}_G^{(D)}(f)$  and  $\hat{S}_G^{(D)}(f)/S_G(f)$  differ only by a constant in the interval  $[f + \eta - W, f + W]$ . Since

$$R(\eta) \approx \text{cov} \{ \hat{S}_G^{(D)}(f + \eta)/S_G(f), \hat{S}_G^{(D)}(f)/S_G(f) \}$$

(i.e., the direct spectral estimator is locally stationary), it follows that the “spectrum” of the spectral estimator  $\hat{S}_G^{(D)}(\cdot)$  is given by  $\{S_G^2(f)r_\tau\}$ , where  $\tau$  is a pseudo-lag variable. Setting  $\eta = 0$  in Equation (213b) yields

$$R(0) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} r_\tau \text{ and hence } S_G^2(f)R(0) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} S_G^2(f)r_\tau,$$

which is a relationship analogous to the usual one between an ACVS and its SDF at zero frequency (see Equation (113d)).

[3] For the practitioner, the effect of the restriction in the preceding discussion to frequencies “not too close to 0 or Nyquist” is that – for *finite* sample sizes  $N$  from a process with zero mean – the  $\chi_2^2$ -based distribution for the periodogram in Equation (203e) is only appropriate for  $1/(N\Delta_t) \leq |f| \leq f_N - 1/(N\Delta_t)$ . Likewise, the  $\chi_2^2$  result in Equation (204b) for general direct spectral estimators  $\hat{S}^{(D)}(f)$  is reasonable for  $B_H/2 \leq |f| \leq f_N - B_H/2$  or, more conservatively,  $B_H \leq |f| \leq f_N - B_H$ , where  $B_H$  is the effective bandwidth of  $\hat{S}^{(D)}(\cdot)$  (see Exercise [6.30] for a specific example).

Another warning about Equations (203e) and (204b) is that the  $\chi_1^2$  result for  $f = 0$  is no longer valid if we center the time series by subtracting the sample mean from each observation prior to forming the periodogram or other direct spectral estimators, as would be called for in practical applications where typically we do not know the process mean (again, see Exercise [6.30] for an example). In particular, as we noted in C&E [3] for Section 6.3, centering a time series forces the value of the periodogram at  $f = 0$  to be identically equal to zero, indicating that the result  $\hat{S}^{(P)}(0) \stackrel{d}{=} S(0)\chi_1^2$  cannot hold (with a possible exception if  $S(0) = 0$ ).

[4] The bandwidth of an SDF estimator is sometimes regarded as a measure of the minimal separation in frequency between approximately uncorrelated SDF estimates (Priestley, 1981, p. 519; Walden et al., 1995). Since, in the Gaussian case, the correlation between  $\hat{S}^{(D)}(f + \eta)$  and  $\hat{S}^{(D)}(f)$  is approximated by  $R(\eta)$  of Equation (212a), we need to determine the point at which  $R(\cdot)$  can be regarded as being close to zero (see Figure 212). A convenient measure of this point is  $\text{width}_e \{R(\cdot)\}$ , the equivalent width of  $R(\cdot)$ . This measure is best suited for a function that is real-valued, positive everywhere, peaked about 0 and continuous at 0. The function  $R(\cdot)$  satisfies these conditions. From Equation (100d), we have

$$\text{width}_e \{R(\cdot)\} = \int_{-f_N}^{f_N} R(\eta) d\eta / R(0) = \int_{-f_N}^{f_N} R(\eta) d\eta$$

	$p \times 100\%$ Cosine Taper					Slepian Taper				
	$p =$	0.0	0.2	0.5	1.0	$NW =$	1	2	4	8
$B_{\mathcal{H}}/\Delta_{\text{f}}$		1.50	1.56	1.72	2.06		1.59	2.07	2.86	4.01
$\text{width}_{\text{e}}\{R(\cdot)\}/\Delta_{\text{f}}$		1.00	1.11	1.35	1.93		1.40	1.99	2.81	3.97
$B_{\mathcal{H}}/\text{width}_{\text{e}}\{R(\cdot)\}$		1.50	1.41	1.27	1.06		1.13	1.04	1.02	1.01

**Table 214.** Comparison of two bandwidth measures for direct spectral estimators  $\hat{S}^{(D)}(\cdot)$ . The first measure  $B_{\mathcal{H}}$  is given by Equation (194) and is the autocorrelation width of the spectral window associated with  $\hat{S}^{(D)}(\cdot)$ . The second measure  $\text{width}_e \{R(\cdot)\}$  is given by Equation (214a) and is essentially the minimal distance  $\eta$  such that  $\hat{S}^{(D)}(f + \eta)$  and  $\hat{S}^{(D)}(f)$  are approximately uncorrelated (assuming that the process is Gaussian and that  $f$  is not too close to either zero or Nyquist frequency). The tabulated values show the ratio of these measures to either  $\Delta_f$  (first two rows) or one another (third row) for the eight data tapers of length  $N = 64$  shown in Figure 190, where  $\Delta_f = 1/(N \Delta_t)$  is the spacing between adjacent Fourier frequencies  $f_k = k/(N \Delta_t)$ .

because  $R(0) = 1$ . The right-hand side is related to the notion of *correlation time*, which, in its time-domain formulation, measures the time needed for any correlation in a stationary process to die out (Yaglom, 1987a, p. 113; other names for this notion are *decorrelation time* and *integral time scale*). Since  $\{r_\tau\} \longleftrightarrow R(\cdot)$  implies that

$$r_\tau = \int_{-f_N}^{f_N} R(\eta) e^{i2\pi\eta\tau \Delta_t} d\eta \text{ and hence } r_0 = \int_{-f_N}^{f_N} R(\eta) d\eta,$$

we find that, in combination with Equation (213a),

$$\text{width}_e \{R(\cdot)\} = r_0 = \frac{1}{\Delta_t} \sum_{t=0}^{N-1} h_t^4, \tag{214a}$$

which is very easy to calculate. The horizontal lines on Figure 212 show  $\text{width}_e \{R(\cdot)\}$  for four Slepian data tapers.

It is of interest to compare  $\text{width}_e \{R(\cdot)\}$  with our standard measure of the effective bandwidth of a direct spectral estimator, namely,  $B_{\mathcal{H}}$  of Equation (194), where  $\mathcal{H}(\cdot)$  is the spectral window associated with  $\hat{S}^{(D)}(\cdot)$ . The first line of Table 214 shows the ratio of  $B_{\mathcal{H}}$  to  $\Delta_f$  for the eight data tapers of length  $N = 64$  shown in Figure 190, where  $\Delta_f = 1/(N \Delta_t)$  is the distance between adjacent Fourier frequencies. The second and third lines show the ratios  $\text{width}_e \{R(\cdot)\} / \Delta_f$  and  $B_{\mathcal{H}}/\text{width}_e \{R(\cdot)\}$ . In all cases  $B_{\mathcal{H}}$  is greater than  $\text{width}_e \{R(\cdot)\}$ , implying our standard measure is such that, if  $0 < f < f' < f_N$  and  $|f - f'| \geq B_{\mathcal{H}}$ , then  $\text{corr} \{ \hat{S}^{(D)}(f), \hat{S}^{(D)}(f') \} \approx 0$ . The difference between the two measures is greatest for the default (rectangular) data taper and decreases as the degree of tapering increases, i.e., as either  $p$  or  $NW$  increases. (The values in the table remain virtually the same if we redo the computations using  $N = 1024$  rather than  $N = 64$ .)

[5] The result stated in Equation (204a), namely, that the periodogram is asymptotically uncorrelated over the grid of Fourier frequencies, needs some clarification. A firm basis for this result is Corollary 7.2.1 of Fuller (1996), from which we can deduce that, if  $\{X_t\}$  can be expressed as

$$X_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} \text{ with } \sum_{j=1}^{\infty} j|\theta_j| < \infty, \tag{214b}$$

where  $\{\epsilon_t\}$  is a Gaussian white noise process with zero mean and finite variance, then

$$|\text{cov} \{ \hat{S}^{(P)}(f_j), \hat{S}^{(P)}(f_k) \}| \leq \frac{C}{N^2},$$

where  $C$  is a finite constant independent of  $f_j$  or  $f_k$ . The conditions for this result to hold are restrictive, but reasonable to assume in certain practical applications. For example, a Gaussian AR(1) process  $X_t = \phi X_{t-1} + \epsilon_t$  with  $|\phi| < 1$  can be represented as above with  $\theta_j = \phi^j$  (see Equation (44b)). This stationary process is commonly used as a simple model for climate time series, in part due to its interpretation as a discretized first-order differential equation (von Storch and Zwiers, 1999).

▷ **Exercise [215]** Show that the summation condition on  $\{\theta_j\}$  holds here since

$$\sum_{j=1}^{\infty} j|\theta_j| = \frac{|\phi|}{(1-|\phi|)^2}. \quad \triangleleft$$

If some of the stipulations surrounding Equation (214b) cannot reasonably be assumed to hold, subtle issues arise. For example, if  $\{X_t\}$  can be expressed as

$$X_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} \quad \text{with} \quad \sum_{j=1}^{\infty} |\theta_j| < \infty,$$

where  $\{\epsilon_t\}$  are now independent and identically distributed RVs with zero mean and finite variance, and if the SDF for  $\{X_t\}$  is positive at all frequencies, then Theorem 7.1.2 of Fuller (1996) implies that, for a finite set of fixed frequencies, say  $0 < f_{(1)} < f_{(2)} < \dots < f_{(J)} < f_N$ , the RVs  $\hat{S}^{(P)}(f_{(1)})$ ,  $\hat{S}^{(P)}(f_{(2)})$ ,  $\dots$ ,  $\hat{S}^{(P)}(f_{(J)})$ , are asymptotically independent with distributions dictated by  $S(f_{(j)})\chi^2_{2}/2$ ,  $j = 1, \dots, J$ . This same result follows from Theorem 5.2.6 of Brillinger (1981a), but now under the assumption that  $\{X_t\}$  is a completely stationary process with all of its moments existing and with its cumulants of all orders being absolutely summable (cumulants can be constructed directly from moments). Unfortunately this result is sometimes misinterpreted as being valid over the grid of Fourier frequencies (see the discussion by Anderson of Diggle and al Wasel, 1997).

For additional discussion on the large-sample distribution of the periodogram and other direct spectral estimators under various assumptions, see Fay and Soulier (2001), Kokoszka and Mikosch (2000) and Lahiri (2003a).

[6] It is sometimes of interest to test the hypothesis that a particular time series can be regarded as a segment  $X_0, \dots, X_{N-1}$  of a realization of a white noise process (see Section 10.15). One of the most popular tests for white noise is known as the *cumulative periodogram test*. This test is based upon the periodogram over the set of Fourier frequencies  $f_k = k/(N \Delta_t)$  such that  $0 < f_k < f_N$ , i.e.,  $f_1, \dots, f_M$  with  $M = \lfloor (N-1)/2 \rfloor$ . As noted in the discussion surrounding Equation (203b), under the null hypothesis that  $\{X_t\}$  is a Gaussian white noise process,  $\hat{S}^{(P)}(f_1), \dots, \hat{S}^{(P)}(f_M)$  constitute a set of independent and identically distributed RVs having a scaled chi-square distribution with two degrees of freedom, i.e., an exponential distribution. If we form the normalized cumulative periodogram

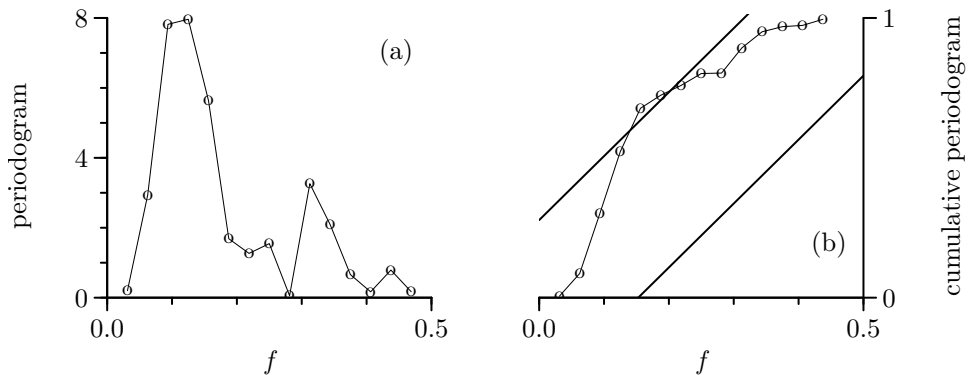
$$\mathcal{P}_k \stackrel{\text{def}}{=} \frac{\sum_{j=1}^k \hat{S}^{(P)}(f_j)}{\sum_{j=1}^M \hat{S}^{(P)}(f_j)}, \quad k = 1, \dots, M-1, \quad (215a)$$

the  $\mathcal{P}_k$  terms have the same distribution as an ordered random sample of  $M-1$  RVs from the uniform distribution over the interval  $[0, 1]$  (Bartlett, 1955). We can thus base our test for white noise on the well-known *Kolmogorov goodness-of-fit test* for a completely specified distribution (Conover, 1999). If we let

$$D^+ \stackrel{\text{def}}{=} \max_{1 \leq k \leq M-1} \left( \frac{k}{M-1} - \mathcal{P}_k \right) \quad \text{and} \quad D^- \stackrel{\text{def}}{=} \max_{1 \leq k \leq M-1} \left( \mathcal{P}_k - \frac{k-1}{M-1} \right),$$

we can reject the null hypothesis of white noise at the  $\alpha$  level of significance if the Kolmogorov test statistic  $D \stackrel{\text{def}}{=} \max(D^+, D^-)$  exceeds  $D(\alpha)$ , the upper  $\alpha \times 100\%$  percentage point for the distribution of  $D$  under the null hypothesis. A simple approximation to  $D(\alpha)$  is given by

$$\tilde{D}(\alpha) \stackrel{\text{def}}{=} \frac{C(\alpha)}{(M-1)^{1/2} + 0.12 + 0.11/(M-1)^{1/2}}, \quad (215b)$$



**Figure 216.** Periodogram (a) and associated normalized cumulative periodogram (b) for the first 32 values of the AR(2) series in Figure 34(a). The periodogram is plotted on a linear/linear scale.

where  $C(0.10) = 1.224$ ,  $C(0.05) = 1.358$  and  $C(0.01) = 1.628$  (Stephens, 1974). For all  $M \geq 7$ , this approximation is good to within 1% for the three stated values of  $\alpha$  (i.e.,  $0.99 < \tilde{D}(\alpha)/D(\alpha) < 1.01$  for  $\alpha = 0.10, 0.05$  and  $0.01$ ). Values for  $D(\alpha)$  can be found in Table A13 of Conover (1999).

We can obtain a useful graphical equivalent of this test by plotting  $\mathcal{P}_k$  versus  $f_k$  for  $k = 1, \dots, M-1$  along with the lines defined by

$$L_u(f) \stackrel{\text{def}}{=} \frac{fN\Delta_t - 1}{M-1} + \tilde{D}(\alpha) \quad \text{and} \quad L_l(f) \stackrel{\text{def}}{=} \frac{fN\Delta_t}{M-1} - \tilde{D}(\alpha). \quad (216)$$

If any of the points given by  $(f_k, \mathcal{P}_k)$  falls outside the region between these two lines, then we reject the null hypothesis at a level of significance of  $\alpha$ ; conversely, if all of the points fall inside this region, then we fail to reject the null hypothesis at the stated level.

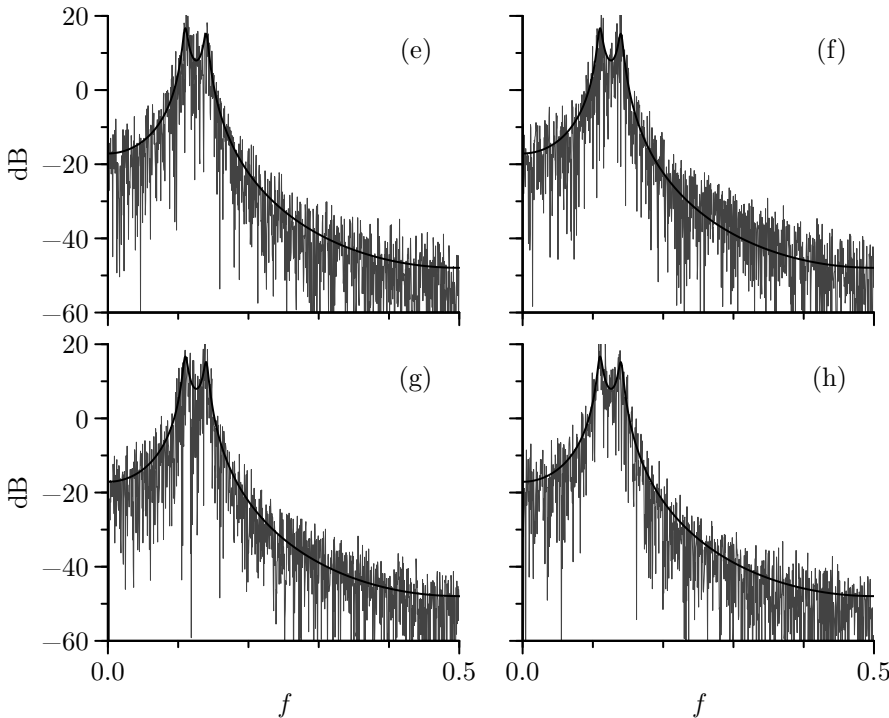
As a concrete example, let us consider a short time series consisting of the first  $N = 32$  values of the AR(2) series of Figure 34(a). Figure 216(a) shows  $\hat{S}^{(P)}(f_j)$  for this series versus  $f_j = j/32$  for  $j = 1, \dots, M = 15$  (recall that  $\Delta_t = 1$ ). The jagged thin curve in 216(b) depicts the corresponding normalized cumulative periodogram  $\mathcal{P}_k$  versus  $f_k$  for  $k = 1, \dots, M-1 = 14$ . The two thick lines are  $L_u(\cdot)$  and  $L_l(\cdot)$  corresponding to a significance level of  $\alpha = 0.05$ ; specifically, since here  $\tilde{D}(0.05) \doteq 0.349$  (for the record,  $D(0.05) \doteq 0.349$  also), we have

$$L_u(f) \doteq \frac{32}{14}f - \frac{1}{14} + 0.349 \quad \text{and} \quad L_l(f) \doteq \frac{32}{14}f - 0.349.$$

Since the normalized cumulative periodogram is not entirely contained in the region between these two lines, we reject the null hypothesis of white noise at the 0.05 level of significance. Equivalently, the values of  $D^+$  and  $D^-$  are, respectively, 0.066 and 0.392, so the value of  $D$  is 0.392, which – since it exceeds  $\tilde{D}(0.05) \doteq 0.349$  – tells us to reject the null hypothesis. We note that  $D^-$  is the value associated with  $f_5 = 5/32$ ; i.e.,  $D^- = \mathcal{P}_5 - 4/14$  with  $\mathcal{P}_5 \doteq 0.6773$ . (For other usages of this test, see the discussions surrounding Figure 231, C&E [1] for Section 10.9, Figure 574 and Exercise [10.13].)

Finally, we note that the cumulative periodogram test requires computation of the periodogram over the Fourier frequencies. When, for some sample sizes  $N$ , we cannot use an FFT algorithm directly to compute the required DFT of  $X_0, \dots, X_{N-1}$ , we can use the chirp transform algorithm to obtain the required DFT efficiently (see the discussion in C&E [1] for Section 3.11).

[7] We noted in C&E [2] for Section 6.3 that leakage in the periodogram can be attributed to a mismatch between the extremes of a time series. While tapering is one approach for eliminating this mismatch, another is to extend the series  $X_0, X_1, \dots, X_{N-1}$  via reflection to obtain the series of length  $2N$  depicted in Equation (183b) and to compute the periodogram of the reflection-extended series (throughout this discussion, we assume  $\Delta_t = 1$  for convenience).



**Figure 217.** DCT-based periodograms (rough-looking light curves) for the four AR(4) time series of length  $N = 1024$  shown in Figure 34, along with the true SDF for the AR(4) process (smooth thick curves). The DCT-periodograms are computed over the grid of frequencies  $\tilde{f}_k = k/(2N)$ . The usual periodograms are shown in Figure 182 over this same grid and in Figure 173 over the grid of Fourier frequencies  $f_k = k/N$ .

▷ **Exercise [217]** Show that the periodogram of the reflection-extended series over its Fourier frequencies  $\tilde{f}_k = k/(2N)$  is given by

$$\tilde{S}^{(\text{DCT})}(\tilde{f}_k) = \frac{2}{N} \left( \sum_{t=0}^{N-1} X_t \cos(\pi \tilde{f}_k [2t+1]) \right)^2 = \frac{2}{N} \left( \sum_{t=0}^{N-1} X_t \cos\left(\frac{\pi k [2t+1]}{2N}\right) \right)^2.$$

Show also that this estimator is useless at  $\tilde{f}_k = 1/2$  since  $\tilde{S}^{(\text{DCT})}(1/2) = 0$  always. ◁

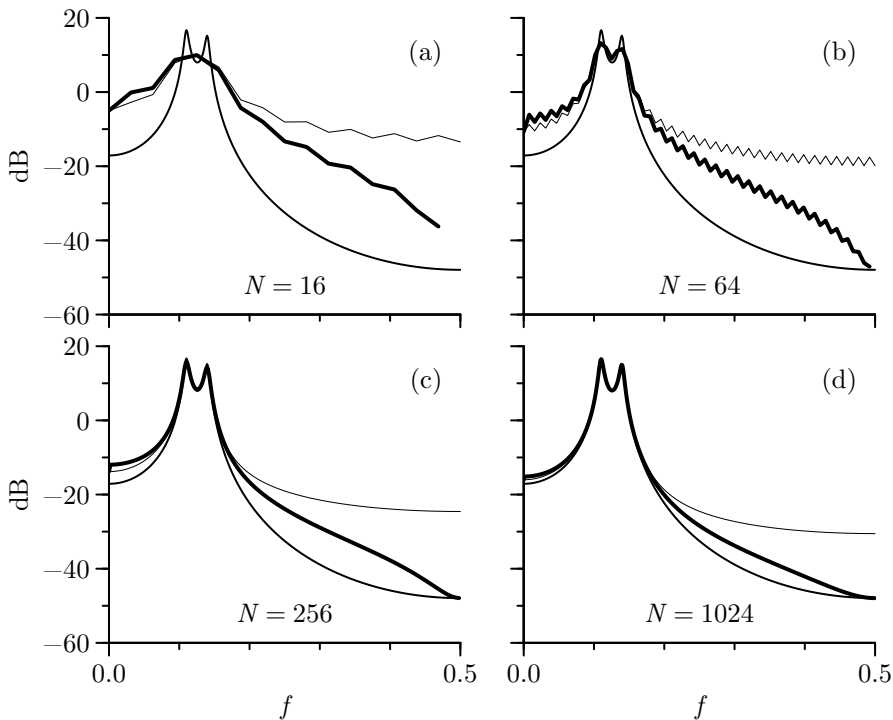
The rationale for the superscript on  $\tilde{S}^{(\text{DCT})}(\cdot)$  is that this estimator can be related to the so-called *type-II discrete cosine transform* (DCT-II; see Ahmed et al., 1974, and Strang, 1999, for background). One definition for the DCT-II of  $X_0, X_1, \dots, X_{N-1}$  is

$$C_k = \left( \frac{2 - \delta_k}{N} \right)^{1/2} \sum_{t=0}^{N-1} X_t \cos\left(\frac{\pi k [2t+1]}{2N}\right), \quad k = 0, 1, \dots, N-1$$

(Gonzalez and Woods, 2007), where  $\delta_k$  is Kronecker's delta function (i.e.,  $\delta_k = 1$  if  $k = 0$  and  $= 0$  otherwise). Hence  $\tilde{S}^{(\text{DCT})}(\tilde{f}_k) = C_k^2$  for  $k = 1, 2, \dots, N-1$ , while  $\tilde{S}^{(\text{DCT})}(\tilde{f}_0) = 2C_0^2$ . Because of this connection, we define the *DCT-based periodogram* to be

$$\hat{S}^{(\text{DCT})}(\tilde{f}_k) = C_k^2, \quad k = 0, 1, \dots, N-1,$$

which differs from  $\tilde{S}^{(\text{DCT})}(\tilde{f}_k)$  only at  $k = 0$  (Chan et al., 1995, use the term *semiperiodogram* for a related concept).



**Figure 218.** Expected value of periodogram  $\hat{S}^{(P)}(\tilde{f}_k)$  (thinnest curve) and of DCT-based periodogram  $\hat{S}^{(\text{DCT})}(\tilde{f}_k)$  (thickest) versus  $\tilde{f}_k = k/(2N)$  for AR(4) process of Equation (35a) with sample sizes  $N = 16, 64, 256$  and  $1024$ . In each plot the true SDF is the smooth curve of medium thickness.

Figure 217 shows  $\hat{S}^{(\text{DCT})}(\cdot)$  for the four AR(4) series displayed in Figure 34, along with the true SDF for the AR(4) process. If we compare these estimates to the corresponding periodograms depicted in Figures 173 and 182, the DCT-based periodograms appear to have less bias overall; however, they are also noisier looking, particularly in the twin-peaks area of the SDF.

To explore the bias and variance properties of  $\hat{S}^{(\text{DCT})}(\cdot)$  in more depth, suppose that  $\{X_t\}$  is a Gaussian stationary process with zero mean and ACVS  $\{s_\tau\}$ . Then, for all  $k$ ,  $C_k$  also obeys a Gaussian distribution with mean zero and variance, say,  $\sigma_k^2$ .

▷ **Exercise [218]** Show that

$$\sigma_0^2 = \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_\tau,$$

while, for  $k = 1, \dots, N-1$ ,

$$\sigma_k^2 = E\{\hat{S}^{(P)}(\tilde{f}_k)\} - \frac{2}{N \sin(\pi k/N)} \sum_{\tau=1}^{N-1} s_\tau \sin(\pi \tau k/N). \quad \triangleleft$$

Under the assumption that the ACVS is absolutely summable, the results of this exercise can be used to argue that, as was true for the usual periodogram, the DCT-based periodogram is asymptotically unbiased for  $0 \leq f < 1/2$ :

$$\lim_{N \rightarrow \infty} E\{\hat{S}^{(\text{DCT})}(f)\} \rightarrow S(f); \quad (218)$$

however, in contrast to the periodogram,  $\hat{S}^{(\text{DCT})}(1/2)$  does not have a useful definition.



Figure 218 compares the expected value of the usual periodogram  $\hat{S}^{(P)}(\cdot)$  (thinnest curve in each plot) with that of the DCT-based periodogram  $\hat{S}^{(\text{DCT})}(\cdot)$  (thickest) over the frequencies  $\tilde{f}_k$  for the AR(4) process of Equation (35a), along with the true SDF. Each plot depicts a different sample size, namely,  $N = 16, 64, 256$  and  $1024$  (cf. Figure 178). At high frequencies  $\hat{S}^{(\text{DCT})}(\cdot)$  is considerably less biased than  $\hat{S}^{(P)}(\cdot)$ , but its bias is somewhat larger at low frequencies. The two estimators have very similar first moments in the high-power twin-peaks portion of the SDF. In keeping with Equation (218), the bias in  $\hat{S}^{(\text{DCT})}(\cdot)$  decreases as  $N$  increases.

The theory above implies that

$$\left(\frac{\mathcal{C}_k}{\sigma_k}\right)^2 \stackrel{\text{d}}{=} \chi_1^2, \text{ i.e., } \hat{S}^{(\text{DCT})}(\tilde{f}_k) \stackrel{\text{d}}{=} \sigma_k^2 \chi_1^2, \quad (219a)$$

and hence  $E\{\hat{S}^{(\text{DCT})}(\tilde{f}_k)\} = \sigma_k^2$  and  $\text{var}\{\hat{S}^{(\text{DCT})}(\tilde{f}_k)\} = 2\sigma_k^4$  for all  $k$ . Specializing to the case of Gaussian white noise with variance  $\sigma^2 = S(f)$ , we find

$$E\{\hat{S}^{(\text{DCT})}(\tilde{f}_k)\} = S(\tilde{f}_k) \text{ and } \text{var}\{\hat{S}^{(\text{DCT})}(\tilde{f}_k)\} = 2S^2(\tilde{f}_k)$$

for  $0 \leq \tilde{f}_k < 1/2$ ; by contrast, we have

$$E\{\hat{S}^{(P)}(f_k)\} = S(f_k) \text{ and } \text{var}\{\hat{S}^{(P)}(f_k)\} = S^2(f_k)$$

for  $0 < f_k < 1/2$ . The DCT-based periodogram thus has twice the variance of the usual periodogram. This result holds approximately for colored Gaussian processes, which explains why the DCT-based periodograms in Figure 217 are markedly choppier looking than their  $\hat{S}^{(P)}(\cdot)$  counterparts in Figures 173 and 182.

We leave it as an exercise for the reader to show that, for Gaussian white noise with mean zero and variance  $\sigma^2$ , the RVs  $\hat{S}^{(\text{DCT})}(\tilde{f}_k)$ ,  $k = 0, 1, \dots, N-1$ , are independent and identically distributed with a distribution dictated by  $\sigma^2 \chi_1^2$ . The collective number of degrees of freedom is thus  $N$ . By contrast, the RVs  $\hat{S}^{(P)}(f_k)$ ,  $k = 0, 1, \dots, \lfloor N/2 \rfloor$ , are independent, but not identically distributed: the distribution for  $k = 1, \dots, \lfloor (N-1)/2 \rfloor$  is given by  $\sigma^2 \chi_2^2/2$ , whereas that for  $k = 0$  and  $k = N/2$  (when  $N$  is an even sample size) is dictated by  $\sigma^2 \chi_1^2$ . For both even and odd sample sizes, the collective number of degrees of freedom is still  $N$ . Ignoring  $f = 0$  and  $f = 1/2$ , while the RVs in the DCT-based periodogram have twice the variability as those of the usual periodogram, this deficiency is compensated for by the fact that there are (approximately) twice as many independent RVs in the DCT-based periodogram.

For additional discussion on the use of the DCT-II for SDF estimation, see Davies (2001) and Narasimhan and Harish (2006). Also the celebrated physicist Albert Einstein wrote a short article on spectral analysis (Einstein, 1914), in which he proposed smoothing a basic SDF estimator that is essentially based on a discrete cosine transform; moreover, his use of this transform evidently arose from the (then) common practice of using reflection-extended time series in conjunction with Fourier analysis (Yaglom, 1987b).

Exercise [6.31] invites the reader to redo the key results above for the alternative way of constructing a reflected series displayed in Equation (183c).

## 6.7 Computational Details

Here we give some details about computing the periodogram and other direct spectral estimators. We shall strive for clarity rather than get into issues of computational efficiency. All computations are done in terms of DFTs and inverse DFTs, so we first review the following pertinent results.

Recall that the DFT for a finite sequence  $g_0, \dots, g_{M-1}$  with a sampling interval of unity is the sequence  $G_0, \dots, G_{M-1}$ , where, with  $\tilde{f}'_k = k/M$ ,

$$G_k = \sum_{t=0}^{M-1} g_t e^{-i2\pi kt/M} = \sum_{t=0}^{M-1} g_t e^{-i2\pi \tilde{f}'_k t} \quad (219b)$$

and

$$g_t = \frac{1}{M} \sum_{k=0}^{M-1} G_k e^{i2\pi kt/M} = \frac{1}{M} \sum_{k=0}^{M-1} G_k e^{i2\pi \tilde{f}'_k t} \quad (220a)$$

(see Equations (91b) and (92a)). This relationship is summarized by the notation  $\{g_t\} \longleftrightarrow \{G_k\}$ . Both of these sequences are defined outside the range 0 to  $M - 1$  by cyclic (periodic) extension; i.e., if  $t < 0$  or  $t \geq M$ , then  $g_t \stackrel{\text{def}}{=} g_{t \bmod M}$  so that, for example,  $g_{-1} = g_{M-1}$  and  $g_M = g_0$ . We define the cyclic autocorrelation of  $\{g_t\}$  by

$$g \star g_t^* = \sum_{s=0}^{M-1} g_{s+t} g_s^*, \quad t = 0, \dots, M-1 \quad (220b)$$

(cf. Equation (101h)). If  $\{h_t\}$  is another sequence of length  $M$  with DFT  $\{H_k\}$ , i.e.,  $\{h_t\} \longleftrightarrow \{H_k\}$ , we define the cyclic convolution of  $\{g_t\}$  and  $\{h_t\}$  by

$$g \star h_t = \sum_{s=0}^{M-1} g_s h_{t-s}, \quad t = 0, \dots, M-1 \quad (220c)$$

(see Equation (101d)). We define the frequency domain cyclic convolution of  $\{G_k\}$  and  $\{H_k\}$  by

$$G \star H_k = \frac{1}{M} \sum_{l=0}^{M-1} G_l H_{k-l}, \quad k = 0, \dots, M-1 \quad (220d)$$

(see Equation (101f)). We note that

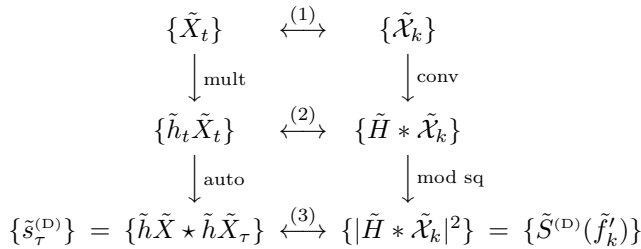
$$\{g \star g_t^*\} \longleftrightarrow \{|G_k|^2\}, \quad \{g \star h_t\} \longleftrightarrow \{G_k H_k\} \quad \text{and} \quad \{g_t h_t\} \longleftrightarrow \{G \star H_k\}$$

(see Equations (102a), (101e) and (101g)). Exercise [3.20a] outlines an efficient method for computing cyclic autocorrelations and convolutions.

We start with a time series of length  $N$  that can be regarded as a realization of  $X_0, X_1, \dots, X_{N-1}$ , a segment of length  $N$  of the stationary process  $\{X_t\}$  with sampling interval  $\Delta_t$  and unknown mean  $\mu$ , SDF  $S(\cdot)$  and ACVS  $\{s_\tau\}$ . Let  $M$  be any integer satisfying  $M \geq 2N - 1$  (the choice  $M = 2N$  is often used, particularly when  $N$  is a power of two). We let

$$\tilde{X}_t \stackrel{\text{def}}{=} \begin{cases} X_t - \bar{X}, & 0 \leq t \leq N-1; \\ 0, & N \leq t \leq M-1; \end{cases} \quad \text{and} \quad \tilde{h}_t \stackrel{\text{def}}{=} \begin{cases} h_t, & 0 \leq t \leq N-1; \\ 0, & N \leq t \leq M-1, \end{cases} \quad (220e)$$

where  $\bar{X} = \sum_{t=0}^{N-1} X_t / N$  is the sample mean, and  $\{h_t\}$  is a data taper of length  $N$  with normalization  $\sum_{t=0}^{N-1} h_t^2 = 1$ . The sequences  $\{\tilde{X}_t\}$  and  $\{\tilde{h}_t\}$  are each of length  $M$  and are implicitly defined outside the range  $t = 0$  to  $M - 1$  by cyclic extension. For convenience, we perform our computations assuming that the sampling interval for these sequences – and all others referred to in Figure 221 – is unity so that the DFT formulae of Equations (219b) and (220a) can be used (after completion of the computations, we indicate the adjustments needed to handle  $\Delta_t \neq 1$ ). We create these two zero padded sequences so that we can compute various noncyclic convolutions using cyclic convolution (see Exercise [3.20b]). We denote the DFTs of  $\{\tilde{X}_t\}$  and  $\{\tilde{h}_t\}$  by, respectively, the sequences  $\{\tilde{\mathcal{X}}_k\}$  and  $\{\tilde{H}(\tilde{f}'_k)\}$ , both of which are of length  $M$  and indexed by  $k = 0, \dots, M - 1$  (note that  $\tilde{\mathcal{X}}_k$  is different from  $\mathcal{X}_k$  of



**Figure 221.** Pathways for computing  $\tilde{S}^{(D)}(\cdot)$ , from which, upon multiplication by  $\Delta_t$ , we can obtain the direct spectral estimate  $\hat{S}^{(D)}(f'_k) = \Delta_t \tilde{S}^{(D)}(\tilde{f}'_k)$ , where  $f'_k = k/(M \Delta_t)$  and  $\tilde{f}'_k = k/M$  (adapted from figure 1 of Van Schooneveld and Frijling, 1981).

Equation (171a) because we have forced  $\{\tilde{X}_t\}$  to have a sampling interval of unity). As usual, these Fourier relationships are summarized by

$$\{\tilde{X}_t\} \longleftrightarrow \{\tilde{\mathcal{X}}_k\} \quad \text{and} \quad \{\tilde{h}_t\} \longleftrightarrow \{\tilde{H}(\tilde{f}'_k)\}.$$

Our goal is to compute either the periodogram  $\hat{S}^{(P)}(\cdot)$  or some other direct spectral estimate  $\hat{S}^{(D)}(\cdot)$  at frequencies  $f'_k = k/(M \Delta_t)$  (note the difference between  $f'_k$  and  $\tilde{f}'_k = k/M$ : the former has physically meaningful units, whereas we take the latter to be unitless). Since the periodogram is a special case of  $\hat{S}^{(D)}(\cdot)$ , we first consider computation of direct spectral estimates in general and then comment afterwards on shortcuts that can be made if only the periodogram is of interest. Figure 221 depicts pathways to compute  $\hat{S}^{(D)}(\cdot)$ . These involve three Fourier transform pairs, the first of which we have already noted. Here are comments regarding the remaining two.

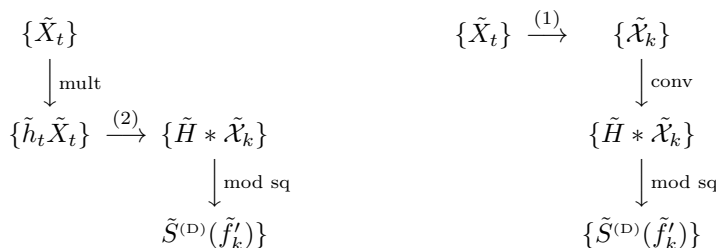
- (2) The left-hand side is obtained from a term by term multiplication of the sequences  $\{\tilde{h}_t\}$  and  $\{\tilde{X}_t\}$ , while its DFT (the right-hand side) can be obtained either directly or by convolving the sequences  $\{\tilde{H}(\tilde{f}'_k)\}$  and  $\{\tilde{\mathcal{X}}_k\}$  using Equation (220d).
- (3) The left-hand side is the autocorrelation of  $\{\tilde{h}_t \tilde{X}_t\}$  and can be obtained using Equation (220b). Its DFT can be obtained either directly or by computing the square modulus of  $\{\tilde{H} * \tilde{\mathcal{X}}_k\}$  term by term. If we let  $\tilde{s}_\tau^{(D)}$  be the  $\tau$ th element of the autocorrelation of  $\{\tilde{h}_t \tilde{X}_t\}$ , then we have

$$\hat{s}_\tau^{(D)} = \begin{cases} \tilde{s}_\tau^{(D)}, & |\tau| \leq N-1; \\ 0, & \text{otherwise.} \end{cases} \quad (221)$$

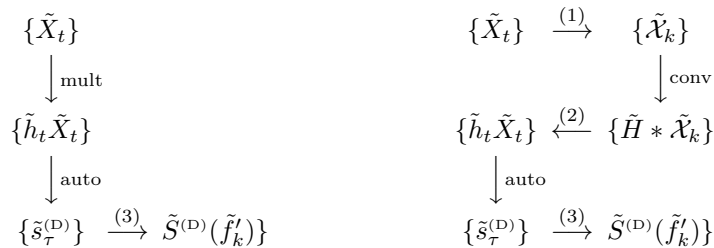
On the right-hand side, if we let  $\tilde{S}^{(D)}(\tilde{f}'_k) \stackrel{\text{def}}{=} |\tilde{H} * \tilde{\mathcal{X}}_k|^2$ , then we have

$$\hat{S}^{(D)}(f'_k) = \Delta_t \tilde{S}^{(D)}(\tilde{f}'_k), \quad k = 0, \dots, \lfloor M/2 \rfloor.$$

As shown in Figure 221, there are four computational pathways to get  $\{\tilde{S}^{(D)}(\tilde{f}'_k)\}$  (and hence  $\{\hat{S}^{(D)}(f'_k)\}$ ) from  $\{X_t\}$ . Here are two of them:



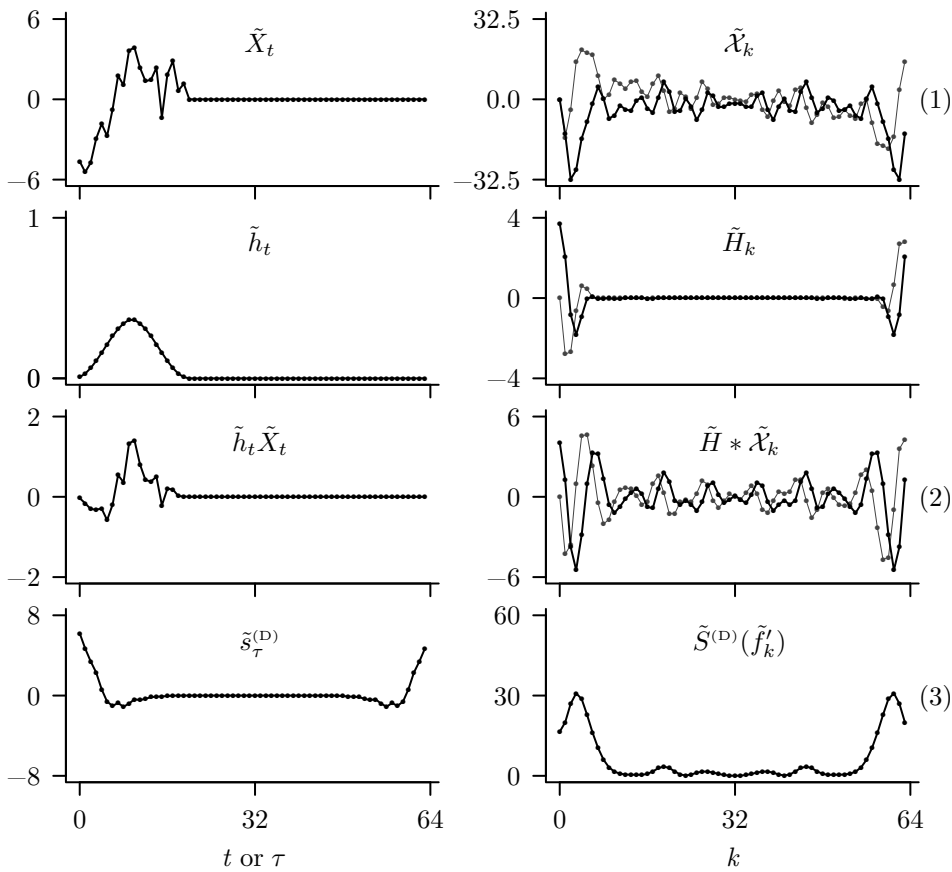
The one on the left avoids convolution and autocorrelation (these operations can be time consuming on a digital computer if implemented via their definitions in Equations (220b) and (220d)). The right-hand pathway makes use of a convolution, which, from a computational point of view, makes it less favorable; however, Exercise [6.18a] notes the theoretically interesting fact that one version of the Hanning data taper has a DFT with just a few nonzero terms, and hence the convolution can be done relatively quickly. If we are also interested in the ACVS estimate  $\{\hat{s}^{(D)}\}$ , we can augment either pathway to obtain it by taking the inverse Fourier transform of  $\{|\tilde{H} * \tilde{\mathcal{X}}_k|^2\}$ . In the final two pathways, we need to create  $\{\hat{s}^{(D)}\}$  to get to  $\{\tilde{S}^{(D)}(\tilde{f}'_k)\}$ :



Neither pathway is attractive computationally, but both are of pedagogical interest.

As an example of these manipulations, let us consider a time series related to the rotation of the earth. The entire series has 100 values. These are shown, after centering by subtracting off 21.573, by the pluses in the bottoms of Figures 149 and 150. We make use of the first  $N = 20$  values prior to centering. For the record, these are 7.1, 6.3, 7.0, 8.8, 9.9, 9.0, 11.0, 13.5, 12.8, 15.4, 15.6, 14.1, 13.1, 13.2, 14.1, 10.4, 13.6, 14.6, 12.4 and 12.9. The sampling interval is  $\Delta_t = 1/4$  year. Figure 223 shows four Fourier transform pairs of interest, one pair per row of plots: the sequence in each right-hand plot is the DFT of the left-hand sequence. The three Fourier transform pairs in Figure 221 are shown in rows 1, 3 and 4 of Figure 223. To use a conventional “power of two” FFT, we let  $M = 64$  (the smallest power of two greater than  $2N - 1 = 39$ ). The left-hand plot of the first row thus shows  $\{\tilde{X}_t\}$ , a sequence of length 64, the first 20 values of which constitute the series centered using its sample mean (11.74), and the last 44 values are all zeros. The corresponding right-hand plot is its DFT – a sequence of 64 complex-valued numbers – the real parts of which are shown by the dark connected dots, and the imaginary parts by the gray ones. The second row of plots shows  $\{\tilde{h}_t\}$  (here a Hanning data taper of length 20 padded with 44 zeros) and its DFT (another complex-valued sequence). The left-hand plot on the third row is  $\{\tilde{h}_t \tilde{X}_t\}$ . The left-hand plot on the bottom row is  $\{\tilde{s}_\tau^{(D)}\}$ , from which we can pick out  $\{\hat{s}_\tau^{(D)}\}$ ; the right-hand plot is  $\{\tilde{S}^{(D)}(\tilde{f}'_k)\}$ , which is real-valued. For  $k = 0$  to 32, multiplication of  $\tilde{S}^{(D)}(\tilde{f}'_k)$  by  $\Delta_t = 1/4$  yields  $\hat{S}^{(D)}(f'_k)$ , a direct spectral estimator of  $S(\cdot)$  at frequency  $f'_k = k/16$  cycles per year (note that  $k = 32$  corresponds to the Nyquist frequency of 2 cycles per year). One comment: a data taper is of questionable utility for this short series and is employed here primarily for illustrative purposes.

Since the periodogram is a direct spectral estimator that makes use of the default (rectangular) data taper  $h_t = 1/\sqrt{N}$ , we could compute it as described by Figure 221. In this case  $\{\tilde{H} * \tilde{\mathcal{X}}_k\}$  ends up being just the Fourier transform of  $\{\tilde{X}_t/\sqrt{N}\}$ , and the periodogram is gotten by multiplying  $\{|\tilde{H} * \tilde{\mathcal{X}}_k|\}$  by  $\Delta_t$ . We can simplify Figure 221 by dropping the second row since it really does nothing more than provide a division by  $\sqrt{N}$ . This leads to the pathways shown in Figure 224a, which yields  $\{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}$ , whereas the pathways in Figure 221 give us  $\{\tilde{S}^{(D)}(\tilde{f}'_k)\} = \{\tilde{S}^{(P)}(\tilde{f}'_k)\}$ . Dropping the division by  $\sqrt{N}$  thus turns into a need to eventually divide by  $N$ . When  $\Delta_t \neq 1$ , we can combine the division by  $N$  with the multiplication by  $\Delta_t$  needed to get the actual periodogram  $\{\hat{S}^{(P)}(f'_k)\} = \{\Delta_t \tilde{S}^{(P)}(\tilde{f}'_k)\}$ ; i.e., a single multiplication by  $\Delta_t/N$  gets us from  $\{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}$  to the desired  $\{\hat{S}^{(P)}(f'_k)\}$ .



**Figure 223.** Illustration of computational pathways for direct spectral estimates (cf. Figure 221).

While there are two pathways from  $\{\tilde{X}_t\}$  to  $\{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}$  embedded in Figure 224a, the one of main computation interest is

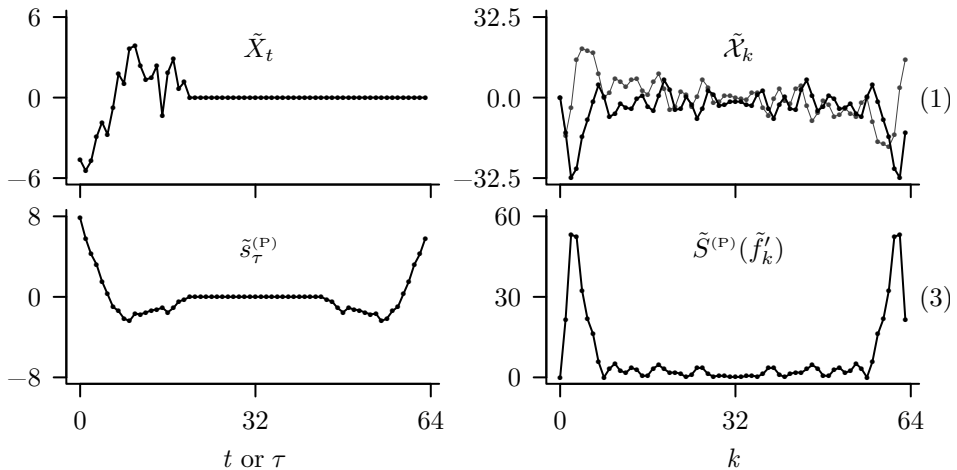
$$\begin{aligned} \{\tilde{X}_t\} &\xrightarrow{(1)} \{\tilde{\mathcal{X}}_k\} \\ &\downarrow \text{mod sq} \\ \{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}. \end{aligned}$$

This pathway avoids the autocorrelation embedded in the other pathway, but does not give us the biased ACVS estimate directly. We can obtain this estimate via the inverse DFT of  $\{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}$ , which yields  $\{N \cdot \tilde{s}_\tau^{(P)}\}$ , from which we can extract  $\{\hat{s}_\tau^{(P)}\}$  through an analog of Equation (221). If we are only interested in computing the periodogram over the grid of Fourier frequencies  $f_k = k/(N \Delta_t)$ , then we could start by defining  $\tilde{X}_t = X_t - \bar{X}$ ,  $0 \leq t \leq N - 1$ ; i.e., we eliminate the zero padding in  $\{\tilde{X}_t\}$ . Letting  $\{\tilde{\mathcal{X}}_k\}$  denote the DFT of  $\{\tilde{X}_t\}$ , we would then entertain the pathway

$$\begin{aligned} \{\tilde{X}_t\} &\xleftrightarrow{(1)} \{\tilde{\mathcal{X}}_k\} \\ &\downarrow \text{mod sq} \\ \{|\tilde{\mathcal{X}}_k|^2\} &= \{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}, \end{aligned}$$

$$\begin{array}{ccc}
 \{\tilde{X}_t\} & \xleftrightarrow{(1)} & \{\tilde{\mathcal{X}}_k\} \\
 \downarrow \text{auto} & & \downarrow \text{mod sq} \\
 \{N \cdot \tilde{s}_\tau^{(P)}\} = \{\tilde{X} \star \tilde{X}_\tau\} & \xleftrightarrow{(3)} & \{|\tilde{\mathcal{X}}_k|^2\} = \{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}
 \end{array}$$

**Figure 224a.** Modification of Figure 221 appropriate for computing  $N \cdot \tilde{S}^{(P)}(\cdot)$ , from which, upon multiplication by  $\Delta_t/N$ , we obtain the periodogram  $\hat{S}^{(P)}(f'_k) = \Delta_t \tilde{S}^{(P)}(\tilde{f}'_k)$ , where, as before,  $f'_k = k/(M \Delta_t)$  and  $\tilde{f}'_k = k/M$ .



**Figure 224b.** Illustration of computational pathways for the periodogram (cf. Figure 224a).

from which we would extract the actual periodogram via  $\hat{S}^{(P)}(f_k) = \Delta_t \tilde{S}^{(P)}(\tilde{f}_k)$ ,  $k = 0, \dots, \lfloor N/2 \rfloor$ , where here  $\tilde{f}_k = k/N$ . We have, however, lost the ability to get the biased estimate of the ACVS; i.e., the inverse DFT of  $\{N \cdot \tilde{S}^{(P)}(\tilde{f}_k)\}$  is *not* equal to  $\{N \cdot \tilde{s}_\tau^{(P)}\}$  in general. We must stick with zero padding if we want to get the biased ACVS estimate.

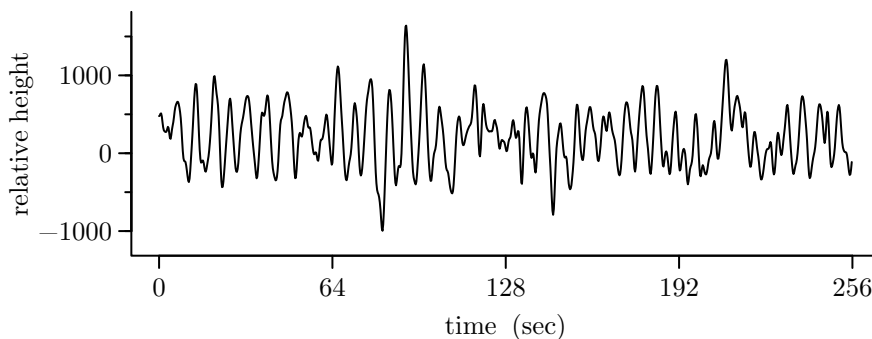
Figure 224b illustrates the pathways in Figure 224a using the same 20-point time series as before. The top row is the same as that in Figure 223. The left-hand plot on the bottom row is  $\{\tilde{s}_\tau^{(P)}\}$ , from which we can get the values for  $\{\hat{s}_\tau^{(P)}\}$  via an analog of Equation (221); the right-hand plot is the real-valued sequence  $\{\tilde{S}^{(P)}(\tilde{f}'_k)\}$  (we show  $\{\tilde{s}_\tau^{(P)}\}$  and  $\{\tilde{S}^{(P)}(\tilde{f}'_k)\}$  rather than  $\{N \cdot \tilde{s}_\tau^{(P)}\}$  and  $\{N \cdot \tilde{S}^{(P)}(\tilde{f}'_k)\}$  to facilitate comparison with Figure 223). For  $k = 0$  to 32, multiplying  $\tilde{S}^{(P)}(\tilde{f}'_k)$  by  $\Delta_t = 1/4$  yields  $\hat{S}^{(P)}(f'_k)$ , the periodogram at frequency  $f'_k = k/16$  cycles per year.

## 6.8 Examples of Periodogram and Other Direct Spectral Estimators

### *Ocean Wave Data*

Figure 225 shows a plot of a time series recorded in the Pacific Ocean by a wave-follower. As the wave-follower moved up and down with the waves in the ocean, it measured the surface displacement (i.e., sea level) as a function of time. The frequency response of the wave-follower was such that – mainly due to its inertia – frequencies higher than 1 Hz could not be reliably measured. The data were originally recorded using an analog device. They were low-pass filtered in analog form using an antialiasing filter with a cutoff of approximately 1 Hz and then sampled every 1/4 sec, yielding a Nyquist frequency of  $f_N = 2$  Hz. The plotted series consists of a 256 sec portion of this data, so there are  $N = 1024$  data points in all.

Figure 226 shows four direct spectral estimates for this ocean wave data, each evaluated on the grid of 512 positive Fourier frequencies. The top plot is the periodogram  $\hat{S}^{(P)}(\cdot)$ ,



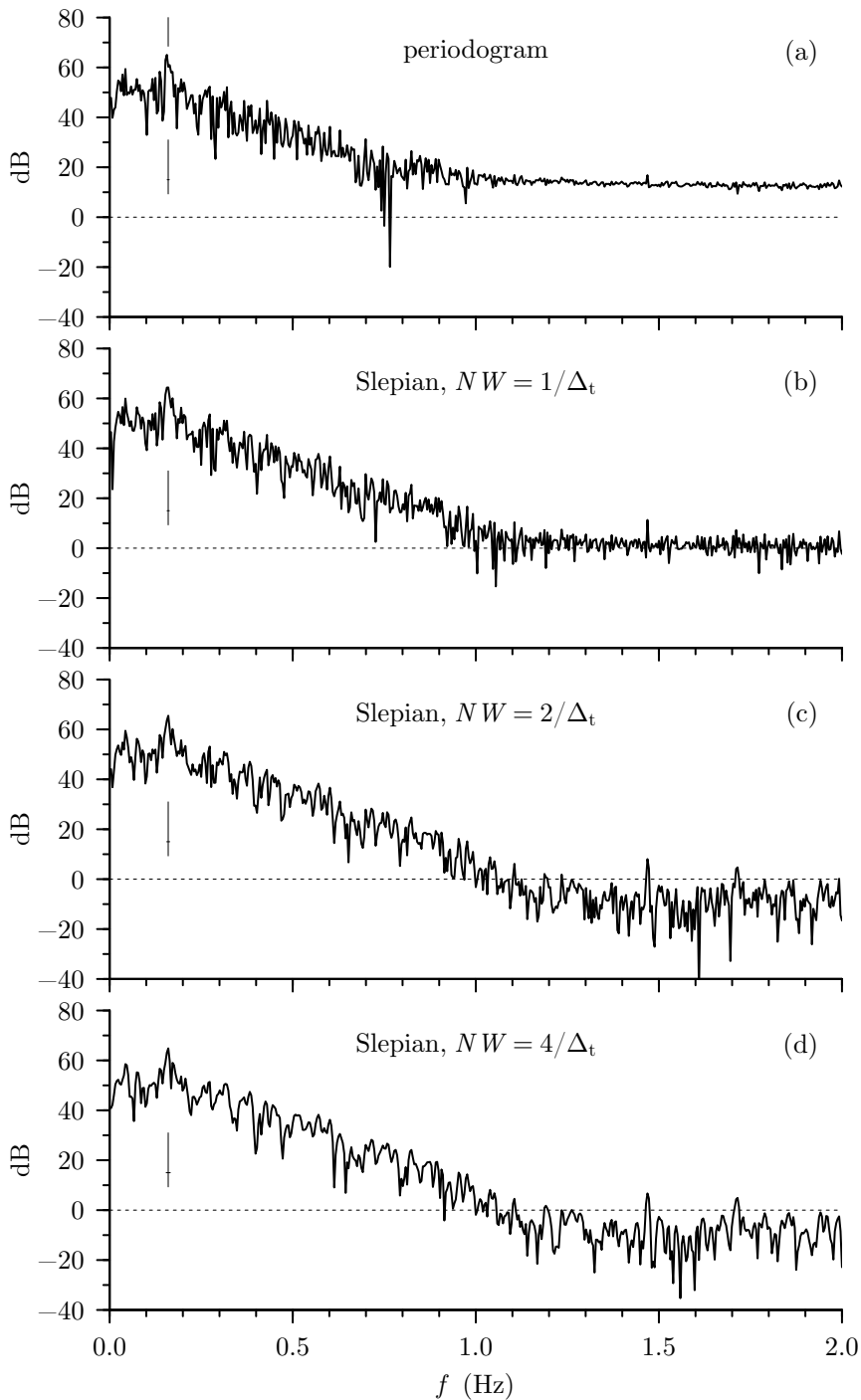
**Figure 225.** Ocean wave data (courtesy of A. Jessup, Applied Physics Laboratory, University of Washington). There are  $N = 1024$  samples taken  $\Delta_t = 1/4$  sec apart.

while the bottom three plots are direct spectral estimates  $\hat{S}^{(D)}(\cdot)$  using a Slepian data taper with  $NW$  parameters of  $1/\Delta_t$ ,  $2/\Delta_t$  and  $4/\Delta_t$ . We centered the time series by subtracting off the sample mean  $\bar{X} \doteq 209.1$  prior to computing all four spectral estimates. We used Equation (196b) to form the Slepian tapers  $\{h_t\}$  and normalized them such that  $\sum h_t^2 = 1$ .

All four spectral estimates show a broad, low frequency peak at 0.160 Hz, corresponding to a period of 6.2 sec (the location of this frequency is marked by a thin vertical line in the upper left-hand corner of the top plot). While the dominant features of the time series can be attributed to this broad peak and other features in the frequency range 0 to 0.2 Hz, the data were actually collected to investigate whether the rate at which the SDF decreases over the range 0.2 to 1.0 Hz is in fact consistent with calculations based upon a physical model. The range from 1 to 2 Hz is of little physical interest because it is dominated by instrumentation and preprocessing (i.e., inertia in the wave-follower and the effect of the antialiasing filter); nonetheless, it is of operational interest to examine this portion of the SDF to check that the spectral levels and shape are in accord with the frequency response claimed by the manufacturer of the wave-follower, the transfer function of the antialiasing filter, and a rough guess at the SDF for ocean waves from 1 to 2 Hz.

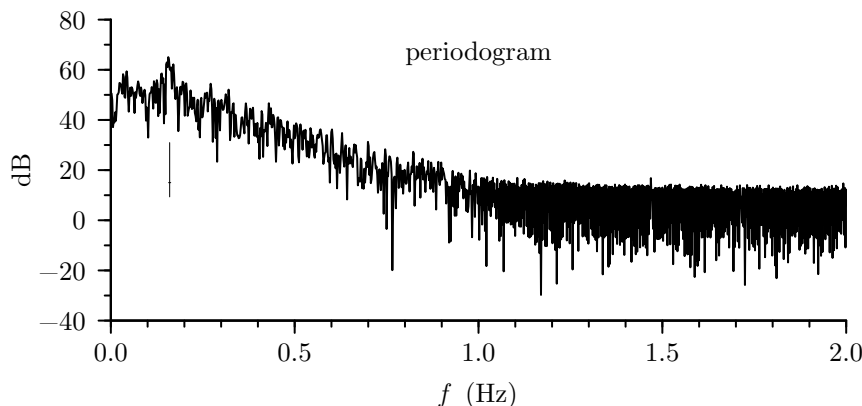
An examination of the high frequency portions of the SDFs in Figure 226 shows evidence of bias due to leakage in both the periodogram and the direct spectral estimate with the  $NW = 1/\Delta_t$  data taper. The evidence is twofold. First, the levels of these two spectral estimates at high frequencies are considerably higher than those for the  $NW = 2/\Delta_t$  and  $NW = 4/\Delta_t$  direct spectral estimates (by about 25 dB for the periodogram and 10 dB for the  $NW = 1/\Delta_t$  spectral estimate). Second, the local variability in the periodogram is markedly less in the high frequencies as compared to the low frequencies, an indication of leakage we have seen previously (see Figures 173(e) and (f)). The same is true to a lesser extent for the  $NW = 1/\Delta_t$  spectral estimate.

As an interesting aside, in Figure 227 we have replotted the periodogram but now over a grid of frequencies twice as fine as the Fourier frequencies – recall that the plots of Figure 226 involve just the Fourier frequencies. This new plot shows an *increase* in the local variability of the periodogram in the high frequencies as compared to the low frequencies. If we were to expand the frequency scale on the plot, we would see that the increased variability is in the form of a ringing; i.e., the SDF alternates between high and low values. Had we seen this plot first, this ringing would have told us that leakage might be a problem. The key point here is that, when plotted on a logarithmic scale (such as decibels), the periodogram and other direct spectral estimators should exhibit approximately the same local variability across all frequencies – if this is not so, it is important to understand why not. Leakage causes the



**Figure 226.** Direct spectral estimates for ocean wave data. The (barely visible) width of the crisscross below each spectral estimate gives the bandwidth measure  $B_{\mathcal{H}}$  as per Equation (194), while its height gives the length of a 95% CI for  $10 \log_{10}(S(f))$  as per Equation (205b).





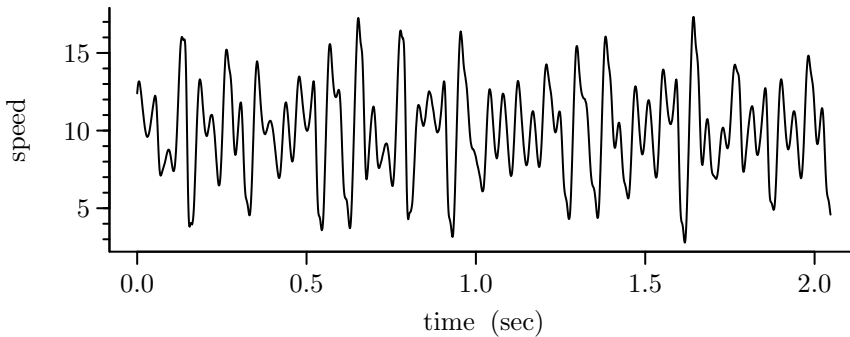
**Figure 227.** Periodogram for ocean wave data. Here the periodogram is plotted over a grid of frequencies twice as fine as the Fourier frequencies, whereas in the top plot of Figure 226 it is plotted over just the Fourier frequencies. The strange appearance of this plot is due to the fact that, because of leakage, the periodogram has a nearly constant value of approximately 12 dB at all high Fourier frequencies, whereas its values are considerably below 12 dB over most of the grid of high frequencies halfway between the Fourier frequencies.

ringing here, but other potential causes are outliers (see, for example, Figure 21 of Martin and Thomson, 1982) and echoes (Bogert et al., 1963).

As in Figures 207 and 208, we have added crisscrosses to the left-hand portions of all the plots in Figures 226 and 227. These are a useful guide in assessing the sampling variability in the depicted spectral estimates because they provide, after being mentally centered on a particular value for  $10 \log_{10}(\hat{S}^{(p)}(f))$  or  $10 \log_{10}(\hat{S}^{(d)}(f))$ , a 95% CI for  $10 \log_{10}(S(f))$ . The horizontal width of each crisscross shows  $B_H$  of Equation (194), which is our standard measure of the effective bandwidth of a direct spectral estimator. Recall that, if  $0 < f < f' < f_N$  and  $|f - f'| \geq B_H$ , we can regard  $\hat{S}^{(d)}(f)$  and  $\hat{S}^{(d)}(f')$  as being approximately uncorrelated. The widths (just barely visible!) of the crisscrosses in the four plots of Figure 226 are, from top to bottom, 0.0059 Hz, 0.0062 Hz, 0.0081 Hz and 0.0113 Hz. Under the assumption that  $S(\cdot)$  is slowly varying, the crisscross gives us a rough idea of the local variability we can expect to see in  $\hat{S}^{(d)}(\cdot)$  across frequencies if all of the assumptions behind spectral analysis hold. If  $f$  and  $f'$  are closer together than the width of the crisscross, we can expect  $\hat{S}^{(d)}(f)$  and  $\hat{S}^{(d)}(f')$  to be similar to some degree because they are positively correlated; on the other hand, if  $f$  and  $f'$  are farther apart than this width,  $\hat{S}^{(d)}(f)$  and  $\hat{S}^{(d)}(f')$  should exhibit a variation consistent with the height of the crisscross. For example, the high frequency portion of the periodogram in the upper plot of Figure 226 shows much less variability from frequency to frequency than we would expect from the height of the crisscross, while, on the finer scale of Figure 227, the local variability is both too large and on too fine a scale (i.e., rapid variations occur on a scale less than the effective bandwidth  $B_H$ ). In contrast, the low frequency portion of the periodogram does not exhibit any gross departure from reasonable variability.

If we now concentrate on the bottom two plots of Figure 226, we see that the  $NW = 2/\Delta_t$  and  $NW = 4/\Delta_t$  spectral estimates look quite similar overall. Increasing the degree of tapering beyond that given by the  $NW = 2/\Delta_t$  data taper thus does not appear to gain us anything. We can conclude that a Slepian data taper with  $NW = 2/\Delta_t$  is sufficient to control leakage over all frequencies for the ocean wave data (if, however, we only want to study the SDF for frequencies less than 1 Hz, the  $NW = 1/\Delta_t$  spectral estimate also has acceptable bias properties).

A careful examination of the  $NW = 2/\Delta_t$  direct spectral estimate shows one interesting feature barely evident in the periodogram, namely, a small peak at  $f = 1.469$  Hz about 15 dB



**Figure 228.** Chaotic beam data (courtesy of W. Constantine, Department of Mechanical Engineering, University of Washington). There are  $N = 2048$  samples taken  $\Delta_t = 0.001$  sec apart.

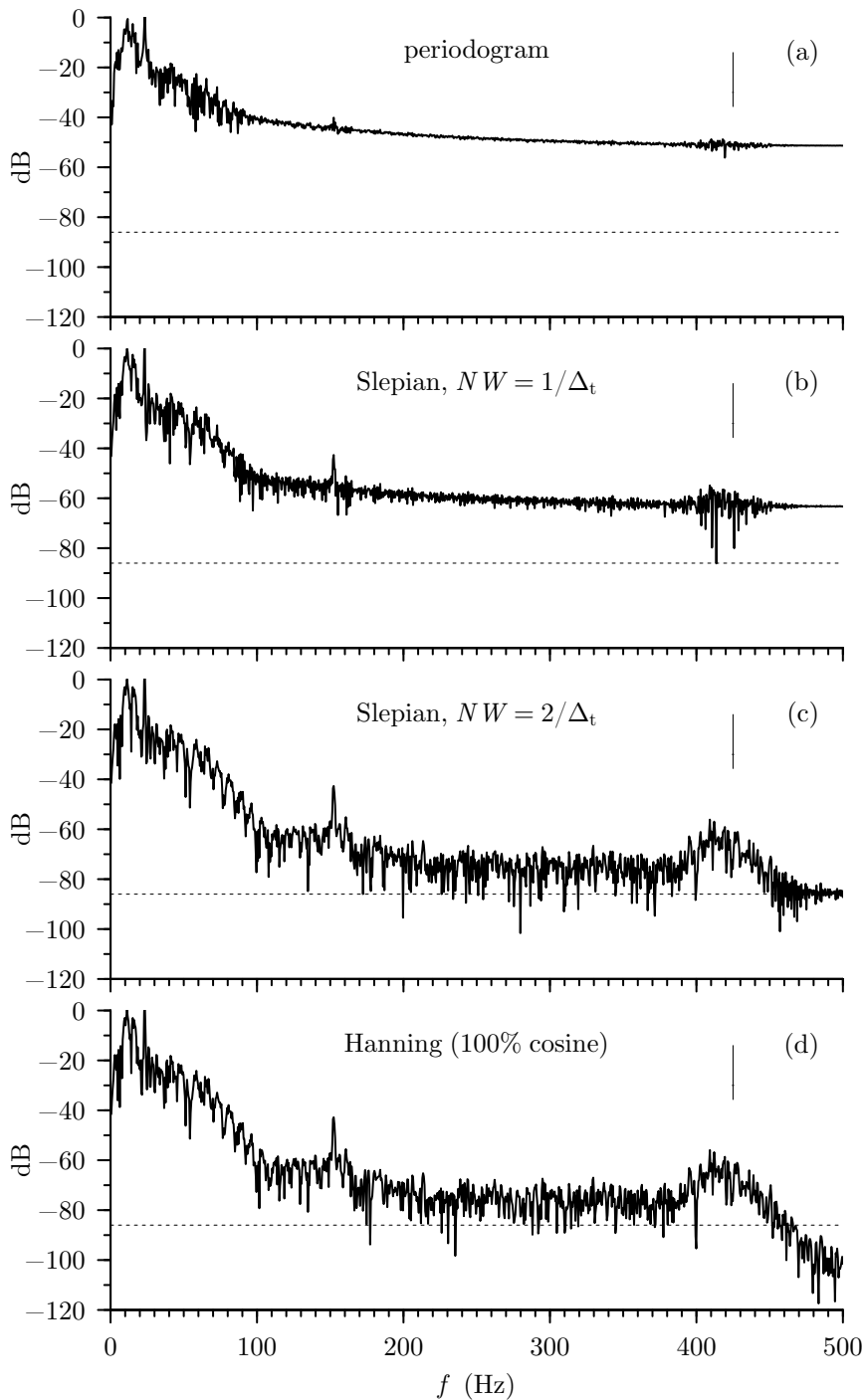
above the local background. This peak evidently corresponds to a 0.7 sec resonance in the wave-follower. That this peak is not attributable to just statistical fluctuations can be assessed by the  $F$ -test for periodicity discussed in Section 10.10 and applied to the ocean wave data in Section 10.15; however, a rough indication that it is significant can be seen by mentally moving the crisscross on the plot to this peak and noting that the lower end of a 95% CI for  $S(1.469)$  is about 10 dB above a reasonable guess at the local level of the SDF. Other than this peak (and a smaller, less significant one at 1.715 Hz), the spectral estimate is relatively flat between 1.0 and 2.0 Hz, a conclusion that does not disagree with factors known to influence this portion of the SDF.

To conclude, this example shows that bias in the periodogram can seriously impact its usefulness as an SDF estimator. Tapering helps reduce the bias, but the sampling variability in the resulting direct spectral estimates is no better than that of the periodogram. Getting a better visual representation of the rate at which the power in the SDF decreases over the frequency range 0.2 to 1.0 Hz requires using SDF estimators with reduced variability. After a discussion of such estimators in the next chapter, we continue our analysis of the ocean wave data in Section 7.12.

### *Chaotic Beam Data*

Our next example is a time series from an experiment designed to capture chaotic motion (Constantine, 1999). The data (Figure 228) are speeds of the bottom of a thin steel beam recorded over time. The beam was suspended vertically by clamping its top to an electromechanical shaker, which subjected the beam to a horizontal excitation through rapid sinusoidal variations, thus causing the bottom of the beam to vibrate. The vertical mounting allowed the beam's bottom to be suspended between two magnets. These magnets provided nonlinear buckling forces as the beam was being excited at a frequency close to its first natural mode of vibration (approximately 10 Hz). The amplitude of the sinusoidal variations was adjusted to yield seemingly chaotic motion. The bottom speeds were recorded at  $\Delta_t = 0.001$  sec intervals for a little over 2 sec ( $N = 2048$  measurements overall). For more details about this time series, see Constantine (1999); for discussions on chaotic dynamics in general, see, e.g., Moon (1992), Acheson (1997) or Chan and Tong (2001).

We start by examining the periodogram  $\{\hat{S}^{(P)}(\cdot)\}$  over the positive Fourier frequencies  $f_k = k/(N \Delta_t)$  (Figure 229(a)). This periodogram is somewhat reminiscent of the one for the ocean wave data (Figure 226(a)): both have fluctuations that are markedly smaller at high frequencies than at low frequencies. As was the case for the ocean wave data, this pattern leads us to suspect that the periodogram for the chaotic beam data suffers from bias due to leakage. By examining direct spectral estimates based upon Slepian data tapers with  $NW = 1/\Delta_t$



**Figure 229.** Direct spectral estimates for chaotic beam data. The (barely visible) width of the crisscrosses gives the bandwidth measure  $B_{\mathcal{H}}$  (Equation (194)), while its height gives the length of a 95% CI for  $10 \log_{10}(S(f))$  (Equation (205b)). The horizontal dashed lines aid in comparing the four estimates (see the text for the rationale for placing them at -86 dB).

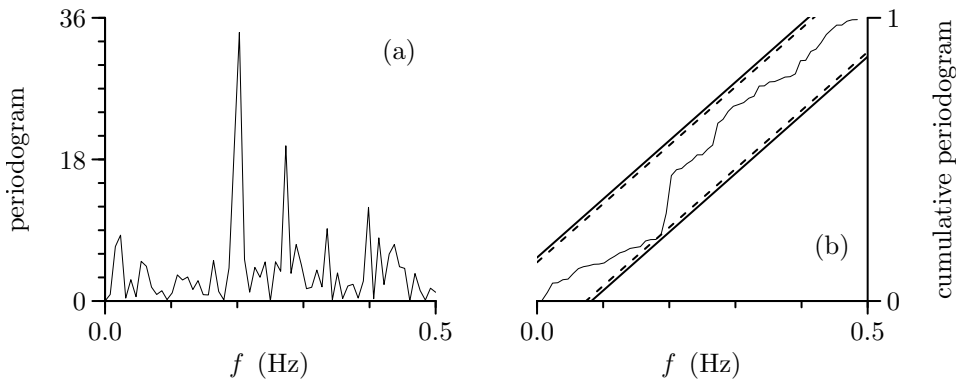
(Figure 229(b)),  $2/\Delta_t$  (229(c)),  $3/\Delta_t$  and  $4/\Delta_t$  (not shown), we find that, while the choice  $NW = 1/\Delta_t$  results in some decrease in bias, there is still leakage over a substantial interval of frequencies; on the other hand, the choice  $NW = 2/\Delta_t$  eliminates most of the bias, but not at frequencies above approximately 450 Hz. The choices  $3/\Delta_t$  and  $4/\Delta_t$  show no obvious signs of bias due to leakage at any frequency. Interestingly enough, the choice of a Hanning data taper leads to a direct spectral estimate (Figure 229(d)) that is virtually the same as the  $3/\Delta_t$  and  $4/\Delta_t$  Slepian-based estimates. While the spectral windows for the Hanning and  $NW = 2/\Delta_t$  Slepian data tapers have central lobes that are almost identical, their sidelobe patterns are markedly different, a fact that Figures 191(d) and (f) illustrate for sample size  $N = 64$ . The sidelobes for the Hanning-based spectral window are smaller at frequencies well separated from the central lobe than they are for the Slepian-based window. Recall that the expected value of a direct spectral estimator is the convolution of the true SDF with the spectral window  $\mathcal{H}(\cdot)$  associated with the data taper (see Equation (186e)). Leakage in a low-power region of the true SDF arises because of an interaction between a high-power region and the sidelobes of  $\mathcal{H}(\cdot)$ . For the chaotic beam data, the high-power region is at low frequencies. For frequencies close to the Nyquist frequency  $f_N = 500$  Hz, the interaction between this region and the sidelobes of  $\mathcal{H}(\cdot)$  far away from the central lobe is evidently not strong enough to produce noticeable leakage in the Hanning-based estimate at high frequencies, but the opposite holds for the Slepian-based estimate. (Prior to forming the Hanning- and Slepian-based direct spectral estimates, we centered the time series by subtracting off its sample mean  $\bar{X} \doteq 10.0$ . We used Equation (189b) for the Hanning data taper and Equation (196b) for the Slepian tapers. We normalized all data tapers  $\{h_t\}$  such that  $\sum_{t=0}^{N-1} h_t^2 = 1$ .)

While it is possible to determine theoretical SDFs for certain chaotic processes (see, e.g., Lopes et al., 1997), the goal of this experiment was not to compare estimated and theoretical SDFs, but rather to investigate the effect of measurement noise on certain statistics of interest in assessing nonlinear dynamics (Constantine, 1999). If we model the chaotic beam time series as a realization of a process defined by  $X_t = Y_t + \epsilon_t$ , where  $\{Y_t\}$  is a stationary process exhibiting chaotic motion and  $\{\epsilon_t\}$  represents measurement noise, and if we assume that  $\{Y_t\}$  and  $\{\epsilon_t\}$  possess SDFs denoted by  $S_Y(\cdot)$  and  $S_\epsilon(\cdot)$  and that  $Y_t$  and  $\epsilon_u$  are uncorrelated for all choices of  $t$  and  $u$ , then the result of Exercise [4.12] says that the SDF for  $\{X_t\}$  is given by  $S_X(f) = S_Y(f) + S_\epsilon(f)$ . If  $\{\epsilon_t\}$  is white noise (often a reasonable model for measurement noise) so that  $S_\epsilon(f) = \sigma_\epsilon^2 \Delta_t$ , where  $\sigma_\epsilon^2$  is the variance of  $\{\epsilon_t\}$ , then we would expect to see  $S_X(\cdot)$  bottom out close to a value of  $\sigma_\epsilon^2 \Delta_t$  (the noise floor) at frequencies where the SDF for measurement noise dominates that for chaotic motion. The Hanning-based direct spectral estimate of Figure 229(d) shows no evidence of a noise floor over any substantial interval of frequencies, indicating that the effect of measurement noise on statistics for assessing nonlinear dynamics might be minimal here.

In conclusion, had we looked at, say, the  $NW = 2/\Delta_t$  Slepian-based direct spectral estimate of Figure 229(c) instead of the Hanning-based estimate, we might have erroneously concluded that a noise floor exists at frequencies higher than approximately 450 Hz at a level of approximately  $-86$  dB (indicated by the horizontal dashed line). We would deduce a much higher noise floor from an uncritical examination of the periodogram in Figure 229(a)). This example shows the key role played by tapering in obtaining SDF estimates with low enough bias to allow proper assessment of the underlying true SDF.

### Comments and Extensions to Section 6.8

[1] In C&E [6] for Section 6.6, we discussed a test for white noise based upon the normalized cumulative periodogram (see Equation (215a)). Here we apply this test to the ocean noise time series (Figure 2(d)), which, for pedagogical purposes, we regarded in Chapter 1 as a realization of a white noise process. Figure 231(a) shows the periodogram for this time series over the grid of Fourier frequencies. The



**Figure 231.** Periodogram (a) and associated normalized cumulative periodogram (b) for the ocean noise series in Figure 2(d). The periodogram is plotted on a linear/linear scale.

jagged curve in Figure 231(b) is the cumulative periodogram, for which the corresponding test statistic  $D$  is 0.1536. This statistic is less than the upper  $\alpha = 0.05$  percentage point  $\tilde{D}(0.05) \doteq 0.1696$  for the distribution of  $D$  under the null hypothesis of white noise, but greater than the  $\alpha = 0.1$  point  $\tilde{D}(0.1) \doteq 0.1528$ . We thus cannot reject the null hypothesis at a level of significance of  $\alpha = 0.05$ , but we can at the higher level of 0.1. The upper and lower solid lines in Figure 231(b) give a visualization of the cumulative periodogram test at the  $\alpha = 0.05$  level of significance. The fact that the cumulative periodogram never crosses either line corresponds to our not being able to reject the null hypothesis at that level of significance. On the other hand, the dashed lines correspond to the  $\alpha = 0.1$  test, and the cumulative periodogram *does* dip below the lower line, but just barely at two frequencies ( $\frac{23}{128} \doteq 0.180$  and  $\frac{24}{128} \doteq 0.188$  Hz). The observed level of significance is quite close to 0.1, which provides cause for caution in regarding the ocean noise as a realization of white noise. (We return to this example in Section 10.15, where we apply a second – arguably more powerful – test for white noise, which casts serious doubt on the white noise hypothesis.)

### 6.9 Comments on Complex-Valued Time Series

All of the previous sections in this chapter assumed that we are dealing with a real-valued time series. Suppose now that our time series is complex-valued and can be regarded as a realization of a portion  $Z_0, Z_1, \dots, Z_{N-1}$  of a complex-valued discrete parameter stationary process  $\{Z_t\}$  with zero mean and SDF  $S(\cdot)$ . If we can make the simplifying assumption that the real and imaginary components of  $\{Z_t\}$  are uncorrelated processes with equal variances, we need only make a surprisingly few minor modifications to the material in the previous sections in order to estimate  $S(\cdot)$ . For example, the biased estimator of the ACVS in Equation (170b) now becomes

$$\hat{s}_\tau^{(P)} = \begin{cases} \sum_{t=0}^{N-\tau-1} Z_{t+\tau} Z_t^* / N, & 0 \leq \tau \leq N-1; \\ (\hat{s}_{-\tau}^{(P)})^*, & -(N-1) \leq \tau < 0; \\ 0, & \text{otherwise.} \end{cases} \quad (231)$$

The main result of Equation (170d) still holds, namely,

$$\Delta_t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(P)} e^{-i2\pi f \tau \Delta_t} = \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} Z_t e^{-i2\pi f t \Delta_t} \right|^2 = \hat{S}^{(P)}(f),$$

but a slightly different proof is required. All of the stated results concerning tapering and prewhitening in Sections 6.4 and 6.5 still hold, with obvious changes to Equations (188b) and (195a). The main results of Section 6.6 concerning the statistical properties of direct spectral

estimators also hold and, in fact, simplify somewhat in that the annoying necessity of special treatment for  $f = 0$  and  $f_N$  disappears; for example, Equation (204b) now becomes

$$\hat{S}^{(D)}(f) \stackrel{d}{=} \frac{S(f)}{2} \chi_2^2 \text{ for } -f_N \leq f \leq f_N$$

to a good approximation for large  $N$ , while Equation (204c) becomes

$$\text{var} \{ \hat{S}^{(D)}(f) \} \approx S^2(f) \text{ for } -f_N \leq f \leq f_N.$$

The grid of frequencies over which distinct  $\hat{S}^{(D)}(f)$  variates are approximately uncorrelated now includes both positive and negative frequencies. The derivation leading up to Equation (211b) is no longer valid since we made use of the fact that  $dZ(-u) = dZ^*(u)$  for real-valued processes; also, we must reformulate the normalized cumulative periodogram of Equation (215a) in an obvious two-sided fashion. Finally, note that we must plot all estimated SDFs over the set of frequencies ranging from  $-f_N$  to  $f_N$  instead of just from 0 to  $f_N$  because the SDF is no longer an even function as in the case of real-valued processes.

The assumption that the real and imaginary components of  $\{Z_t\}$  are uncorrelated with equal variances is key to our ability to make straightforward adjustments to the real-valued theory to handle the complex-valued case. This assumption is often reasonable when, e.g., we create  $\{Z_t\}$  via complex demodulation and low-pass filtering of a real-valued process (for details, see Bloomfield, 2000, Hasan, 1983, or Stoica and Moses, 2005); however, examples abound of complex-valued processes for which this assumption is *not* reasonable, rendering the corresponding complex-valued theory markedly different from the real-valued theory (Picinbono and Bondon, 1997, and Schreier and Scharf, 2003). One example is a process whose real and imaginary components are geometrically related to motion taking place in a plane. Here the physics governing the motion can dictate a nonzero correlation between the components. For details about these types of complex-valued time series, see Schreier and Scharf (2010, chapter 10) and Walden (2013).

### 6.10 Summary of Periodogram and Other Direct Spectral Estimators

Given a time series that we can regard as a portion  $X_0, X_1, \dots, X_{N-1}$  of a real-valued stationary process  $\{X_t\}$  with a sampling interval of  $\Delta_t$  and with an autocovariance sequence (ACVS)  $\{s_{X,\tau}\}$  and spectral density function (SDF)  $S_X(\cdot)$  related by  $\{s_{X,\tau}\} \longleftrightarrow S_X(\cdot)$ , we can use the fact that

$$S_X(f) = \Delta_t \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f \tau \Delta_t} \quad (232)$$

to formulate an estimator for  $S_X(\cdot)$  based upon an estimator of  $\{s_{X,\tau}\}$  (see Equation (163)). Under the simplifying assumption that  $\{X_t\}$  has zero mean, we can estimate  $s_{X,\tau}$  for  $|\tau| < N$  using the so-called biased ACVS estimator

$$\hat{s}_{X,\tau}^{(P)} = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} X_{t+|\tau|} X_t \quad (\text{see (170b)})$$

(when, as is usually the case, we cannot assume a priori that  $\{X_t\}$  has zero mean, we must center the time series by subtracting off its sample mean  $\bar{X} = \frac{1}{N} \sum_{t=0}^{N-1} X_t$  so that the summand  $X_{t+|\tau|} X_t$  becomes  $(X_{t+|\tau|} - \bar{X})(X_t - \bar{X})$ ). If we define  $\hat{s}_{X,\tau}^{(P)} = 0$  for  $|\tau| \geq N$ , the resulting estimator  $\{\hat{s}_{X,\tau}^{(P)}\}$  is always a positive semidefinite sequence, which means that it is a valid theoretical ACVS; i.e., it is the ACVS for a well-defined stationary process. Substituting  $\{\hat{s}_{X,\tau}^{(P)}\}$  for  $\{s_{X,\tau}\}$  in Equation (232) yields the SDF estimator known as the periodogram:

$$\hat{S}_X^{(P)}(f) \stackrel{\text{def}}{=} \Delta_t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{X,\tau}^{(P)} e^{-i2\pi f \tau \Delta_t}. \quad (\text{see (170c)})$$

Another formulation for the periodogram that is fully equivalent to the above is

$$\hat{S}_X^{(P)}(f) = \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi f t \Delta_t} \right|^2. \quad (\text{see (170d)})$$

Just as the theoretical ACVS and SDF satisfy  $\{s_{X,\tau}\} \longleftrightarrow S_X(\cdot)$ , we have  $\{\hat{s}_{X,\tau}^{(P)}\} \longleftrightarrow \hat{S}_X^{(P)}(\cdot)$ . In this Fourier transform pair, the inverse relationship tells us that

$$\hat{s}_{X,0}^{(P)} = \int_{-f_N}^{f_N} \hat{S}_X^{(P)}(f) df, \quad \text{which mimics } s_{X,0} = \int_{-f_N}^{f_N} S_X(f) df. \quad (\text{see (171d)})$$

Thus, just as  $S_X(\cdot)$  is an analysis of the process variance  $s_{X,0}$ , so is  $\hat{S}_X^{(P)}(\cdot)$  an analysis of the sample variance  $\hat{s}_{X,0}^{(P)}$  across a continuum of frequencies. We also have

$$\{\hat{s}_{X,\tau}^{(P)} : \tau = -(N-1), \dots, N-1\} \longleftrightarrow \{\hat{S}_X^{(P)}(f'_k) : k = -(N-1), \dots, N-1\}, \quad (\text{see (171e)})$$

where  $f'_k \stackrel{\text{def}}{=} k / [(2N-1) \Delta_t]$ . The inverse Fourier transform (Equation (171f)) leads to

$$\hat{s}_{X,0}^{(P)} = \frac{1}{(2N-1) \Delta_t} \sum_{k=-(N-1)}^{N-1} \hat{S}_X^{(P)}(f'_k),$$

and hence  $\{\hat{S}_X^{(P)}(f'_k)\}$  is an analysis of  $\hat{s}_{X,0}^{(P)}$  across a discrete set of frequencies. More generally, for any  $N' \geq N$  and with  $f'_k$  redefined to be  $k / (N' \Delta_t)$ , we have

$$\hat{s}_{X,0}^{(P)} = \frac{1}{N' \Delta_t} \sum_{k=0}^{N'-1} \hat{S}_X^{(P)}(f'_k). \quad (\text{see (237b)})$$

The statistical properties of the periodogram are unfortunately such as to limit its value as an estimator of the SDF. Turning first to its first-moment properties, we have

$$E\{\hat{S}_X^{(P)}(f)\} = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_{X,\tau} = \int_{-f_N}^{f_N} \mathcal{F}(f-f') S_X(f') df', \quad (233)$$

where

$$\mathcal{F}(f) \stackrel{\text{def}}{=} \Delta_t N \mathcal{D}_N^2(f \Delta_t)$$

is Fejér's kernel, which is defined in terms of Dirichlet's kernel  $\mathcal{D}_N(\cdot)$  (see Equations (174a), (174b) and (174c)). As  $N \rightarrow \infty$ , this kernel acts as a Dirac delta function, and hence, for  $N$  large enough,  $E\{\hat{S}_X^{(P)}(f)\} \approx S_X(f)$ ; however, for certain processes, this approximation is not accurate even for samples sizes that superficially seem large. The periodogram is badly biased for finite  $N$  when there is a deleterious interaction in the integrand of Equation (233) between a highly structured  $S_X(\cdot)$  and the sidelobes of Fejér's kernel (these sidelobes are in evidence in Figure 176). In other words,  $\mathcal{F}(\cdot)$  fails to act as a decent approximation to a delta function in Equation (233) because it is not sufficiently concentrated around  $f - f' = 0$ . This form of bias goes by the name of leakage.

Two approaches for dealing with leakage are tapering and prewhitening. Tapering consists of multiplying the values of a time series by a data taper  $\{h_t : t = 0, 1, \dots, N-1\}$ , a suitably chosen sequence of real-valued constants usually normalized such that  $\sum_t h_t^2 = 1$  (when needed, we define  $h_t$  to be zero for  $t < 0$  and  $t \geq N$ ). Tapering leads to the family of so-called direct spectral estimators:

$$\hat{S}_X^{(D)}(f) \stackrel{\text{def}}{=} \Delta_t \left| \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi f t \Delta_t} \right|^2 \quad (\text{see (186b)})$$

(note that use of the rectangular or default taper  $h_t = 1/\sqrt{N}$  turns  $\hat{S}_X^{(D)}(\cdot)$  back into the periodogram). In contrast to the periodogram, we now have

$$\begin{aligned} E\{\hat{S}_X^{(D)}(f)\} &= \Delta_t \sum_{\tau=-(N-1)}^{N-1} \left( \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t \right) s_{X,\tau} e^{-i2\pi f \tau \Delta_t} \\ &= \int_{-f_N}^{f_N} \mathcal{H}(f - f') S_X(f') df', \end{aligned} \quad (234)$$

where

$$\mathcal{H}(f) \stackrel{\text{def}}{=} \Delta_t \left| \sum_{t=0}^{N-1} h_t e^{-i2\pi f t \Delta_t} \right|^2$$

is the spectral window associated with  $\{h_t\}$  (see Equations (186g), (186e) and (186f)). For certain processes and sample sizes for which the periodogram suffers from unacceptable leakage, we can have  $E\{\hat{S}_X^{(D)}(f)\} \approx S_X(f)$  to a good approximation for a suitably chosen data taper. Examples of suitable tapers include the Hanning data taper

$$h_t = \begin{cases} C \left[ 1 - \cos \left( \frac{\pi(t+1)}{\lfloor N/2 \rfloor + 1} \right) \right], & 0 \leq t \leq \lfloor \frac{N+1}{2} \rfloor - 1; \\ h_{N-t-1}, & \lfloor \frac{N+1}{2} \rfloor \leq t \leq N-1, \end{cases} \quad (\text{see (189b)})$$

and tapers based on an approximation to zeroth-order Slepian sequences, namely,

$$h_t = C I_0 \left( \pi W (N-1) \left( 1 - \frac{(2t+1-N)^2}{N^2} \right)^{1/2} \right) \quad (\text{see (196b)})$$

(in the above equations,  $C$  is a generic constant enforcing the normalization  $\sum_t h_t^2 = 1$ ;  $\lfloor x \rfloor$  is the largest integer less than or equal to  $x$ ;  $I_0(\cdot)$  is the modified Bessel function of the first kind and zeroth order; and  $W$  defines the interval of frequencies  $[-W, W]$  over which the spectral window  $\mathcal{H}(\cdot)$  is maximally concentrated – this is usually specified indirectly by setting  $NW$ , for which typical values are 1, 2, 4 and 8).

Tapering attempts to compensate for leakage in the periodogram by replacing Fejér's kernel in Equation (233) with a spectral window  $\mathcal{H}(\cdot)$  that is effectively a better approximation to a delta function (cf. Equations (233) and (234)). Prewhitening takes the complementary approach of replacing a highly structured SDF with one having less structure. The idea is to apply a prewhitening filter  $\{g_l : l = 0, 1, \dots, L-1\}$  to the time series, yielding a new series

$$Y_t \stackrel{\text{def}}{=} \sum_{l=0}^{L-1} g_l X_{t+L-1-l}, \quad 0 \leq t \leq N-L. \quad (\text{see (197a)})$$

The new series is a realization of a stationary process with an SDF given by

$$S_Y(f) = \left| \sum_{l=0}^{L-1} g_l e^{-i2\pi f l \Delta_t} \right|^2 S_X(f). \quad (\text{see (197b)})$$

If  $\{g_l\}$  is chosen such that  $S_Y(\cdot)$  is less highly structured than  $S_X(\cdot)$ , the periodogram  $\hat{S}_Y^{(P)}(\cdot)$  for the  $Y_t$ 's might have acceptable bias properties, in which case we can entertain a “postcolored” estimate of  $S_X(\cdot)$ , namely,

$$\hat{S}_X^{(PC)}(f) = \frac{\hat{S}_Y^{(P)}(f)}{\left| \sum_{l=0}^{L-1} g_l e^{-i2\pi f l \Delta_t} \right|^2}. \quad (\text{cf. (198)})$$



Turning now to statistical properties other than the first moment, we can state that, subject to certain technical assumptions,

$$\hat{S}_X^{(D)}(f) \stackrel{d}{=} \begin{cases} S_X(f)\chi_{\nu}^2/2, & 0 < |f| < f_N; \\ S_X(f)\chi_1^2, & |f| = 0 \text{ or } f_N, \end{cases} \quad (\text{see (204b)})$$

asymptotically as  $N \rightarrow \infty$  for all direct spectral estimators (including the periodogram). Here  $\stackrel{d}{=}$  means “equal in distribution,” and  $\chi_{\nu}^2$  is a chi-square random variable (RV) with  $\nu$  degrees of freedom (thus the statement  $U \stackrel{d}{=} c\chi_{\nu}^2$  means that the RV  $U$  has the same distribution as a chi-square RV after multiplication by a constant  $c$ ). While the univariate properties of direct spectral estimators (including the periodogram) are the same, their bivariate properties are taper-dependent. Focusing on the periodogram first and considering the subset of Fourier frequencies  $f_k \stackrel{\text{def}}{=} k/(N\Delta_t)$  satisfying  $0 \leq f_k \leq f_N$ , we can claim that, under suitable conditions, the  $\lfloor N/2 \rfloor + 1$  RVs  $\hat{S}_X^{(P)}(f_0), \hat{S}_X^{(P)}(f_1), \dots, \hat{S}_X^{(P)}(f_{\lfloor N/2 \rfloor})$  are all approximately pairwise uncorrelated for  $N$  large enough. A complementary claim is that, if  $0 \leq f < f' \leq f_N$ , then

$$\text{cov}\{\hat{S}_X^{(P)}(f), \hat{S}_X^{(P)}(f')\} \approx 0 \text{ as long as } |f - f'| \geq \frac{1}{N\Delta_t},$$

i.e., a spacing greater than between adjacent Fourier frequencies. For other direct spectral estimators, the corresponding statement is that

$$\text{cov}\{\hat{S}_X^{(D)}(f), \hat{S}_X^{(D)}(f')\} \approx 0 \text{ as long as } |f - f'| \geq B_{\mathcal{H}},$$

where  $B_{\mathcal{H}}$  is our standard measure of the effective bandwidth of  $\hat{S}_X^{(D)}(\cdot)$ :

$$B_{\mathcal{H}} \stackrel{\text{def}}{=} \frac{\left(\int_{-f_N}^{f_N} \mathcal{H}(f) df\right)^2}{\int_{-f_N}^{f_N} \mathcal{H}^2(f) df} = \frac{\Delta_t}{\sum_{\tau=-(N-1)}^{N-1} (h \star h_{\tau})^2} \quad (\text{see (194)})$$

(here  $\{h \star h_{\tau}\}$  is the autocorrelation of  $\{h_t\}$  defined as per Equation (99f)).

An important aspect of the large-sample distribution for direct spectral estimators is its lack of dependence on the sample size  $N$ . Since the variance of  $\chi_{\nu}^2$  is  $2\nu$ , the distributional properties of  $\hat{S}_X^{(D)}(\cdot)$  imply that

$$\text{var}\{\hat{S}_X^{(D)}(f)\} = \begin{cases} S_X^2(f), & 0 < |f| < f_N; \\ 2S_X^2(f), & |f| = 0 \text{ or } f_N. \end{cases} \quad (\text{see (204c)})$$

Ignoring the pathological case  $S_X(f) = 0$ , neither the periodogram nor any other direct spectral estimator is such that  $\text{var}\{\hat{S}_X^{(D)}(f)\} \rightarrow 0$  as  $N \rightarrow \infty$ . In statistical parlance, these SDF estimators are *inconsistent* because they do not get closer and closer to the true SDF with high probability as the sample size gets arbitrarily large. While direct spectral estimators are fundamentally flawed in this aspect, they are the basic building blocks for consistent SDF estimators to be discussed in subsequent chapters.

## 6.11 Exercises

[6.1] Suppose  $\{\epsilon_t\}$  is a white noise process with mean zero and variance  $0 < \sigma_{\epsilon}^2 < \infty$ , and consider the following three stationary processes constructed from it:

- (a)  $X_t = \mu + \epsilon_0$ ;
- (b)  $X_t = \mu + \epsilon_t + \epsilon_{t-1}$ ; and
- (c)  $X_t = \mu + \epsilon_t - \epsilon_{t-1}$ .

Let  $\{s_{\tau}\}$  be the ACVS for any one of the three processes, and let  $\bar{X} = \frac{1}{N} \sum_{t=0}^{N-1} X_t$  be the sample mean based upon a portion of one of the processes. For all three processes, express  $\bar{X}$  as a linear combination of RVs from  $\{\epsilon_t\}$ , and determine  $\text{var}\{\bar{X}\}$  directly with this expression. In each case,

verify that the equation for  $\text{var}\{\bar{X}\}$  is in agreement with Equation (165a). Equations (165b) and (165c) suggest the approximation

$$\text{var}\{\bar{X}\} \approx \frac{1}{N} \sum_{\tau=-\infty}^{\infty} s_{\tau}.$$

Comment on how well the above works in each case.

- [6.2] Suppose that  $X_0, \dots, X_{N-1}$  is a sample of size  $N$  from a white noise process with unknown mean  $\mu$  and variance  $\sigma^2$ . With  $\hat{s}_{\tau}^{(U)}$  and  $\hat{s}_{\tau}^{(P)}$  defined as in Equations (166a) and (166b), show that, for  $0 < |\tau| \leq N-1$ ,

$$E\{\hat{s}_{\tau}^{(U)}\} = -\frac{\sigma^2}{N} \quad \text{and} \quad E\{\hat{s}_{\tau}^{(P)}\} = -\left(1 - \frac{|\tau|}{N}\right) \frac{\sigma^2}{N}.$$

Use this to argue (a) that the magnitude of the bias of the biased estimator  $\hat{s}_{\tau}^{(P)}$  can be less than that of the “unbiased” estimator  $\hat{s}_{\tau}^{(U)}$  when  $\mu$  is estimated by  $\bar{X}$  and (b) that, for this example, the mean square error of  $\hat{s}_{\tau}^{(P)}$  is less than that of  $\hat{s}_{\tau}^{(U)}$  for  $0 < |\tau| \leq N-1$ .

- [6.3] Suppose that the RVs  $X_0$  and  $X_1$  are part of a real-valued Gaussian stationary process  $\{X_t\}$  with unknown mean  $\mu$  and unknown ACVS  $\{s_{\tau}\}$ , which is arbitrary except for the mild restriction  $s_1 < s_0$ . Based upon just these two RVs, consider the so-called unbiased and biased estimators of  $s_1$  dictated by, respectively, Equations (166a) and (166b):

$$\hat{s}_1^{(U)} = (X_0 - \bar{X})(X_1 - \bar{X}) \quad \text{and} \quad \hat{s}_1^{(P)} = \frac{1}{2}(X_0 - \bar{X})(X_1 - \bar{X}), \quad \text{where} \quad \bar{X} = \frac{X_0 + X_1}{2}.$$

- For this special case of  $N = 2$ , determine  $E\{\hat{s}_1^{(U)}\}$  and  $E\{\hat{s}_1^{(P)}\}$ , and use these to derive the biases of  $\hat{s}_1^{(U)}$  and  $\hat{s}_1^{(P)}$  as estimators of  $s_1$ . How does the squared bias of  $\hat{s}_1^{(U)}$  compare with that of  $\hat{s}_1^{(P)}$ ? Under what conditions (if any) are  $\hat{s}_1^{(U)}$  and  $\hat{s}_1^{(P)}$  unbiased?
- Determine the variances of  $\hat{s}_1^{(U)}$  and  $\hat{s}_1^{(P)}$ . How does  $\text{var}\{\hat{s}_1^{(U)}\}$  compare with  $\text{var}\{\hat{s}_1^{(P)}\}$ ? Hint: recall the Isserlis theorem (Equation (30)).
- Using the results of parts (a) and (b), verify that, for this special case of  $N = 2$ , the inequality of Equation (167a) always holds:

$$\text{MSE}\{\hat{s}_1^{(P)}\} = E\{(\hat{s}_1^{(P)} - s_1)^2\} < E\{(\hat{s}_1^{(U)} - s_1)^2\} = \text{MSE}\{\hat{s}_1^{(U)}\}.$$

Investigate the relative contribution of  $\text{var}\{\hat{s}_1^{(U)}\}$  to  $\text{MSE}\{\hat{s}_1^{(U)}\}$ , and do the same for  $\hat{s}_1^{(P)}$ .

- Calculate  $\mathbf{P}[\hat{s}_1^{(U)} < -1]$  when  $\text{var}\{X_0\} = 4/5$  and  $\rho_1 \stackrel{\text{def}}{=} s_1/s_0 = 1/2$ , and calculate  $\mathbf{P}[\hat{s}_1^{(U)} > 1]$  also. Hint: if  $Z$  is a Gaussian RV with zero mean and unit variance, then  $Z^2$  obeys a chi-square distribution with one degree of freedom.
- With  $\hat{s}_{\tau}^{(U)}$  and  $\hat{s}_{\tau}^{(P)}$  being dictated by Equations (166a) and (166b) for  $\tau = 0$  in addition to  $\tau = 1$ , verify that, for the special case of  $N = 2$ ,

$$\hat{\rho}_1^{(U)} \stackrel{\text{def}}{=} \frac{\hat{s}_1^{(U)}}{\hat{s}_0^{(U)}} = -1 \quad \text{and} \quad \hat{\rho}_1^{(P)} \stackrel{\text{def}}{=} \frac{\hat{s}_1^{(P)}}{\hat{s}_0^{(P)}} = -\frac{1}{2}$$

no matter what values the RVs  $X_0$  and  $X_1$  assume (we can ignore the special case  $X_0 = X_1 = 0$  since it has probability zero). Thus, while part (d) indicates that realizations of  $\hat{s}_1^{(U)}$  can take on different values, those of  $\hat{\rho}_1^{(U)}$  cannot.

- [6.4] Equations (174d) and (175a) show two ways of expressing Fejér’s kernel. Verify the following third expression:

$$\mathcal{F}(f) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) e^{-i2\pi f \tau \Delta_t}. \quad (236)$$

- [6.5] Let  $X_0, \dots, X_{N-1}$  be a sample of size  $N$  from a stationary process with an unknown mean and an SDF  $S(\cdot)$  defined over  $[-1/2, 1/2]$  (i.e., the sampling interval  $\Delta_t$  is taken to be 1). At lag  $\tau = 0$  both the unbiased and biased estimators of the ACVS reduce to

$$\hat{s}_0 = \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2.$$

- (a) Show that  $E\{\hat{s}_0\} = s_0 - \text{var}\{\bar{X}\} \leq s_0$ .  
 (b) Show that

$$E\{\hat{s}_0\} = \int_{-1/2}^{1/2} \left(1 - \frac{1}{N} \mathcal{F}(f)\right) S(f) df,$$

where  $\mathcal{F}(\cdot)$  is Fejér's kernel (see Equation (174c)). Hint: consider Equations (165a), (174a) and (174b).

- (c) Plot  $(1 - \mathcal{F}(f)/N)$  versus  $f$  for, say,  $N = 64$ . Based upon this plot, for what kind of  $S(\cdot)$  would there be a large discrepancy between  $E\{\hat{s}_0\}$  and  $s_0$ ?

- [6.6] Construct an example of a time series of one specific length  $N$  for which the unbiased estimate of the ACVS

$$\hat{s}_{-(N-1)}^{(U)}, \hat{s}_{-(N-2)}^{(U)}, \dots, \hat{s}_{-1}^{(U)}, \hat{s}_0^{(U)}, \hat{s}_1^{(U)}, \dots, \hat{s}_{N-2}^{(U)}, \hat{s}_{N-1}^{(U)}$$

given by Equation (166a) is *not* a valid ACVS for some stationary process (here  $\hat{s}_\tau^{(U)} \stackrel{\text{def}}{=} 0$  for  $|\tau| \geq N$ ). Note:  $N$  can be chosen in whatever manner so desired.

- [6.7] Let  $X_0, \dots, X_{N-1}$  be a portion of a stationary process with sampling interval  $\Delta_t$ , and let  $\hat{s}_\tau^{(P)}$  for  $\tau = 0, \pm 1, \dots, \pm(N-1)$  be the biased estimator of the ACVS defined by either Equation (170b) – if  $E\{X_t\}$  is known to be zero – or Equation (166b) – if  $E\{X_t\}$  is unknown and hence estimated by the sample mean  $\bar{X}$ . For  $|\tau| \geq N$ , let  $\hat{s}_\tau^{(P)} \stackrel{\text{def}}{=} 0$ . Let  $f'_k = k/(N' \Delta_t)$ , where  $N'$  is any integer greater than or equal to  $N$ .

- (a) Show that

$$\hat{S}^{(P)}(f'_k) = \Delta_t \sum_{\tau=0}^{N'-1} (\hat{s}_\tau^{(P)} + \hat{s}_{N'-\tau}^{(P)}) e^{-i2\pi k\tau/N'}$$

and hence that

$$\hat{s}_\tau^{(P)} + \hat{s}_{N'-\tau}^{(P)} = \frac{1}{N' \Delta_t} \sum_{k=0}^{N'-1} \hat{S}^{(P)}(f'_k) e^{i2\pi k\tau/N'} \quad (237a)$$

(Bloomfield, 2000, section 8.4).

- (b) Show that, when  $N' = 2N - 1$ , part (a) can be manipulated to yield Equation (171f). Argue that, if  $N' \geq 2N - 2$ , we can still recover  $\{\hat{s}_\tau^{(P)}\}$  from the inverse DFT of  $\{\hat{S}^{(P)}(f'_k)\}$ , but that, if  $N' < 2N - 2$ , we cannot do so in general.  
 (c) Use part (a) to note that, for any  $N' \geq N$ ,

$$\hat{s}_0^{(P)} = \frac{1}{N' \Delta_t} \sum_{k=0}^{N'-1} \hat{S}^{(P)}(f'_k). \quad (237b)$$

- (d) A DFT or inverse DFT of a sequence with  $N'$  terms often can be computed efficiently using an FFT algorithm (see Section 3.11). Such an algorithm typically requires a number of arithmetic operations (additions, subtractions, multiplications and divisions) proportional to  $N' \log_2(N')$ . Compare the computational procedure for evaluating  $\hat{s}_\tau^{(P)}$  for  $\tau = 0, \dots, N - 1$  via Equation (171f) with direct use of Equation (166b) in terms of the number of arithmetic operations required.

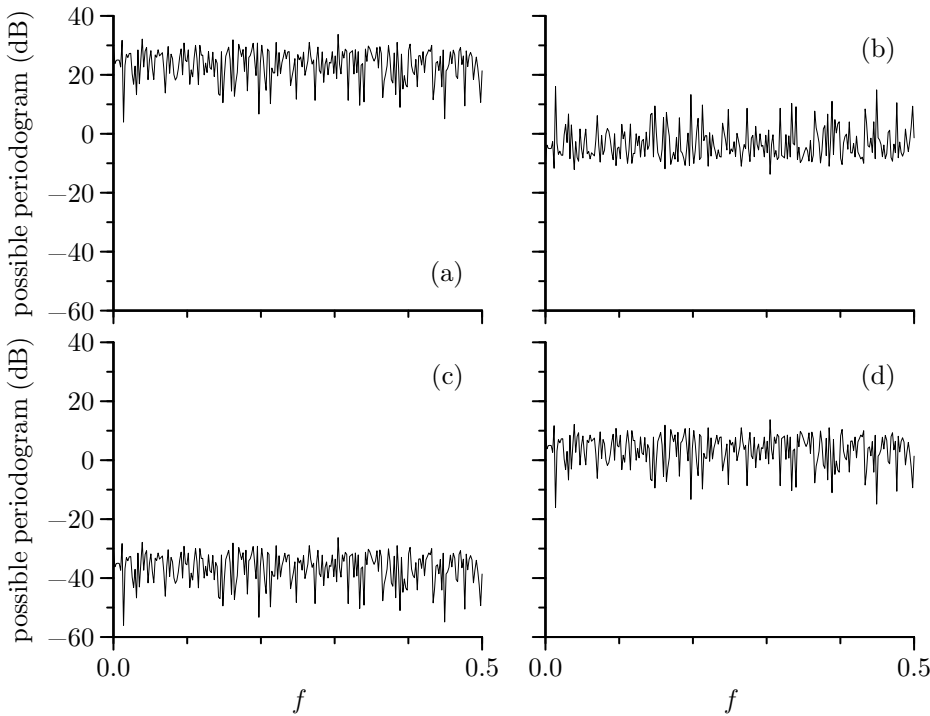
- [6.8] From Equation (170d) we know that, given  $\hat{s}_0^{(P)}, \dots, \hat{s}_{N-1}^{(P)}$ , we can compute the periodogram  $\hat{S}^{(P)}(\cdot)$  at any given frequency. Suppose that we are given the periodogram on the grid of Fourier frequencies, i.e.,  $\hat{S}^{(P)}(f_k)$  for  $k = 0, 1, \dots, \lfloor N/2 \rfloor$ , where, as usual,  $f_k = k/(N \Delta_t)$  and  $\lfloor N/2 \rfloor$  is the largest integer less than or equal to  $N/2$ . Using these  $\lfloor N/2 \rfloor + 1$  values, is it possible in general to recover  $\hat{s}_0^{(P)}, \dots, \hat{s}_{N-1}^{(P)}$  in some manner (i.e., not necessarily via an inverse DFT of  $\{\hat{S}^{(P)}(f_k)\}$ )?
- [6.9] Consider the sequence  $\{\hat{s}_\tau^{(P)} : \tau = -(N-1), \dots, N-1\}$ , which contains the biased estimator of the ACVS. Note that, because  $\hat{s}_{-\tau}^{(P)} = \hat{s}_\tau^{(P)}$ , there are only  $N$  distinct variables in this sequence of length  $2N-1$ . Equation (171e) says that the DFT of this sequence yields the periodogram over the grid of frequencies  $f'_k = k/[(2N-1) \Delta_t]$ ,  $k = -(N-1), \dots, N-1$ . Let  $\{\hat{S}_k : k = 0, \dots, N-1\}$  be the DFT of  $\{\hat{s}_\tau^{(P)} : \tau = 0, \dots, N-1\}$ . Show that we can form the periodogram over the grid of Fourier frequencies  $f_k = k/(N \Delta_t)$  directly in terms of the  $\hat{S}_k$  variables. Use this formulation to argue that, given the periodogram at just the Fourier frequencies, it is not in general possible to recover  $\{\hat{s}_\tau^{(P)} : \tau = 0, \dots, N-1\}$  (you may assume the validity of Equation (237b)).
- [6.10] The intent of this exercise is to verify some key results about the periodogram using a simple example. Let  $a$  be any real-valued nonzero constant, and suppose that  $\{a, 0, -a\}$  is a realization of length  $N = 3$  of a portion  $X_0, X_1, X_2$  of a stationary process with sampling interval  $\Delta_t = 1$ , unknown mean  $\mu$ , SDF  $S(\cdot)$  and ACVS  $\{s_\tau\}$ .
- Determine the biased estimator  $\{\hat{s}_\tau^{(P)}\}$  of the ACVS for  $\{a, 0, -a\}$ . For comparison, also determine the unbiased estimator  $\{\hat{s}_\tau^{(U)}\}$  of the ACVS. Is  $\{\hat{s}_\tau^{(U)}\}$  necessarily a valid ACVS for some stationary process? Explain why or why not.
  - Determine the periodogram  $\hat{S}^{(P)}(\cdot)$  for  $\{a, 0, -a\}$  using its definition in Equation (170c) as the Fourier transform of  $\{\hat{s}_\tau^{(P)}\}$ .
  - An equivalent way of obtaining the periodogram is via Equation (170d), which is based on the DFT of  $\{a, 0, -a\}$ . Verify that computing the periodogram in this alternative manner gives the same result as in part (b).
  - As noted in Equations (171c) and (171f), the biased estimator of the ACVS can be recovered from the periodogram via either an integral of  $\hat{S}^{(P)}(\cdot)$  over  $[-1/2, 1/2]$  or a summation involving  $\hat{S}^{(P)}(f'_k)$ , where  $f'_k = k/(2N-1)$ . Verify that these two equations hold for the example considered here when  $\tau = 0$  and  $\tau = 2$ .
  - Based upon Equation (181), show that the biased estimator of the ACVS can also be recovered from a summation involving  $\hat{S}^{(P)}(\tilde{f}_k)$ , where  $\tilde{f}_k = k/(2N)$ . Verify that the equation expressing  $\hat{s}_\tau^{(P)}$  in terms of the  $\hat{S}^{(P)}(\tilde{f}_k)$ 's holds for the example considered here when  $\tau = 0$  and  $\tau = 2$ .
- [6.11] Let  $x_0, x_1, \dots, x_{N-1}$  represent a particular time series of interest. Let  $\hat{S}_x^{(P)}(\cdot)$  be the periodogram for  $\{x_t\}$ . Define two new time series  $\{v_t\}$  and  $\{w_t\}$  via  $v_t = -10x_t$  and  $w_t = cx_t$ , where  $c$  is a real-valued constant such that  $w_t \neq x_t$  and  $w_t \neq v_t$ . Let  $\hat{S}_v^{(P)}(\cdot)$  and  $\hat{S}_w^{(P)}(\cdot)$  denote the periodograms for  $\{v_t\}$  and  $\{w_t\}$ . The periodograms for  $\{x_t\}$ ,  $\{v_t\}$  and  $\{w_t\}$  are each shown in one of the four plots making up Figure 239 (as usual, the periodograms are displayed on a decibel scale). Which plot corresponds to which periodogram, and what value(s) can the constant  $c$  assume? Briefly explain the rationale for your choices.
- [6.12] Suppose that  $x_0, x_1, \dots, x_{N-1}$  represents one particular time series and that we form its periodogram using

$$\hat{S}_x^{(P)}(f) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x_t e^{-i2\pi f t} \right|^2.$$

For any integer  $0 < m \leq N-1$ , define  $y_t = x_{t+m \bmod N}$ ; i.e.,  $y_0, y_1, \dots, y_{N-1}$  is a circularly shifted version of  $x_0, x_1, \dots, x_{N-1}$  (see Exercise [2.19] for a discussion of the “mod” operator). Let

$$\hat{S}_y^{(P)}(f) = \frac{1}{N} \left| \sum_{t=0}^{N-1} y_t e^{-i2\pi f t} \right|^2$$

be the periodogram for  $y_t$ .



**Figure 239.** Plots of periodograms for a time series  $\{x_t\}$  and for two rescaled versions given by  $v_t = -10x_t$  and  $w_t = cx_t$ , along with a spurious plot.

- (a) Show that  $\hat{S}_y^{(P)}(f_k) = \hat{S}_x^{(P)}(f_k)$ , where  $f_k = k/N$  is a Fourier frequency.  
 (b) Prove or disprove the claim that  $\hat{S}_y^{(P)}(f) = \hat{S}_x^{(P)}(f)$  at all frequencies (i.e.,  $f$  need not be a Fourier frequency).

[6.13] Let  $X_0, X_1, \dots, X_{N-1}$  be a portion of a zero mean stationary process with SDF  $S(\cdot)$  and sampling interval  $\Delta_t$ , from which we form a periodogram  $\hat{S}^{(P)}(f)$ . Let

$$b(f) \stackrel{\text{def}}{=} E\{\hat{S}^{(P)}(f)\} - S(f)$$

represent the bias in the periodogram at frequency  $f$ . Recall that, if  $g(\cdot)$  is a function defined over the interval  $[a, b]$ , then, by definition,

$$\frac{1}{b-a} \int_a^b g(x) dx \quad (239)$$

is the average value of  $g(\cdot)$  over  $[a, b]$ .

- (a) Show that the average value of the bias in the periodogram over  $f \in [-f_N, f_N]$  is zero.  
 (b) Show that the average value of the bias in the periodogram over  $f \in [0, f_N]$  is also zero.
- [6.14] For each of the four AR(4) time series of length  $N = 1024$  displayed in Figures 34(e), (f), (g) and (h), compute the periodogram after subjecting the series to the end-point matching dictated by Equation (183a), followed by centering the end-point matched series by subtracting off its sample mean (the four time series are downloadable from the “Data” part of the website for the book). Compute each periodogram in two ways, namely, across just the nonzero Fourier frequencies  $f_k = k/N$  and across a grid of nonzero frequencies  $\tilde{f}_k = k/(2N)$  whose spacing is twice as fine as the Fourier frequencies. Plot your results along with the true AR(4) SDF on a decibel scale for comparison with Figures 173(e), (f), (g) and (h) and Figures 182(e), (f), (g) and (h). The goal of

end-point matching is to create an SDF estimator with less bias than in the periodogram for the original unaltered series. Comment on how well the goal is achieved for the four AR(4) series.

- [6.15] (a) Given a time series  $Y_0, \dots, Y_{N-1}$ , show that, if  $C$  does not depend on  $t$ ,

$$\tilde{S}^{(P)}(f_k) \stackrel{\text{def}}{=} \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} (Y_t - C) e^{-i2\pi f_k t \Delta_t} \right|^2 = \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} Y_t e^{-i2\pi f_k t \Delta_t} \right|^2, \quad (240a)$$

where  $k$  is an integer,  $f_k = k/(N \Delta_t)$  and  $0 < f_k \leq f_N$ .

- (b) Show that, in the above,  $\tilde{S}^{(P)}(0) = 0$  when  $C = \bar{Y} = \frac{1}{N} \sum_{t=0}^{N-1} Y_t$ ; i.e., the periodogram of a centered time series is necessarily zero at the origin. What does this imply about  $\sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(P)}$ , where  $\hat{s}_\tau^{(P)}$  is defined in keeping with Equation (166b)?
- (c) Here we examine one of the consequences of *not* centering a time series before computing the periodogram. Suppose that  $Y_0, \dots, Y_{N-1}$  is a segment of length  $N$  of a stationary process with ACVS  $\{s_\tau\}$ , SDF  $S(\cdot)$  and nonzero mean  $\mu$ . Define  $X_t = Y_t - \mu$  so that  $\{X_t\}$  is also a stationary process with ACVS  $\{s_\tau\}$  and SDF  $S(\cdot)$ , but now with zero mean. Let

$$\tilde{S}_Y^{(P)}(f) \stackrel{\text{def}}{=} \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} Y_t e^{-i2\pi f t \Delta_t} \right|^2 \quad \text{and} \quad \hat{S}_X^{(P)}(f) \stackrel{\text{def}}{=} \frac{\Delta_t}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi f t \Delta_t} \right|^2$$

(since  $\{X_t\}$  is a zero mean process, the latter is in keeping with Equation (170d)). Show that

$$E\{\tilde{S}_Y^{(P)}(f)\} = E\{\hat{S}_X^{(P)}(f)\} + \mu^2 \mathcal{F}(f), \quad (240b)$$

where  $\mathcal{F}(\cdot)$  is Fejér's kernel (defined in Equation (175a)).

- (d) As an example of the practical impact of (240a) and (240b), consider the Willamette River time series, which is downloadable from the "Data" part of the website for the book (Figures 2(c) and 568a show, respectively, the first 128 values and all  $N = 395$  values of this series – the entire series is to be considered for this exercise). Here the sampling interval and Nyquist frequency are  $\Delta_t = 1/12$  year and  $f_N = 6$  cycles per year. Figure 568b shows the periodogram over the grid of frequencies  $k/(1024 \Delta_t)$ ,  $k = 1, \dots, 512$ , after centering the series by subtracting off its sample mean of 9.825 (see Exercise [179] for details about how to compute the periodogram over this grid of frequencies). Compute and plot the periodogram for the time series *without* centering it (1) over the nonzero Fourier frequencies  $k/(395 \Delta_t)$ ,  $k = 1, \dots, 197$ , and (2) over the same grid of frequencies used in Figure 568b. Taking into consideration Equations (240a) and (240b), comment on how these two plots compare with Figure 568b.

- [6.16] Verify Equation (186g). Hint: consider replacing  $X_t/\sqrt{N}$  in the solution to Exercise [170] with  $h_t X_t$ .

- [6.17] Show that Equation (188a) is true.

- [6.18] (a) Define  $h_t = C [1 - \cos(2\pi t/N)]$ ,  $t = 0, 1, \dots, N-1$ , where  $C$  is a constant chosen such that  $\sum_{t=0}^{N-1} h_t^2 = 1$ . This is essentially the definition for the Hanning data taper given by Priestley (1981, p. 562), but it is slightly different from the one in Equation (189b); in particular, here  $h_0 = 0$ , and the taper is asymmetric in the sense that  $h_t \neq h_{N-t-1}$  in general, whereas in (189b)  $h_t > 0$  and  $h_t = h_{N-t-1}$  for  $0 \leq t \leq N-1$ . Let

$$I(f_k) \stackrel{\text{def}}{=} \left(\frac{\Delta_t}{N}\right)^{1/2} \sum_{t=0}^{N-1} X_t e^{-i2\pi f_k t \Delta_t} \quad \text{and} \quad J(f_k) \stackrel{\text{def}}{=} \Delta_t^{1/2} \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi f_k t \Delta_t},$$

where, as usual,  $f_k = k/(N \Delta_t)$ . Show that

$$J(f_k) = C N^{1/2} \left( -\frac{1}{2} I(f_{k-1}) + I(f_k) - \frac{1}{2} I(f_{k+1}) \right).$$

- (b) Noting that  $\hat{S}_G^{(P)}(f_k) = |I(f_k)|^2$  and  $\hat{S}_G^{(D)}(f_k) = |J(f_k)|^2$ , show that, when  $\{X_t\}$  is Gaussian white noise with mean zero and variance  $\sigma^2$ ,

$$\text{corr}\{\hat{S}_G^{(D)}(f_{k+\tau}), \hat{S}_G^{(D)}(f_k)\} = \begin{cases} 4/9, & \tau = 1; \\ 1/36, & \tau = 2; \\ 0, & \tau = 3, \end{cases} \quad (241)$$

for  $2 \leq k < \lfloor N/2 \rfloor - 3$ , whereas  $\text{corr}\{\hat{S}_G^{(P)}(f_{k+\tau}), \hat{S}_G^{(P)}(f_k)\} = 0$  for  $\tau = 1, 2$  and  $3$ .  
Hint: make use of the Isserlis theorem (Equation (30)) in a manner similar to how it is used to derive Equation (211a).

- [6.19] Verify the computational formula in Equation (192b).  
[6.20] Consider a time series  $X_0, \dots, X_{N-1}$  that is a portion of a stationary process with mean  $\mu$ . Let  $\{h_t\}$  be a data taper whose associated spectral window is  $\mathcal{H}(\cdot)$ . Consider

$$\hat{S}^{(D)}(f) = \Delta_t \left| \sum_{t=0}^{N-1} h_t(X_t - \mu) e^{-i2\pi f t \Delta_t} \right|^2 \quad \text{and} \quad \tilde{S}^{(D)}(f) \stackrel{\text{def}}{=} \Delta_t \left| \sum_{t=0}^{N-1} (h_t X_t - \mu) e^{-i2\pi f t \Delta_t} \right|^2,$$

i.e., a direct spectral estimator and a related estimator in which the process mean is subtracted *after* the time series is tapered.

- (a) Develop an equation that relates  $E\{\hat{S}^{(D)}(f)\}$  and  $E\{\tilde{S}^{(D)}(f)\}$  for  $0 \leq f \leq f_N$ , where, as usual,  $f_N = 1/(2\Delta_t)$  is the Nyquist frequency.  
(b) Using the equation called for in part (a), show that, at Fourier frequencies  $f_k = k/(N\Delta_t)$  satisfying  $0 < f_k \leq f_N$ , we have  $E\{\tilde{S}^{(D)}(f_k)\} = E\{\hat{S}^{(D)}(f_k)\} + \mu^2 \mathcal{H}(f_k)$ .  
(c) Briefly discuss the implication of part (a) for SDF estimation when the operations of centering and tapering a white noise process with nonzero mean are inadvertently interchanged. Briefly discuss the implications of part (b) for both white and colored processes  $\{X_t\}$  when  $\{h_t\}$  is the default (rectangular) data taper.
- [6.21] Recall the definition for the average value of a function given in Equation (239). Let  $\hat{S}^{(P)}(\cdot)$  and  $\hat{S}^{(D)}(\cdot)$  represent, respectively, the periodogram and a direct spectral estimator based upon the Hanning data taper for a time series with Nyquist frequency  $f_N$ . Can we potentially determine the need for tapering by comparing the average values of  $\hat{S}^{(P)}(\cdot)$  and  $\hat{S}^{(D)}(\cdot)$  over the interval  $[-f_N, f_N]$ ? If we replace  $\hat{S}^{(P)}(\cdot)$  and  $\hat{S}^{(D)}(\cdot)$  with  $\log(\hat{S}^{(P)}(\cdot))$  and  $\log(\hat{S}^{(D)}(\cdot))$ , can we potentially determine the need for tapering by comparing their average values? Illustrate your answers by considering the periodograms (Figure 173) and Hanning-based direct spectral estimates (Figure 187) for the four AR(4) time series displayed in Figure 34 (these are available on the website for the book).
- [6.22] Here we explore the use of short prewhitening filters to compensate for bias in the periodogram of the AR(4) time series  $\{X_t\}$  displayed in Figure 34(f) (this series is available on the website for the book). Use the prewhitening filter  $\{g_0 = 1, g_1 = -0.9\}$  to obtain a prewhitened series  $\{Y_t\}$ . Compute the periodogram for  $\{Y_t\}$  over the grid of frequencies  $f'_k = k/2048$ ,  $k = 0, \dots, 1024$ , and then postcolor the periodogram to create an estimate for the SDF of  $\{X_t\}$ . Plot this estimate in a manner similar to Figure 199(d). Repeat using the prewhitening filter  $\{g_0 = 1, g_1 = -1\}$  in place of  $\{g_0 = 1, g_1 = -0.9\}$ . Finally, repeat using the filter  $\{g_0 = 1, g_1 = -0.7165\}$ . How well do these three estimates of the SDF for  $\{X_t\}$  compare to the estimate shown in Figure 199(d)?
- [6.23] (a) Show that, if we let  $0 < f < f_N$  rather than restricting  $f$  to be a Fourier frequency, then Equation (203c) becomes  $E\{\hat{S}_G^{(P)}(f)\} = \sigma^2 \Delta_t = S_G(f)$  while Equation (203d) is then

$$\text{var}\{\hat{S}_G^{(P)}(f)\} = \sigma^4 \Delta_t^2 (1 + \mathcal{D}_N^2(2f\Delta_t)) = S_G^2(f) (1 + \mathcal{D}_N^2(2f\Delta_t)),$$

where  $\mathcal{D}_N(\cdot)$  is Dirichlet's kernel as defined by Equation (17c) (Proakis and Manolakis, 2007, Problem 14.3(c)).

- (b) For what values of  $f$  does  $\text{var}\{\hat{S}_G^{(P)}(f)\}$  achieve its minimum value of  $\sigma^4 \Delta_t^2$ ?  
(c) Argue that, as  $N \rightarrow \infty$ ,  $\text{var}\{\hat{S}_G^{(P)}(f)\} \rightarrow \sigma^4 \Delta_t^2$  for all  $0 < f < f_N$ .  
(d) Plot  $\text{var}\{\hat{S}_G^{(P)}(f)\}$  versus  $f$  for the cases  $N = 16$  and  $64$  assuming that  $\sigma^2 = 1$  and  $\Delta_t = 1$ .

- [6.24] (a) Generate a realization of a portion  $X_0, X_1, \dots, X_{63}$  of the zero mean AR(2) process of Equation (34) (see Exercise [597] for a description of how to generate realizations from this process). Assuming for convenience that  $\Delta_t = 1$ , compute the periodogram for  $\{X_t\}$  at three adjacent Fourier frequencies, namely,  $f_6 = 6/64$ ,  $f_7 = 7/64$  and  $f_8 = 8/64$ , and call these values  $\hat{S}_0^{(P)}(f_6)$ ,  $\hat{S}_0^{(P)}(f_7)$  and  $\hat{S}_0^{(P)}(f_8)$ . Repeat what we just described a large number  $N_R$  of times (using a different realization of  $\{X_t\}$  each time) to obtain the sequences  $\{\hat{S}_j^{(P)}(f_6) : j = 0, \dots, N_R - 1\}$ ,  $\{\hat{S}_j^{(P)}(f_7) : j = 0, \dots, N_R - 1\}$  and  $\{\hat{S}_j^{(P)}(f_8) : j = 0, \dots, N_R - 1\}$  (take “large” to mean something from 1,000 up to 100,000). Compute the sample mean and sample variance for the three sequences, and compute the sample correlation coefficient  $\hat{\rho}$  between  $\{\hat{S}_j^{(P)}(f_6)\}$  and  $\{\hat{S}_j^{(P)}(f_7)\}$ , between  $\{\hat{S}_j^{(P)}(f_6)\}$  and  $\{\hat{S}_j^{(P)}(f_8)\}$  and between  $\{\hat{S}_j^{(P)}(f_7)\}$  and  $\{\hat{S}_j^{(P)}(f_8)\}$  (see Equation (3a)). Compare these sample values with theoretical values suggested by Equations (174a), (175c), (145a), (203f) and (203g) (Equation (174a) calls for the ACVS for the AR(2) process – Exercise [9.6] states how to form it). Form histograms for  $\{\hat{S}_j^{(P)}(f_6)\}$ ,  $\{\hat{S}_j^{(P)}(f_7)\}$  and  $\{\hat{S}_j^{(P)}(f_8)\}$ , and compare them to the PDFs suggested by Equations (203e) and (209a). Form histograms also for  $\{10 \log_{10}(\hat{S}_j^{(P)}(f_6))\}$ ,  $\{10 \log_{10}(\hat{S}_j^{(P)}(f_7))\}$  and  $\{10 \log_{10}(\hat{S}_j^{(P)}(f_8))\}$ , and compare them to the PDFs suggested by Equation (209b).
- (b) Repeat part (a), but now use the Fourier frequencies  $f_{29} = 29/64$ ,  $f_{30} = 30/64$  and  $f_{31} = 31/64$  in place of  $f_6$ ,  $f_7$  and  $f_8$ .
- [6.25] Repeat Exercise [6.24], but with the following modifications:
- (a) use a direct spectral estimate with a Hanning data taper (Equation (189b)) instead of the periodogram and,
- (b) in comparing the sample values to theory, use Equation (186g) rather than (174a); use Equation (204c) rather than (203f); and Equation (207) with  $c = 2$  rather than (204a).
- [6.26] Repeat Exercises [6.24] and [6.25], but with the following modifications. In place of the AR(2) process, generate realizations of a portion  $X_0, X_1, \dots, X_{255}$  of the zero mean AR(4) process of Equation (35a) (Exercise [11.1] describes how to generate these realizations). For part (a) of Exercise [6.24] take the three adjacent Fourier frequencies to be  $27/256$ ,  $28/256$  and  $29/256$ . Exercise [9.6] states how to form the ACVS for the AR(4) process, which is needed for part (a). For part (b) take the three frequencies to be  $125/256$ ,  $126/256$  and  $127/256$ .
- [6.27] Suppose that  $\{X_t\}$  is a real-valued stationary process with zero mean, sampling interval  $\Delta_t$ , ACVS  $\{s_\tau\}$  and SDF  $S(\cdot)$ . Let  $a_0, \dots, a_{N-1}$  and  $b_0, \dots, b_{N-1}$  be any two sequences of complex-valued numbers of finite length  $N$ . Define

$$A(f) = \Delta_t \sum_{t=0}^{N-1} a_t e^{-i2\pi f t \Delta_t} \quad \text{and} \quad B(f) = \Delta_t \sum_{t=0}^{N-1} b_t e^{-i2\pi f t \Delta_t};$$

i.e., with  $a_t \stackrel{\text{def}}{=} 0$  and  $b_t \stackrel{\text{def}}{=} 0$  for all  $t < 0$  and  $t \geq N$ , then  $\{a_t\} \longleftrightarrow A(\cdot)$  and  $\{b_t\} \longleftrightarrow B(\cdot)$ .

- (a) Use the spectral representation theorem for  $\{X_t\}$  to show that

$$\text{cov} \left\{ \sum_{t=0}^{N-1} a_t X_t, \sum_{t=0}^{N-1} b_t X_t \right\} = \frac{1}{\Delta_t^2} \int_{-f_N}^{f_N} A(f) B^*(f) S(f) df. \quad (242a)$$

- (b) Show that we also have

$$\text{cov} \left\{ \sum_{t=0}^{N-1} a_t X_t, \sum_{t=0}^{N-1} b_t X_t \right\} = \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} a_t b_u^* s_{t-u}. \quad (242b)$$



(c) Now suppose that  $\{X_t\}$  is a Gaussian process. Show that

$$\begin{aligned} \text{cov} \left\{ \left| \sum_{t=0}^{N-1} a_t X_t \right|^2, \left| \sum_{t=0}^{N-1} b_t X_t \right|^2 \right\} \\ = \frac{1}{\Delta_t^4} \left| \int_{-f_N}^{f_N} A(f) B^*(f) S(f) df \right|^2 + \frac{1}{\Delta_t^4} \left| \int_{-f_N}^{f_N} A(f) B(-f) S(f) df \right|^2 \end{aligned} \quad (243a)$$

(hint: generalize the argument leading to Equation (211a)).

(d) Show that Equation (243a) can be rewritten as

$$\text{cov} \left\{ \left| \sum_{t=0}^{N-1} a_t X_t \right|^2, \left| \sum_{t=0}^{N-1} b_t X_t \right|^2 \right\} = \left| \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} a_t b_u^* s_{t-u} \right|^2 + \left| \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} a_t b_u s_{t-u} \right|^2. \quad (243b)$$

(e) Use the results of part (c) to verify Equation (211b) (hint: let  $a_t$  and  $b_t$  be, respectively,  $\Delta_t^{1/2} h_t e^{-i2\pi(f+\eta)t \Delta_t}$  and  $\Delta_t^{1/2} h_t e^{-i2\pi f t \Delta_t}$ ).

(f) Consider a time series  $X_0, X_1, \dots, X_{N-1}$  that is a portion of a Gaussian AR(1) process  $X_t = \phi X_{t-1} + \epsilon_t$  with an ACVS given by  $s_\tau = \phi^{|\tau|}/(1 - \phi^2)$ , where  $\phi = 0.8$ , and  $\{\epsilon_t\}$  is a zero mean Gaussian white noise process with variance  $\sigma_\epsilon^2 = 1$  (see Exercise [2.17a]). Given this series (and assuming  $\Delta_t = 1$  for convenience), consider the periodogram  $\hat{S}^{(P)}(\tilde{f}_k)$  of Equation (170d) over the grid of frequencies  $\tilde{f}_k = k/(2N)$ ,  $k = 0, 1, \dots, N$ , with  $N$  set initially to 16, then to 64 and finally to 256. Use part (d) of this exercise to compute the variance of  $\hat{S}^{(P)}(\tilde{f}_k)$  for all  $N+1$  settings for  $\tilde{f}_k$  and for all three choices of sample size  $N$ . An approximation to this variance is obtainable from Equation (203f) in conjunction with (145a) (in the latter,  $p$  should be set to 1, and  $\phi_{1,1}$  to  $\phi$ ). Plot the ratio of the approximate to the exact variance versus  $\tilde{f}_k$  for each of the three sample sizes. Repeat what we just described, but now with  $\phi$  set to 0.9, 0.99 and finally  $-0.99$ . Comment on how well the approximation works.

[6.28] (a) If  $\{h_t\} \longleftrightarrow H(\cdot)$  and  $\{s_{G,\tau}\} \longleftrightarrow S_G(\cdot)$ , show that Equation (211b) can be rewritten as

$$\begin{aligned} \text{cov} \{ \hat{S}_G^{(D)}(f + \eta), \hat{S}_G^{(D)}(f) \} \\ = \Delta_t^2 \left| \sum_{j=0}^{N-1} h_j C_j^*(f + \eta) e^{-i2\pi f j \Delta_t} \right|^2 + \Delta_t^2 \left| \sum_{j=0}^{N-1} h_j C_j(f + \eta) e^{-i2\pi f j \Delta_t} \right|^2, \end{aligned}$$

where

$$C_j(f) \stackrel{\text{def}}{=} \sum_{k=0}^{N-1} h_k s_{G,j-k} e^{-i2\pi f k \Delta_t}.$$

(b) As per Equation (212a), let

$$R(\eta) \stackrel{\text{def}}{=} \left| \sum_{t=0}^{N-1} h_t^2 e^{-i2\pi \eta t \Delta_t} \right|^2, \quad \text{and, as per usual, assume } \sum_{t=0}^{N-1} h_t^2 = 1.$$

Show that, for Gaussian white noise with variance  $\sigma^2$ ,

$$\text{cov} \{ \hat{S}_G^{(D)}(f + \eta), \hat{S}_G^{(D)}(f) \} = \sigma^4 [R(\eta) + R(2f + \eta)] \Delta_t^2$$

and hence that

$$\text{corr} \{ \hat{S}_G^{(D)}(f + \eta), \hat{S}_G^{(D)}(f) \} = \frac{R(\eta) + R(2f + \eta)}{\sqrt{[1 + R(2f)] [1 + R(2f + 2\eta)]}}.$$

- (c) Using the Hanning data taper, create a plot of  $\text{corr}\{\hat{S}_G^{(D)}(f+\eta), \hat{S}_G^{(D)}(f)\}$  versus  $\eta \in [0, 1/4]$  for the case  $N = 64$  and  $f = 1/4$  with  $\Delta_t$  assumed to be unity so that  $f_N = 1/2$ .
- [6.29] Suppose that  $\{h_t\}$  is the rectangular data taper. What does  $R(\eta)$  in Equation (212a) reduce to? For what values of  $\eta$  is  $R(\eta) = 0$ ?
- [6.30] Generate a realization of a portion  $X_0, X_1, \dots, X_{63}$  of the zero mean AR(2) process of Equation (34) (see Exercise [597] for a description of how to generate realizations from this process). Assuming  $\Delta_t = 1$  and using the Hanning data taper  $\{h_t\}$ , compute a direct spectral estimate for  $\{X_t\}$  at the frequencies  $f'_k = k/1024$ ,  $k = 0, \dots, 48$ , and denote the  $k$ th of these 49 values by  $\hat{S}_0^{(D)}(f'_k)$  (hint: to compute  $\hat{S}_0^{(D)}(\cdot)$  over the required grid of frequencies, consider the effect of zero padding  $\{h_t X_t\}$  with  $1024 - 64 = 960$  zeros.) Repeat what we just described a large number  $N_R$  of times to obtain  $\{\hat{S}_j^{(D)}(f'_k) : j = 0, \dots, N_R - 1, k = 0, \dots, 48\}$  (take “large” to mean something from 1,000 up to 100,000). Compute the sample mean and sample variance for each of these 49 sequences. Compare these sample values with theoretical values suggested by Equation (204b). Repeat again, but now center each simulated series by subtracting off its sample mean prior to applying the Hanning data taper. Compute the effective bandwidth  $B_H$  for  $\hat{S}_j^{(D)}(\cdot)$  as per Equations (194) and (192c), and comment upon the claim that the  $\chi_2^2$ -based result stated in Equation (204b) is reasonable at low frequencies but with the restriction that  $f \geq B_H/2$  or, more conservatively,  $f \geq B_H$  (see C&E [3] for Section 6.6). Comment also on the applicability of the  $\chi_1^2$ -based result stated in Equation (204b) when the series are centered.
- [6.31] Here we consider some properties of an SDF estimator based on the periodogram of the reflected series shown in Equation (183c) (C&E [7] for Section 6.6 discusses use of the alternative scheme displayed in Equation (183b), which leads to the estimator  $\hat{S}^{(DCT)}(\cdot)$  defined in Exercise [217]).

- (a) Show that the periodogram of the reflected series over  $\check{f}_k \stackrel{\text{def}}{=} k/(2N - 2)$  takes the form

$$\hat{S}(\check{f}_k) = \frac{1}{2N - 2} \left( X_0 + 2 \sum_{t=1}^{N-2} X_t \cos(2\pi \check{f}_k t) + (-1)^k X_{N-1} \right)^2.$$

- (b) Create a plot similar to Figure 217 that shows  $\hat{S}(\check{f}_k)$  versus  $\check{f}_k$  for the four AR(4) time series displayed in Figure 34 (these series are available on the website for the book). How does  $\hat{S}(\cdot)$  compare to the DCT-based periodograms  $\hat{S}^{(DCT)}(\cdot)$  shown in Figure 217?
- (c) Assuming  $\{X_t\}$  is a stationary process with mean zero and ACVS  $\{s_\tau\}$ , show that

$$E\{\hat{S}(\check{f}_k)\} = \frac{1}{N - 1} \left( C_k + D_k + 2 \sum_{\tau=1}^{N-2} \left( s_\tau + (-1)^k s_{N-1-\tau} \right) \cos(2\pi \check{f}_k \tau) \right)$$

for  $k = 0, 1, \dots, N - 1$ , where  $C_k \stackrel{\text{def}}{=} s_0 + (-1)^k s_{N-1}$  and

$$\begin{aligned} D_0 &\stackrel{\text{def}}{=} 2 \sum_{\tau=-(N-3)}^{N-3} (N - |\tau| - 2) s_\tau, \\ D_k &\stackrel{\text{def}}{=} (N - 3) s_0 + 2 \sum_{\tau=1}^{N-3} (N - \tau - 2) s_\tau \cos(2\pi \check{f}_k \tau) \\ &\quad - \frac{2}{\sin(2\pi \check{f}_k)} \sum_{\tau=1}^{N-3} s_\tau \sin(2\pi \check{f}_k [\tau + 1]), \quad k = 1, \dots, N - 2, \\ D_{N-1} &\stackrel{\text{def}}{=} 2 \sum_{\tau=-(N-3)}^{N-3} (N - |\tau| - 2) (-1)^\tau s_\tau. \end{aligned}$$

- (d) Create a plot similar to Figure 218 that shows  $E\{\hat{S}(\check{f}_k)\}$  versus  $\check{f}_k$  for sample sizes  $N = 16, 64, 256$  and  $1024$  for the AR(4) process of Equation (35a). How does  $E\{\hat{S}(\cdot)\}$  compare to  $E\{\hat{S}^{(DCT)}(\cdot)\}$  shown in Figure 218? Note: Exercise [9.6] tells how to compute the ACVS for the AR(4) process.