1

Introduction to Spectral Analysis

1.0 Introduction

This chapter provides a quick introduction to the subject of spectral analysis. Except for some later references to the exercises of Section 1.6, this material is independent of the rest of the book and can be skipped without loss of continuity. Our intent is to use some simple examples to motivate the key ideas. Since our purpose is to view the forest before we get lost in the trees, the particular analysis techniques we use here have been chosen for their simplicity rather than their appropriateness.

1.1 Some Aspects of Time Series Analysis

Spectral analysis is part of time series analysis, so the natural place to start our discussion is with the notion of a time series. The quip (attributed to R. A. Fisher) that a time series is "one damned thing after another" is not far from the truth: loosely speaking, a time series is a set of observations made sequentially in time (but "time" series are also often recorded sequentially in, e.g., distance or depth). Examples abound in the real world, and Figure 2 shows plots of small portions of four actual time series:

(a) the speed of the wind in a certain direction, measured every 0.025 sec;
(b) the daily record of a quantity (to be precise, the change in average daily frequency) that tells how well an atomic clock keeps time on a day-to-day basis (a constant value of zero would indicate that the clock agreed perfectly with a time scale maintained by the US Naval Observatory);
(c) monthly average measurements related to the flow of water in the Willamette River at Salem, Oregon; and
(d) the change in the level of ambient noise in the ocean from one second to the next.

For each of these plots, the values of the time series at 128 successive times are connected by lines to help the eye follow the variations in the series. The visual appearances of these four series are quite different.

The chief aim of time series analysis is to develop quantitative means to allow us to characterize time series, e.g., to say quantitatively how one series differs from another or how two series are related. There are two broad classes of characterizations, namely, time domain techniques and frequency domain techniques. Spectral analysis is the prime example of a frequency domain technique. Before we introduce it, we will first consider a popular time

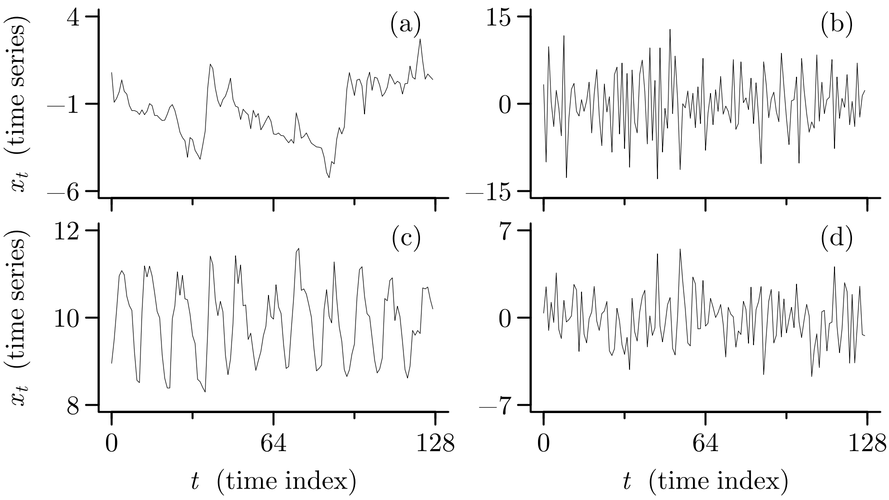

Figure 2 Plots of portions of four time series related to (a) wind speed, (b) an atomic clock, (c) the Willamette River and (d) ocean noise. For each series the vertical axis is the value of the time series (in unspecified units), while the horizontal axis is a unitless time index (the actual time between adjacent observations is 0.025 sec for the wind speed series, one day for the atomic clock data, one month for the Willamette River series and one second for the ocean noise series).

domain technique. We contend that this latter technique is not completely satisfactory and that spectral analysis is a useful and complementary alternative to it.

Let us concentrate for the moment on the wind speed and atomic clock data (top row of plots in Figure 2). How do these two series differ? In the wind speed series adjacent points of the time series tend to be close in value, while in the atomic clock series positive values tend to be followed by negative values, and vice versa. To see this effect graphically, we can plot x_{t+1} versus x_t as the time index t varies from 0 to $N-2$, where we let $x_0, x_1, \ldots, x_{N-1}$ represent any one of our series and let N represent the sample size, i.e., the number of data points in a time series, 128 in our case. Such a plot is called a "lag 1 scatter plot," and Figure 3 shows this plot for each of our four series. We note the following:

(a) For the wind speed series, the points tend to fall about a line of positive slope. Thus a wind speed with a certain value tends to be followed by one near that same value.
(b) For the atomic clock data, the points fall loosely about a line with a negative slope.
(c) The plot for the Willamette River data resembles that of the wind speed series except that the points are more spread out.
(d) For the ocean noise data, it is not obvious that there is a tendency of the points to cluster about a line in one direction or another.

We could create a lag τ scatter plot by plotting $x_{t+\tau}$ versus x_t, but, while such plots are informative, they are unwieldy to work with. To summarize the information in scatter plots similar to those in Figure 3, note that these plots indicate a roughly linear relationship between x_{t+1} and x_t; i.e., with $\tau = 1$, we can write

$$x_{t+\tau} = \alpha_\tau + \beta_\tau x_t + \epsilon_{\tau,t}$$

for some intercept α_τ and slope β_τ (possibly equal to 0), where $\epsilon_{\tau,t}$ represents an "error" term that models deviations from strict linearity. If we make the assumption that a linear relationship holds approximately between $x_{t+\tau}$ and x_t for all τ, we can use as a summary statistic a

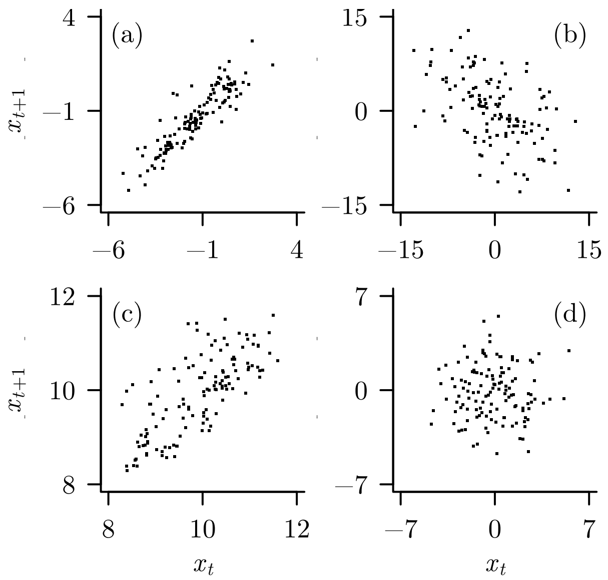

Figure 3 Lag 1 scatter plots for four time series in Figure 2. In each of plot, the value of the time series at time index $t + 1$ is plotted on the vertical axis versus the value at time index t on the horizontal axis (t ranges from 0 to 126).

well-known measure of the strength of the linear association between two ordered collections of variables $\{y_t\}$ and $\{z_t\}$, namely, the Pearson product moment correlation coefficient:

$$\hat{\rho} = \frac{\Sigma(y_t - \overline{y})(z_t - \overline{z})}{\left[\Sigma(y_t - \overline{y})^2 \Sigma(z_t - \overline{z})^2\right]^{1/2}}, \tag{3a}$$

where $\overline{y}$ and $\overline{z}$ are the sample means of the y_t and z_t terms, respectively. This coefficient can be interpreted in many ways (Rogers and Nicewander, 1988; Falk and Well, 1997; Rovine and von Eye, 1997; Nelsen, 1998). For example, if we use $\{y_t\}$ and $\{z_t\}$ to form two vectors, then $\hat{\rho}$ is the cosine of the angle between them. If we let $y_t = x_{t+\tau}$ and $z_t = x_t$, and if we adjust the summations in the denominator to make use of all available data, we are led to the lag τ sample autocorrelation for a time series:

$$\hat{\rho}_\tau = \frac{\sum_{t=0}^{N-\tau-1}(x_{t+\tau} - \bar{x})(x_t - \bar{x})}{\sum_{t=0}^{N-1}(x_t - \bar{x})^2}. \tag{3b}$$

Note that $\hat{\rho}_0 = 1$ and that, as τ increases, the numerator is based on fewer and fewer cross products. As a sequence indexed by the lag τ, the quantity $\{\hat{\rho}_\tau\}$ is called the sample autocorrelation sequence (sample ACS) for the time series x_t. (See Exercise [1.6] for a caveat about interpreting $\hat{\rho}_\tau$ as a true correlation coefficient.)

The sample ACS up to lag 32 is plotted for our four time series in Figure 4. A careful study of these plots can reveal a lot about these series. For example, from the ACS in (c) for the Willamette River data, we see that x_t and x_{t+6} are negatively correlated, while x_t and x_{t+12} are positively correlated. This pattern is consistent with the visual evidence in Figure 2 that the river flow varies with a period of roughly 12 months.

Let us now assume that the time series $x_0, x_1, \ldots, x_{N-1}$ can be regarded as observed values (i.e., realizations) of corresponding random variables (RVs) $X_0, X_1, \ldots, X_{N-1}$. We use the term "modeling of a time series" for the procedure by which we specify the properties

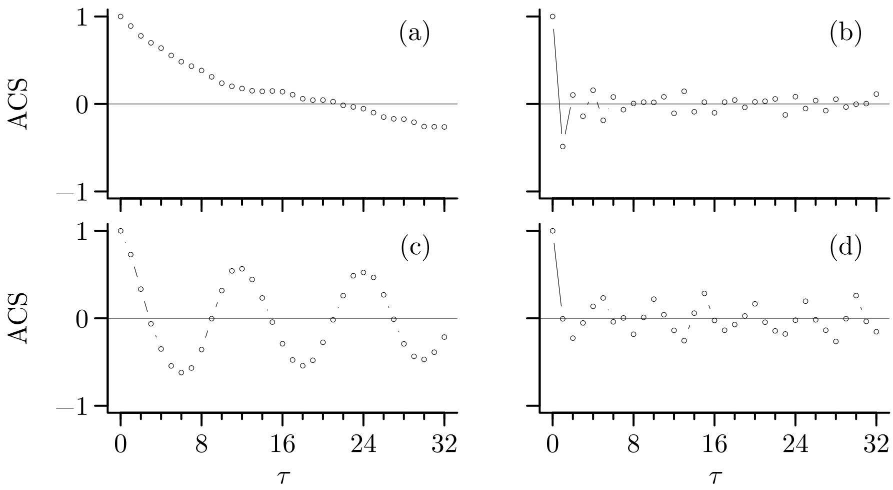

Figure 4 Sample autocorrelation sequences $\{\hat{\rho}_\tau\}$ for the time series of Figure 2. The value of the ACS at lag τ is plotted versus τ for τ ranging from 0 to 32. By definition the ACS for lag 0 is 1.

of these N RVs. For a class of models reasonable for time series such as those in Figure 2, $\hat{\rho}_\tau$ is an estimate of a corresponding population quantity called the lag τ theoretical autocorrelation, defined as

$$\rho_\tau = E\{(X_{t+\tau} - \mu)(X_t - \mu)\}/\sigma^2,$$

where $E\{W\}$ is our notation for the expectation operator applied to the RV W; $\mu = E\{X_t\}$ is the population mean of the time series; and $\sigma^2 = E\{(X_t - \mu)^2\}$ is the corresponding population variance. (Note, in particular, that ρ_τ, μ and σ^2 do not depend on t. As we shall see later, models for which this is true play a central role in spectral analysis and are called stationary.) Moreover, if we make an additional assumption, namely, that the X_t terms follow a multivariate Gaussian (normal) distribution, knowledge of the ρ_τ terms, σ^2 and μ completely specifies our model. Thus, to fit such a model to a time series, we need only estimate ρ_τ, σ^2 and μ from the available data.

As a set of parameters, the ρ_τ terms, σ^2 and μ constitute a time domain characterization of a model. Since a model is completely specified by these parameters in the Gaussian case and since these parameters can all be estimated from a time series, why would we want to consider other characterizations? There are several reasons:

[1] The parameters of a model should ideally make it easy for us to visualize typical time series that can be generated by the model. Unfortunately, it takes a fair amount of experience to be able to look at a theoretical ACS and visualize what kind of time series it corresponds to.

[2] For a lag τ that is a substantial proportion of the length N of a time series, it is often hard to get reliable estimates of ρ_τ (and even more difficult to do so for τ greater than N). This is evident from Equation (3b) since the number of cross products that are used in forming the numerator decreases as τ increases. The variance of $\hat{\rho}_\tau$ depends upon τ and the true ACS in a complicated way – typically it increases as τ increases. Moreover, for most cases of interest the estimators $\hat{\rho}_\tau$ and $\hat{\rho}_{\tau+1}$ are highly correlated. This lack of homogeneity of variance and the correlation between nearby estimators make a plot of $\hat{\rho}_\tau$ versus τ hard to interpret.

[3] Because of these sampling problems, it is difficult to devise good statistical tests based upon $\hat{\rho}_\tau$ for various hypotheses of interest. For example, suppose we entertain a hypoth-

esis that specifies values for ρ_1 and ρ_2. To evaluate to what extent the sample values $\hat{\rho}_1$ and $\hat{\rho}_2$ offer evidence for or against this hypothesis, we need to derive their statistical properties. In general this is not an easy task because it can be difficult to determine the variances of $\hat{\rho}_1$ and $\hat{\rho}_2$ and the degree to which they are correlated.

[4] Even in the rare instances where we believe we have enough data to estimate ρ_τ reliably, a second model characterization can be useful as a complementary way of viewing the properties of our data. In particular, in contrast to time domain models, the characterization behind spectral analysis makes it much easier to visualize the kinds of time series that would be generated by the model (we return to this point in Section 4.6).

Comments and Extensions to Section 1.1

[1] The reader might well ask whether the right-hand side of Equation (3b) should be multiplied by a factor of $N/(N-\tau)$ to compensate for the different number of terms in the summations in the numerator and denominator. Most time series analysts would answer "no." As discussed in Chapter 6, estimation of the ACS via Equation (3b) yields a sequence that corresponds to the ACS for some theoretical stationary process. If we introduce the factor of $N/(N-\tau)$, we cannot make a similar statement, and we could run into practical problems in using the resulting ACS estimates. For example, based upon these estimates, we could in fact obtain a negative value when attempting to compute the variance of certain linear combinations of our time series – this would obviously be nonsense since variances must always be nonnegative.

[2] The use of the term "sample autocorrelation" for the right-hand side of Equation (3b) conforms to that of the statistical literature. Unfortunately this conflicts with the engineering literature, in which sometimes either the quantity

$$\frac{1}{N}\sum_{t=0}^{N-\tau-1}(x_{t+\tau}-\bar{x})(x_t-\bar{x}) \quad \text{or} \quad \frac{1}{N}\sum_{t=0}^{N-\tau-1}x_{t+\tau}x_t$$

is called the "lag τ sample autocorrelation." This latter notation can lead to unnecessary confusion between correlations and covariances and cause nonzero means to be ignored.

[3] We do not want to leave the impression that lag τ scatter plots for time series always indicate an approximately linear relationship between $x_{t+\tau}$ and x_t. As a simple counterexample, Figure 6 shows the first 24 years of a time series of monthly average temperatures at St. Paul, Minnesota, as well as the lag 6 and 9 scatter plots for the entire time series (this extends from 1820 to 1983), both of which are highly nonlinear. In these cases the summary given by the sample ACS does not give the full story about the relationship between $x_{t+\tau}$ and x_t.

[4] While it is reasonable that $\mu = E\{X_t\}$ is independent of the time index t for the wind speed, atomic clock and ocean noise series, it would seem to be an unreasonable assumption for the Willamette River data, which varies with a prominent annual pattern. A more natural assumption is that $E\{X_t\}$ is a function of which month X_t occurs in. As we shall see, the key concept of stationarity assumes that certain quantities – including $E\{X_t\}$ – are independent of the index t. It would appear at first that we cannot assume a stationary model for the Willamette River data as we have implied above. In fact, as we shall discuss later (see Sections 2.6 and 2.8), there is a mathematical trick that allows us to treat such data in the context of a stationary model (the trick involves assuming that the time origin of a periodic phenomenon can be regarded as being picked at random).

1.2 Spectral Analysis for a Simple Time Series Model

Some of the problems of estimation and interpretation that are associated with the ACS are lessened (but not completely alleviated) when we deal with a frequency domain characterization called the "spectrum." The spectrum is simply a second way of characterizing models for time series. The objective of spectral analysis is to study and estimate the spectrum.

Cambridge University Press
978-1-107-02814-2 — Spectral Analysis for Univariate Time Series
Donald B. Percival , Andrew T. Walden
Excerpt
More Information

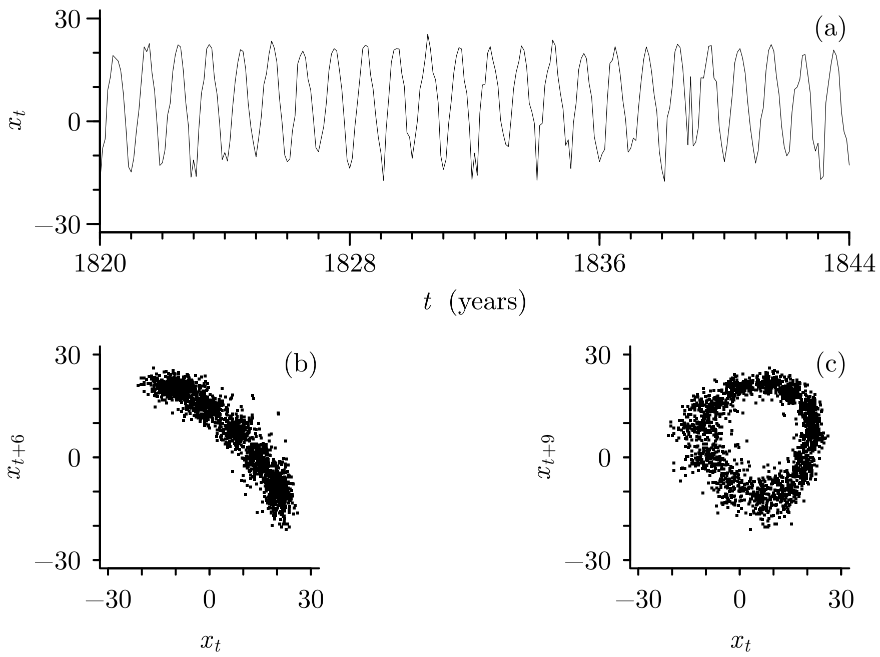

Figure 6 Plots of (a) the first 24 years of the St. Paul temperature time series, (b) the lag 6 scatter plot for the entire series and (c) the corresponding lag 9 scatter plot. The temperature series is measured in degrees centigrade. For the lag τ scatter plot (τ = 6, 9), the value $x_{t+\tau}$ is plotted on the vertical axis versus x_t on the horizontal axis.

How exactly we define the spectrum depends upon what class of models we assume for a time series. A detailed definition for a useful class of models is presented in Chapter 4, but the key idea behind the spectrum is based upon a model for a time series consisting of a linear combination of cosines and sines with different frequencies; i.e.,

$$X_t = \mu + \sum_f \left[A(f)\cos(2\pi f t) + B(f)\sin(2\pi f t)\right]. \tag{6}$$

At first glance it might not seem possible to express time series such as those in Figure 2 in terms of sinusoids: these series appear to have random irregular bumps, whereas cosines and sines are deterministic regular oscillations. Figure 7 demonstrates that we can indeed get random-looking time series by combining sinusoids in a special way. The upper ten plots in the left-hand column show cosines (thick curves) and sines (thin) with equal amplitudes and with frequencies $f = (2l-1)/128$, $l = 1, 2, \ldots, 10$ (top to bottom). The bottom plot in that column shows a series $\{x_t\}$ that is equal to the sum of these twenty sinusoids; i.e.,

$$x_t = \sum_{l=1}^{10} \left[\cos(2\pi \tfrac{2l-1}{128} t) + \sin(2\pi \tfrac{2l-1}{128} t)\right], \quad t = 0, 1, \ldots, 127.$$

This artificial series is highly structured and does not particularly resemble any of our four actual series. We can, however, create series that are more random in appearance by introducing random amplitudes; i.e., we form

$$x_t = \sum_{l=1}^{10} \left[a_l\cos(2\pi \tfrac{2l-1}{128} t) + b_l\sin(2\pi \tfrac{2l-1}{128} t)\right], \quad t = 0, 1, \ldots, 127,$$

Cambridge University Press
978-1-107-02814-2 — Spectral Analysis for Univariate Time Series
Donald B. Percival , Andrew T. Walden
Excerpt
More Information

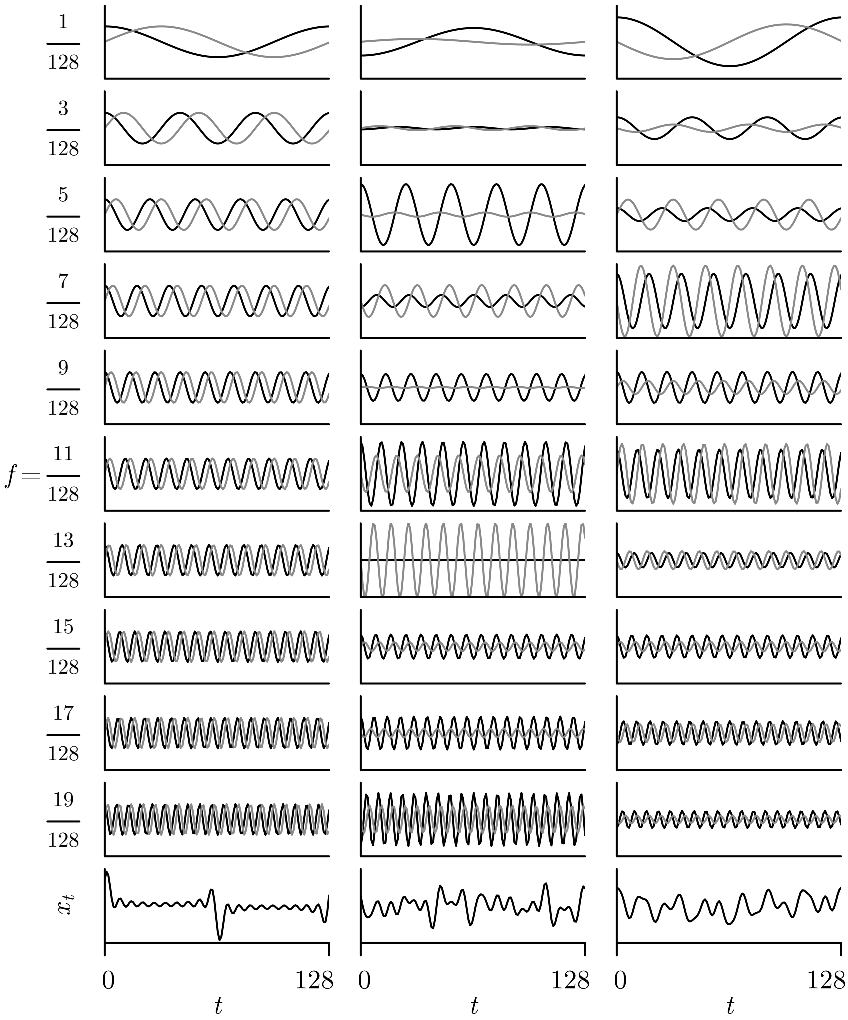

Figure 7 Sums of sinusoids with fixed and random amplitudes (see text for details).

where the a_l and b_l terms are twenty different realizations of uncorrelated Gaussian RVs with zero mean and unit variance. The upper ten plots in the middle column show $a_l \cos(2\pi \frac{2l-1}{128} t)$ (black curves) and $b_l \sin(2\pi \frac{2l-1}{128} t)$ (gray) versus t for $l = 1, 2, \ldots, 10$ (top to bottom) for one particular set of realizations, while the right-hand column shows similar plots for a second set. The bottom plots in each column show the sum x_t versus t. These artificial series are much more random in appearance, so indeed it does seem possible to construct irregular looking series out of sinusoids.

In general the summation in Equation (6) is a rather special one. To say what it means for the class of stationary processes is the subject of the spectral representation theorem (see Section 4.1). Fortunately, if we deal with a particularly simple (but unrealistic) model, we can say exactly what the summation means, define the spectrum in terms of elements involved in the summation, and thereby get an idea of what spectral analysis is all about. Let us assume that our time series can be modeled by a sum of a constant term μ and sinusoids with different fixed frequencies $\{f_j\}$ and random amplitudes $\{A_j\}$ and $\{B_j\}$ (the notation $\lfloor N/2 \rfloor$ refers to the greatest integer less than or equal to $N/2$):

$$X_t = \mu + \sum_{j=1}^{\lfloor N/2 \rfloor} \left[A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t) \right], \quad t = 0, 1, \ldots, N-1. \tag{8a}$$

Here we require that the frequencies of the sinusoids have a very special form, namely, that they be related to the sample size N in the following way:

$$f_j \stackrel{\text{def}}{=} j/N, \quad 1 \le j \le \lfloor N/2 \rfloor$$

(here and throughout this book the symbol $\stackrel{\text{def}}{=}$ means "equal by definition"). The frequency f_j is often called the jth standard (or Fourier) frequency; it is a cyclical frequency measured in cycles per unit time as opposed to an angular frequency $\omega_j \stackrel{\text{def}}{=} 2\pi f_j$, measured in radians per unit time. For example, f_j is measured in cycles per 0.025 sec for the wind speed series, while its units are cycles per month for the Willamette River series. We also assume that the amplitudes $\{A_j\}$ and $\{B_j\}$ are RVs with the following stipulations: for all j

$$E\{A_j\} = E\{B_j\} = 0 \text{ and } E\{A_j^2\} = E\{B_j^2\} = \sigma_j^2.$$

Thus the variance of the amplitudes associated with the jth standard frequency is just σ_j^2. We further assume that the A_j and B_j RVs are all mutually uncorrelated; i.e.,

$$E\{A_j A_k\} = E\{B_j B_k\} = 0 \text{ for } j \ne k \text{ and } E\{A_j B_k\} = 0 \text{ for all } j, k.$$

▷ **Exercise [8]** Show that $E\{X_t\} = \mu$, and then show that

$$E\{(X_{t+\tau} - \mu)(X_t - \mu)\} = \sum_{j=1}^{\lfloor N/2 \rfloor} \sigma_j^2 \cos(2\pi f_j \tau) \text{ and } \sigma^2 \stackrel{\text{def}}{=} E\{(X_t - \mu)^2\} = \sum_{j=1}^{\lfloor N/2 \rfloor} \sigma_j^2, \tag{8b}$$

from which we can conclude that

$$\rho_\tau = \frac{\sum_{j=1}^{\lfloor N/2 \rfloor} \sigma_j^2 \cos(2\pi f_j \tau)}{\sum_{j=1}^{\lfloor N/2 \rfloor} \sigma_j^2}. \tag{8c}$$ ◁

(We emphasize that we are considering models defined by Equation (8a) for pedagogical purposes only. These have a number of undesirable features, not the least of which is an explicit dependence of the component frequencies f_j on the sample size N.)

For this model we *define* the spectrum by

$$S_j \stackrel{\text{def}}{=} \sigma_j^2, \quad 1 \le j \le \lfloor N/2 \rfloor.$$

A plot of S_j versus f_j merely shows us the variances of the RVs that determine the amplitudes of the sinusoidal terms at the standard frequencies. From Equation (8b), we have the following fundamental relationship:

$$\sum_{j=1}^{\lfloor N/2 \rfloor} S_j = \sigma^2.$$

Thus, for a time series generated by the model in Equation (8a), the population variance, σ^2, can be regarded as being composed of a sum of a number of components, each of which is associated with a different nonzero standard frequency. The contribution to the variance due to the sinusoidal terms with frequency f_j is given by S_j. A study of S_j versus f_j indicates where the variability in a time series is likely to come from. In other words, the spectrum represents an analysis of the process variance σ^2 as the sum of variances associated with the Fourier frequencies.

Equation (8c) and the definition of the spectrum tell us that we can determine the ACS and σ^2 if we know the spectrum. Conversely, it can be shown (see Exercise [1.4]) that we can determine the spectrum if we know the ACS and σ^2. The spectrum is a frequency domain characterization for a model of a time series and is fully equivalent to the time domain characterization given by the ACS and σ^2.

For a model given by Equation (8a), it is easy to simulate a typical time series: as we did earlier in creating two of the time series shown in the bottom row of Figure 7, we use a random number generator on a computer to pick values for A_j and B_j and plug these into (8a) to generate a simulated time series. To illustrate this procedure, we will generate four such series using four different spectra. This exercise will show how a spectrum can be used to tell us something about the structure of an associated time series. The four spectra that we will use are actually rough models for the four time series in Figure 2 (for the moment we ignore the question of where these models came from). Figure 10a shows the four theoretical spectra; Figure 10b shows the corresponding ACSs (calculated via Equation (8c)); and Figure 11 shows a simulated time series that corresponds to each of the four spectra (we have set μ in Equation (8a) equal to the sample mean $\bar{x}$ for the corresponding series). If a proposed spectrum is a reasonable model for a time series, the corresponding theoretical ACS should resemble the sample ACS for the series, and simulated time series from that spectrum should have roughly the same visual properties as the actual time series. Here are some specifics about our four time series and these figures.

(a) For the wind speed data, we assume that S_j is large for $j = 1$ and then tapers off rapidly as j gets large. Thus, the low frequency terms in Equation (8a) – these correspond to sinusoids with long periods – should predominate. The theoretical ACS in plot (a) of Figure 10b is positive until lag 18. This picture agrees fairly well with the corresponding sample ACS in Figure 4, which is positive until lag 22. The appearance of the simulated time series is one of rather broad swoops together with some choppiness (evidently due to the higher frequencies in Equation (8a)). The wind speed series and the corresponding simulated series appear to have the same kind of bumpiness.

(b) For the atomic clock data, we assume S_j is large for $j = \lfloor N/2 \rfloor = 64$ and then tapers off rapidly as j decreases. Thus the high frequency terms (i.e., sinusoids with short periods) should predominate. The theoretical ACS oscillates between positive and negative values with an amplitude that is close to zero after the first few lags. The sample ACS in plot (b) of Figure 4 for these data shows more variability than this theoretical ACS (particularly for the higher lags), but the discrepancy might be due to sampling variation. The appearance of the simulated time series in Figure 11 is one of choppiness as the series swings back and forth from positive to negative values. The atomic clock data and the simulated series have the same "feel" to them.

(c) For the Willamette River data, we assume a spectrum that is constant except for a spike at $j = 11$. Since $f_{11} = 11/128$, this frequency corresponds to a period of $1/f_{11} = 128/11 \approx 11.6$ months. This is the frequency with a period closest to 1 year in our model (the next closest is f_{10} with a corresponding period of 12.8 months). We would thus expect terms with about this period to be predominant in Equation (8a). The generated

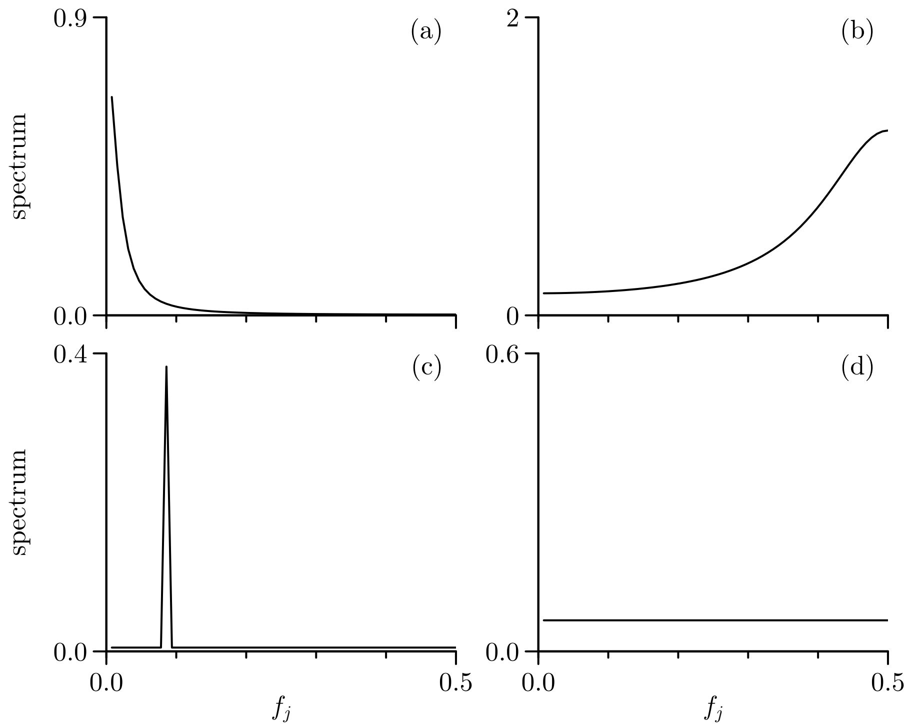

Figure 10a Plots of theoretical spectra S_j versus frequency f_j for models of four time series in Figure 2. The 64 values that determine each spectra are connected by solid lines. The horizontal axis represents frequency measured in cycles per unit time.

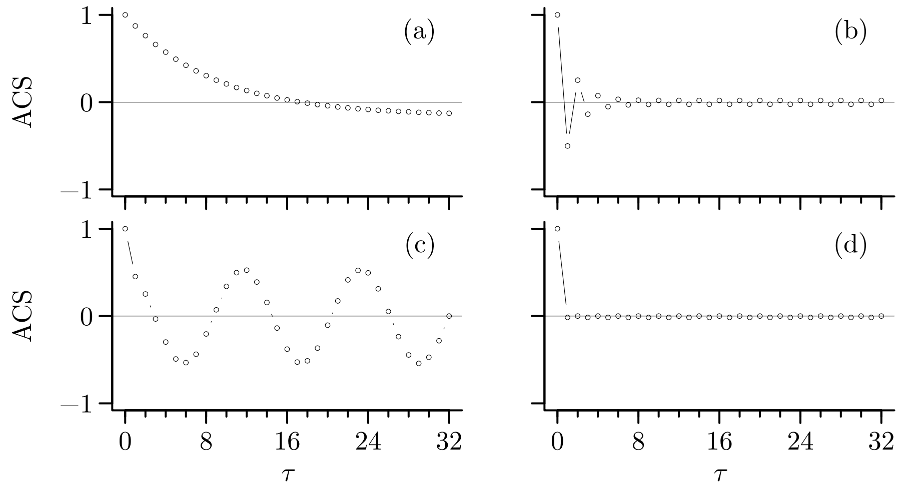

Figure 10b Plots of theoretical autocorrelation sequences of models for four time series in Figure 2 (cf. Figure 4).

time series should have a tendency to fluctuate with this period. This is roughly true for both the simulated series (Figure 11) and the Willamette River series. The sample ACS for the river flow data (Figure 4) and the theoretical ACS (Figure 10b) look fairly similar. (Here one of the limitations of our simple model is apparent: from physical considerations, it would make more sense to have a term corresponding to a frequency