Chapter 10

Time Series

10.1. BACKGROUND

This chapter presents methods for characterizing and analyzing the time variations of data series. Often we encounter data sets consisting of consecutive realizations of atmospheric variables. When the ordering of the data in time is important to their information content, summarization and analysis using time series methods are appropriate.

As has been illustrated earlier, atmospheric observations separated by relatively short times tend to be similar or correlated. Analyzing and characterizing the nature of these temporal correlations, or relationships through time, can be useful both for understanding the underlying physical processes and for forecasting future atmospheric events. Accounting for these correlations is also necessary if valid statistical inferences about time-series data are to be made (see Chapter 5).

10.1.1. Stationarity

Of course, we do not expect the future values of a data series to be identical to some past series of existing observations. However, in many instances it may be very reasonable to assume that their statistical properties will be the same. The idea that past and future values of a time series will be similar statistically is an informal expression of what is called *stationarity*. Usually, the term stationarity is understood to mean *weak stationarity* or *covariance stationarity*. In this sense, stationarity implies that the mean and autocovariance function (Equation 3.39) of the data series do not change through time. Different time slices of a stationary data series (e.g., the data observed to date and the data to be observed in the future) can be regarded as having the same underlying mean, variance, and covariances. Furthermore, the correlations between variables in a stationary series are determined only by their separation in time (i.e., their lag, k, in Equation 3.37) and not their absolute positions in time. Qualitatively, different portions of a stationary time series look alike statistically, even though the individual data values may be very different. Covariance stationarity is a less restrictive assumption than *strict stationarity*, which implies that the full joint distribution of the variables in the series does not change through time. More technical expositions of the concept of stationarity can be found in, for example, Fuller (1996) or Kendall and Ord (1990).

Most methods for analyzing time series assume stationarity of the data. However, many atmospheric processes are distinctly not stationary. Obvious examples of nonstationary atmospheric data series are those exhibiting annual or diurnal cycles. For example, temperatures typically exhibit very strong annual cycles in mid- and high-latitude climates, and we expect the average of the distribution of January temperature to be very different from that for July temperature. Similarly, time series of wind speeds often

Statistical Methods in the Atmospheric Sciences. https://doi.org/10.1016/B978-0-12-815823-4.00010-9

exhibit a diurnal cycle, which derives physically from the tendency for diurnal changes in static stability, imposing a diurnal cycle on the downward momentum transport.

There are two approaches to dealing with nonstationary series. Both aim to process the data in a way that will subsequently allow stationarity to be reasonably assumed. The first approach is to mathematically transform the nonstationary data to approximate stationarity. For example, subtracting a periodic mean function from data subject to an annual cycle would produce a transformed data series with constant (zero) mean. In order to produce a series with both constant mean and variance, it might be necessary to further transform these anomalies to standardized anomalies (Equation 3.27), that is, to divide the values in the anomaly series by standard deviations that also vary through an annual cycle. Not only do temperatures tend to be colder in winter, but the variability of temperature also tends to be higher. Data that become stationary after such cycles have been removed are said to exhibit *cyclostationarity*. A possible approach to transforming a monthly cyclostationary temperature series to (at least approximate) stationarity could be to compute the 12 monthly mean values and 12 monthly standard deviation values, and then to apply Equation 3.27 using the different means and standard deviations for the appropriate calendar month. This was the first step used to construct the time series of SOI values in Figure 3.16.

The alternative to data transformation is to stratify the data. That is, we can conduct separate analyses of subsets of the data record that are short enough to be regarded as nearly stationary. We might analyze daily observations for all available January records at a given location, assuming that each 31-day data record is a sample from the same physical process, but not necessarily assuming that process to be the same as for July, or even for February, data.

10.1.2. Time-Series Models

Characterization of the properties of a time series often is achieved by invoking a mathematical model for the data variations. Having obtained a time-series model for an observed data set, that model might then be viewed as a generating process, or algorithm, that could have produced the data. A mathematical model for the time variations of a data set can allow compact representation of the characteristics of those data in terms of a few parameters. This approach is entirely analogous to the fitting of parametric probability distributions, which constitute another kind of probability model, in Chapter 4. The distinction is that the distributions in Chapter 4 are used without regard to the ordering of the data, whereas the motivation for using time-series models is specifically to characterize the nature of the ordering. Time-series methods are thus appropriate when the ordering of the data values in time is important to a given application.

Regarding an observed time series as having been generated by a theoretical (i.e., model) process is convenient because it allows characteristics of future, and so still unobserved, values of a time series to be inferred from the inevitably limited data in hand. That is, characteristics of an observed time series are summarized by the parameters of a time-series model. Invoking the assumption of stationarity, future values of the time series should then also exhibit the statistical properties implied by the model, so that the properties of the model generating process can be used to infer characteristics of yet unobserved values of the series.

10.1.3. Time-Domain vs. Frequency-Domain Approaches

There are two fundamental approaches to time series analysis: *time-domain* analysis and *frequency-domain* analysis. Although these two approaches proceed very differently and may seem quite

distinct, they are not mutually independent. Rather, they are complementary views that are linked mathematically.

Time-domain methods seek to characterize data series in the same terms in which they are observed and reported. The autocorrelation function (Section 3.5.5) is a primary tool for characterization of relationships among data values in the time-domain approach. Mathematically, time-domain analyses operate in the same space as the data values. Separate sections in this chapter describe different time-domain methods for use with discrete and continuous data. Here the terms discrete and continuous are used in the same sense as in Chapter 4: discrete random variables are allowed to take on only a finite (or possibly countably infinite) number of values, and continuous random variables may take on any of the infinitely many real values within their range.

Frequency-domain analyses represent data series in terms of contributions occurring at different timescales or characteristic frequencies. Most commonly, each timescale is represented by a pair of sine and cosine functions. In that case, the overall time series is regarded as having arisen from the combined effects of a collection of sine and cosine waves oscillating at different rates. The sum of these waves reproduces the original data, but it is often the relative strengths of the individual component waves that are of primary interest. Frequency-domain analyses take place in the mathematical space defined by this collection of sine and cosine waves. That is, such frequency-domain analyses involve transformation of the n original data values into coefficients that multiply an equal number of periodic (the sine and cosine) functions. At first exposure this process can seem very strange and possibly difficult to grasp. However, frequency-domain methods very commonly are applied to atmospheric time series, and important insights can be gained from frequency-domain analyses.

10.2. TIME DOMAIN—I. DISCRETE DATA

10.2.1. Markov Chains

Recall that a discrete random variable is one that can take on only values from among a defined, finite or countably infinite set. The most common class of model, or stochastic process, used to represent time series of discrete variables is known as the *Markov chain*. A Markov chain can be imagined as being based on collection of "states" of a model system. Each state corresponds to one of the elements of the MECE partition of the sample space describing the random variable in question.

For each time period, the length of which is equal to the time separation between observations in the time series, the Markov chain can either remain in the same state or change to one of the other states. Remaining in the same state corresponds to two successive observations of the same value of the discrete random variable in the time series, and a change of state implies two successive values of the time series that are different.

The behavior of a Markov chain is governed by a set of probabilities for these transitions, called the *transition probabilities*. The transition probabilities specify conditional probabilities for the system being in each of its possible states during the next time period. The simplest form is called a first-order Markov chain, for which the transition probabilities controlling the next state of the system depend only on the current state of the system. That is, knowing the current state of the system and the full sequence of states leading up to the current state, provides no more information about the probability distribution for the states at the next observation time than does knowledge of the current state alone. This characteristic of first-order Markov chains is known as the *Markovian property*, which can be expressed more formally as

$$\Pr\{X_{t+1} | X_t, X_{t-1}, X_{t-2}, \ldots, X_1\} = \Pr\{X_{t+1} | X_t\}. \quad (10.1)$$

The probabilities of future states depend on the present state, but they do not depend on the particular way that the model system arrived at the present state. In terms of a time series of observed data the Markovian property means, for example, that forecasts of tomorrow's data value can be made on the basis of today's observation, and that also knowing yesterday's data value provides no additional information.

The transition probabilities of a Markov chain are conditional probabilities. That is, there is a conditional probability distribution pertaining to each possible current state, and each of these distributions specifies probabilities for the states of the system in the next time period. To say that these probability distributions are conditional allows for the possibility that the transition probabilities can be different, depending on the current state. The fact that these distributions can be different is the essence of the capacity of a Markov chain to represent the serial correlation, or persistence, often exhibited by atmospheric variables. If probabilities for future states are the same, regardless of the current state, then the time series consists of independent values. In that case the probability of occurrence of any given state in the upcoming time period would not be affected by the occurrence or nonoccurrence of a particular state in the current time period. If the time series being modeled exhibits persistence, the probability of the system staying in a given state will be higher than the probabilities of arriving at that state from other states, and higher than the corresponding unconditional probability.

If the transition probabilities of a Markov chain do not change through time and none of them are zero, then the resulting time series will be stationary. Modeling nonstationary data series exhibiting, for example, an annual cycle can require allowing the transition probabilities to vary through an annual cycle as well. One way to capture this kind of nonstationarity is to specify that the probabilities vary according to some smooth periodic curve, such as a cosine function, as described in Section 10.4. Alternatively, separate transition probabilities can be used for nearly stationary portions of the cycle, for example, each of the 12 calendar months.

Certain classes of Markov chains are described more concretely, but relatively informally, in the following sections. More formal and comprehensive treatments can be found in, for example, Feller (1970), Karlin and Taylor (1975), or Katz (1985).

10.2.2. Two-State, First-Order Markov Chains

The simplest kind of discrete random variable pertains to dichotomous (yes/no) events. The behavior of a stationary sequence of independent (exhibiting no serial correlation) values of a dichotomous discrete random variable is described by the binomial distribution (Equation 4.1). That is, for serially independent events, the ordering in time is of no importance from the perspective of specifying probabilities for future events, so that a time-series model for their behavior does not provide more information than does the binomial distribution.

A two-state Markov chain is a statistical model for the persistence of binary events. The occurrence or nonoccurrence of rain on a given day is a simple meteorological example of a binary random event, and a sequence of daily observations of "rain" and "no rain" for a particular location would constitute a time series of that variable. Consider a series where the random variable takes on the values $x_t = 1$ if precipitation occurs on day t and $x_t = 0$ if it does not. For the January 1987 Ithaca precipitation data in Table A.1, this time series would consist of the values presented in Table 10.1. That is, $x_1 = 0$, $x_2 = 1$, $x_3 = 1$, $x_4 = 0$, … , and $x_{31} = 1$. It is evident from looking at this series of numbers that the 1's and 0's

TABLE 10.1 Time Series of a Dichotomous Random Variable Derived from the January 1987 Ithaca Precipitation Data in Table A.1

Date, t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
x_t	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	1

Days on which nonzero precipitation was reported yield $x_t = 1$, and days with zero precipitation yield $x_t = 0$.

tend to cluster in time. As was illustrated in Example 2.2, this clustering is an expression of the serial correlation present in the time series, which in turn reflects the fact that the daily measurement interval is shorter than the characteristic timescale of the underlying physical data-generating process. That is, the probability of a 1 following a 1 is apparently higher than the probability of a 1 following a 0, and the probability of a 0 following a 0 is apparently higher than the probability of a 0 following a 1.

A first-order, two-state Markov chain is a common and often quite good stochastic model for data of this kind. A two-state Markov chain is natural for dichotomous data since each of the two states will pertain to one of the two possible data values. A first-order Markov chain has the property that the transition probabilities governing each observation in the time series depend only on the value of the previous member of the time series.

Figure 10.1 illustrates schematically the nature of a two-state, first-order Markov chain. In order to help fix ideas, the two states are labeled in a manner consistent with the data in Table 10.1. For each value of the time series, the stochastic process is either in state 0 (no precipitation occurs and $x_t = 0$) or in state 1 (precipitation occurs and $x_t = 1$). At each time step the process can either stay in the same state or switch to the other state. Therefore four distinct transitions are possible, corresponding to a dry day following a dry day (p_{00}), a wet day following a dry day (p_{01}), a dry day following a wet day (p_{10}), and a wet day following a wet day (p_{11}). Each of these four transitions is represented in Figure 10.1 by arrows, labeled with the appropriate transition probabilities. Here the notation is such that the first subscript on the probability is the state at time t, and the second subscript is the state at time $t+1$.

The transition probabilities are conditional probabilities for the state at time $t+1$ (e.g., whether precipitation will occur tomorrow) given the state at time t (e.g., whether or not precipitation occurred today). That is,

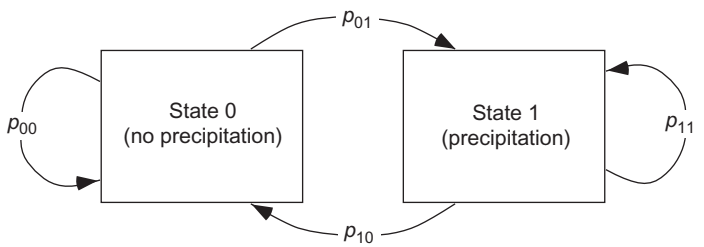

FIGURE 10.1 Schematic representation of a two-state, first-order Markov chain, illustrated in terms of daily precipitation occurrence or nonoccurrence. The two states are labeled "0" for no precipitation, and "1" for precipitation occurrence. For a first-order Markov chain, there are four transition probabilities controlling the state of the system in the next time period. Since these four probabilities are pairs of conditional probabilities, $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$. For quantities like day-to-day precipitation occurrence that exhibit positive serial correlation, $p_{01} < p_{00}$, and $p_{01} < p_{11}$.

$$p_{00} = \Pr\{X_{t+1} = 0 | X_t = 0\} \tag{10.2a}$$

$$p_{01} = \Pr\{X_{t+1} = 1 | X_t = 0\} \tag{10.2b}$$

$$p_{10} = \Pr\{X_{t+1} = 0 | X_t = 1\} \tag{10.2c}$$

and

$$p_{11} = \Pr\{X_{t+1} = 1 | X_t = 1\}. \tag{10.2d}$$

Together, Equations 10.2a and 10.2b define the conditional probability distribution for the value of the time series at time $t+1$, given that $X_t = 0$ at time t. Similarly, Equations 10.2c and 10.2d express the conditional probability distribution for the next value of the time series given that the current value is $X_t = 1$.

Notice that the four probabilities in Equation 10.2 provide some redundant information. Given that the Markov chain is in one state or the other at time t, the sample space for X_{t+1} consists of only two MECE events. Therefore $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$, so that it is really only necessary to focus on one of each of the pairs of transition probabilities, say p_{01} and p_{11}. In particular, it is sufficient to estimate only two parameters for a two-state first-order Markov chain, since each of the two pairs of conditional probabilities must sum to 1. The parameter estimation procedure usually consists simply of computing the conditional relative frequencies, which yield the maximum likelihood estimators (MLEs)

$$\hat{p}_{01} = \frac{\#\text{ of 1's following 0's}}{\text{Total \# of 0's}} = \frac{n_{01}}{n_{0\bullet}} = \frac{n_{01}}{n_{00} + n_{01}} \tag{10.3a}$$

and

$$\hat{p}_{11} = \frac{\#\text{ of 1's following 1's}}{\text{Total \# of 1's}} = \frac{n_{11}}{n_{1\bullet}} = \frac{n_{11}}{n_{10} + n_{11}}. \tag{10.3b}$$

Here n_{01} is the number of transitions from State 0 to State 1, n_{11} is the number of pairs of time steps in which there are two consecutive 1's in the series, $n_{0\bullet}$ is the number of 0's in the series followed by another data point, and $n_{1\bullet}$ is the number of 1's in the series followed by another data point. That is, the subscript "$\bullet$" indicates the total over all values of the index replaced by this symbol, so that $n_{1\bullet} = n_{10} + n_{11}$ and $n_{0\bullet} = n_{00} + n_{01}$. Equations 10.3 state that the parameter p_{01} is estimated by looking at the relative frequency of the event $X_{t+1} = 1$ considering only those points in the time series following data values for which $X_t = 0$. Similarly, p_{11} is estimated as the fraction of points for which $X_t = 1$ that are followed by points with $X_{t+1} = 1$. These somewhat labored definitions of $n_{0\bullet}$ and $n_{1\bullet}$ are necessary to account for the edge effects in a finite sample. The final point in the time series is not counted in the denominator of Equation 10.3a or 10.3b, whichever is appropriate, because there is no available data value following it to be incorporated into the counts in one of the numerators. These definitions also cover cases of missing values, and stratified samples such as 30 years of January data, for example.

Equation 10.3 suggests that parameter estimation for a two-state first-order Markov chain is equivalent to fitting two Bernoulli distributions (i.e., binomial distributions with $N = 1$). One of these binomial distributions pertains to points in the time series preceded by 0's, and the other describes the behavior of points in the time series preceded by 1's. Knowing that the process is currently in state 0 (e.g., no precipitation today), the probability distribution for the event $X_{t+1} = 1$ (precipitation tomorrow) is simply binomial (Equation 4.1) with $p = p_{01}$. The second binomial parameter is $N = 1$, because there is only one data point in the series for each time step. Similarly, if $X_t = 1$, then the distribution for the event $X_{t+1} = 1$ is binomial with $N = 1$ and $p = p_{11}$. The conditional dichotomous events of a stationary Markov chain satisfy the requirements listed in Section 4.2.1 for the binomial distribution. For a stationary process the probabilities do not change through time, and conditioning on the current value of the time

series satisfies the independence assumption for the binomial distribution because of the Markovian property. It is the fitting of two Bernoulli distributions that allows the time dependence in the data series to be represented.

Certain properties are implied for a time series described by a Markov chain. These properties are controlled by the values of the transition probabilities and can be computed from them. First, the long-run relative frequencies of the events corresponding to the two states of the Markov chain are called the *stationary probabilities*. For a Markov chain describing the daily occurrence or nonoccurrence of precipitation, the stationary probability for precipitation, π_1, corresponds to the (unconditional) climatological probability of precipitation. In terms of the transition probabilities p_{01} and p_{11},

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}}, \tag{10.4}$$

with the stationary probability for state 0 being simply $\pi_0 = 1 - \pi_1$. The usual situation of positive serial correlation or persistence produces $p_{01} < \pi_1 < p_{11}$. Applied to daily precipitation occurrence, this relationship means that the conditional probability of a wet day following a dry day is less than the overall climatological relative frequency, which in turn is less than the conditional probability of a wet day following a wet day.

The transition probabilities also imply a specific degree of serial correlation, or persistence, for the binary time series. In terms of the transition probabilities, the lag-1 autocorrelation (Equation 3.36) of the binary time series is simply

$$r_1 = p_{11} - p_{01}. \tag{10.5}$$

In the context of Markov chains, r_1 is sometimes known as the *persistence parameter*. As the correlation r_1 increases, the difference between p_{11} and p_{01} widens, so that state 1 is more and more likely to follow state 1, and less and less likely to follow state 0. That is, there is an increasing tendency for 0's and 1's to cluster in time, or occur in runs. A time series exhibiting no autocorrelation would be characterized by $r_1 = p_{11} - p_{01} = 0$, and $p_{11} = p_{01} = \pi_1$. In that case the two conditional probability distributions specified by Equation 10.2 are the same, and the time series is simply a string of independent Bernoulli realizations. The Bernoulli distribution can be viewed as defining a two-state, zero-order Markov chain.

Once the state of a Markov chain has changed, the number of time periods it will remain in the new state is a random variable, with a probability distribution function. Because the conditional independence implies conditional Bernoulli distributions, this probability distribution function for numbers of consecutive time periods in the same state, or the "spell length," will be the geometric distribution (Equation 4.5), with $p = p_{01}$ for sequences of 0's (dry spells), and $p = p_{10} = 1 - p_{11}$ for sequences of 1's (wet spells).

The full autocorrelation function, Equation 3.37, for the first-order Markov chain follows easily from the lag-1 autocorrelation r_1. Because of the Markovian property, the autocorrelation between members of the time series separated by k time steps is simply the lag-1 autocorrelation multiplied by itself k times,

$$r_k = (r_1)^k. \tag{10.6}$$

A common misconception is that the Markovian property implies independence of values in a first-order Markov chain that are separated by more than one time period. Equation 10.6 shows that the correlation, and hence the statistical dependence, among elements of the time series tails off at increasing lags, but it is never exactly zero unless $r_1 = 0$. Rather, the Markovian property implies conditional independence of data values separated by more than one time period, as expressed by Equation 10.1. Given a particular value for x_t, the different possible values for x_{t-1}, x_{t-2}, x_{t-3}, and so on, do not affect the probabilities for

x_{t+1}. However, for example, $\Pr\{x_{t+1} = 1 \mid x_{t-1} = 1\} \neq \Pr\{x_{t+1} = 1 \mid x_{t-1} = 0\}$, indicating statistical dependence among members of a Markov chain separated by more than one time period. Put another way, it is not that the Markov chain has no memory of the past, but rather that it is only the recent past that matters.

10.2.3. Test for Independence vs. First-Order Serial Dependence

Even if a series of binary data is generated by a mechanism producing serially independent values, the sample lag-one autocorrelation (Equation 3.36) computed from a finite sample is unlikely to be exactly zero. A formal test, similar to the χ^2 goodness-of-fit test (Equation 5.14), can be computed to investigate the statistical significance of the sample autocorrelation for a binary data series. The null hypothesis for this test is that the data series is serially independent (i.e., the data are independent Bernoulli variables), with the alternative being that the series was generated by a first-order Markov chain.

The test is based on a contingency table of the observed transition counts n_{00}, n_{01}, n_{10}, and n_{11}, in relation to the numbers of transitions expected under the null hypothesis. The corresponding expected counts, e_{00}, e_{01}, e_{10}, and e_{11}, are computed from the observed transition counts under the constraint that the marginal totals of the expected counts are the same as for the observed transitions. The comparison is illustrated in Figure 10.2, which shows generic contingency tables for the observed transition counts (a) and those expected under the null hypothesis of independence (b). For example, the transition count n_{00} specifies the number of consecutive pairs of 0's in the time series. This is related to the joint probability $\Pr\{X_t = 0 \cap X_{t+1} = 0\}$. Under the null hypothesis of independence this joint probability is simply the product of the two event probabilities, or in relative frequency terms, $\Pr\{X_t = 0\}\ \Pr\{X_{t+1} = 0\} = (n_{0\bullet}/n)\ (n_{\bullet 0}/n)$. Thus the corresponding number of expected transition counts is simply this product multiplied by the sample size, or $e_{00} = (n_{0\bullet})(n_{\bullet 0})/n$. More generally,

$$e_{ij} = \frac{n_{i\bullet}\, n_{\bullet j}}{n}. \tag{10.7}$$

The test statistic is computed from the observed and expected transition counts using

$$\chi^2 = \sum_i \sum_j \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}}, \tag{10.8}$$

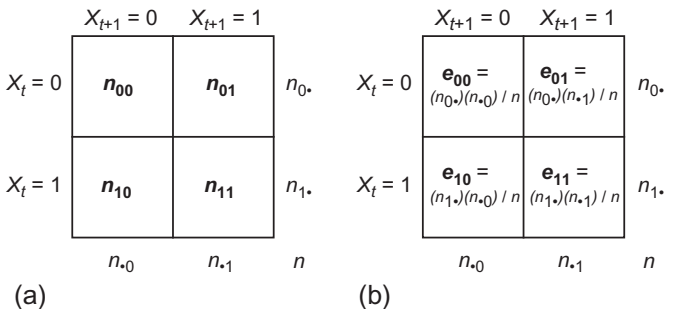

FIGURE 10.2 Contingency tables of observed transition counts n_{ij} (a) for a binary time series, and (b) transition counts e_{ij} expected if the time series actually consists of serially independent values with the same marginal totals. The transition counts are shown in boldface, and the marginal totals are in plain type.

where, for the 2×2 contingency table appropriate for dichotomous data, the summations are for $i = 0$ to 1 and $j = 0$ to 1. That is, there is a separate term in Equation 10.8 for each of the four pairs of contingency table cells in Figure 10.2. Note that Equation 10.8 is analogous to Equation 5.14, with the n_{ij} being the observed counts, and the e_{ij} being the expected counts. Under the null hypothesis, the test statistic follows the χ^2 distribution with $\nu = 1$ degree of freedom. This value of the degrees-of-freedom parameter is appropriate because, given that the marginal totals are fixed, arbitrarily specifying one of the transition counts completely determines the other three.

The fact that the numerator in Equation 10.8 is squared implies that values of the test statistic on the left tail of the null distribution are favorable to H_0, because small values of the test statistic are produced by pairs of observed and expected transition counts of similar magnitudes. Therefore the test is one-tailed. The p value associated with a particular test can be assessed using the χ^2 quantiles in Table B.3.

Example 10.1. Fitting a Two-State, First Order Markov Chain

Consider summarizing the time series in Table 10.1, derived from the January 1987 Ithaca precipitation series in Table A1, using a first-order Markov chain. The parameter estimates in Equation 10.3 are obtained easily from the transition counts. For example, the number of 1's following 0's in the time series of Table 10.1 is $n_{01} = 5$. Similarly, $n_{00} = 11$, $n_{10} = 4$, and $n_{11} = 10$. The resulting sample estimates for the transition probabilities (Equations 10.3) are $p_{01} = 5/16 = 0.312$, and $p_{11} = 10/14 = 0.714$. Note that these are identical to the conditional probabilities computed in Example 2.2.

Whether fitting a first-order Markov chain to the data in Table 10.1 is justified can be investigated using the χ^2 test in Equation 10.8. Here the null hypothesis is that these data resulted from an independent (i.e., Bernoulli) process, and the expected transition counts e_{ij} that must be computed are those consistent with this null hypothesis. These are obtained from the marginal totals $n_{0\bullet} = 11+5 = 16$, $n_{1\bullet} = 4+10 = 14$, $n_{\bullet 0} = 11+4 = 15$, and $n_{\bullet 1} = 5+10 = 15$. The expected transition counts follow easily as $e_{00} = (16)(15)/30 = 8$, $e_{01} = (16)(15)/30 = 8$, $e_{10} = (14)(15)/30 = 7$, and $e_{11} = (14)(15)/30 = 7$. Note that usually the expected transition counts will be different from each other and need not be integer values.

Computing the test statistic in Equation 10.8, we find $\chi^2 = (11–8)^2/8 + (5–8)^2/8 + (4–7)^2/7 + (10–7)^2/7 = 4.82$. The degree of unusualness of this result with reference to the null hypothesis can be assessed with the aid of Table B.3. Looking on the $\nu = 1$ row, the result lies between the 95th and 99th percentiles of the appropriate χ^2 distribution. Thus even for this rather small sample size, the null hypothesis of serial independence would be rejected at the 5% level, although not at the 1% level.

The degree of serial correlation exhibited by this data sample can be summarized using the persistence parameter, which is also the lag-one autocorrelation, $r_1 = p_{11}–p_{01} = 0.714 – 0.312 = 0.402$. This value could also be obtained by operating on the series of 0's and 1's in Table 10.1, using Equation 3.36. This lag-1 autocorrelation is fairly large, indicating substantial serial correlation in the time series. Assuming first-order Markov dependence, it also implies the full autocorrelation function, through Equation 10.6. Figure 10.3 shows that the implied theoretical correlation function for this Markov process, indicated by the dashed line, agrees very closely with the sample autocorrelation function shown by the solid line, for the first few lags. This agreement provides qualitative support for the first-order Markov chain as an appropriate model for the data series.

Finally, the stationary (i.e., climatological) probability for precipitation implied for these data by the Markov chain model is, using Equation 10.4, $\pi_1 = 0.312/(1+0.312–0.714) = 0.522$. This value agrees closely with the relative frequency $16/30 = 0.533$, obtained by counting the number of 1's in the last 30 values of the series in Table 10.1. ◇

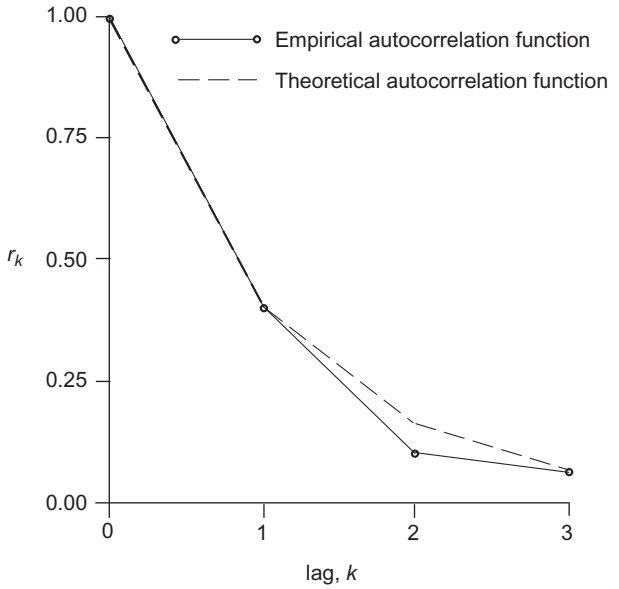

FIGURE 10.3 Sample autocorrelation function for the January 1987 Ithaca binary precipitation occurrence series, Table 10.1 (solid, with circles), and theoretical autocorrelation function (dashed) specified by the fitted first-order Markov chain model (Equation 10.6). The correlations are 1.00 for $k = 0$, since the unlagged data are perfectly correlated with themselves.

10.2.4. Some Applications of Two-State Markov Chains

One interesting application of the Markov chain model is in the computer generation of synthetic rainfall series. Time series of random binary numbers, statistically resembling real rainfall occurrence data, can be produced using a Markov chain as the generating algorithm. This procedure is an extension of the ideas presented in Section 4.7, to time-series data. To generate sequences of numbers statistically resembling those in Table 10.1, for example, the parameters $p_{01} = 0.312$ and $p_{11} = 0.714$, estimated in Example 10.1, would be used together with a uniform [0, 1] random number generator (see Section 4.7.1). The synthetic time series would begin using the stationary probability $\pi_1 = 0.522$. If the first uniform number generated were less than π_1, then $x_1 = 1$, meaning that the first simulated day would be wet. For subsequent values in the series, corresponding to day $t+1$, each new uniform random number would be compared to the appropriate transition probability, depending on whether the most recently generated number, corresponding to day t, was wet or dry. That is, the transition probability p_{01} would be used to generate x_{t+1} if $x_t = 0$, and p_{11} would be used if $x_t = 1$. A wet day ($x_{t+1} = 1$) is simulated if the next uniform random number is less than the transition probability, and a dry day ($x_{t+1} = 0$) is generated if it is not. Since typically $p_{11} > p_{01}$ for daily precipitation occurrence data, simulated wet days are more likely to follow wet days than dry days, as is the case in the real data series.

The Markov chain approach for simulating precipitation occurrence can be extended to include simulation of daily precipitation amounts. This is accomplished by adopting a statistical model for the nonzero rainfall amounts, yielding a sequence of random variables defined on the Markov chain, called a *chain-dependent process* (Katz, 1977; Todorovic and Woolhiser, 1975). Often a gamma distribution (Section 4.4.5) is fit to the precipitation amounts on wet days in the data record (e.g., Katz, 1977; Richardson, 1981; Stern and Coe, 1984), although the mixed exponential distribution (Equation 4.78) often provides a better fit to nonzero daily precipitation data (e.g., Foufoula-Georgiou and

Lettenmaier, 1987; Wilks, 1999a; Woolhiser and Roldan, 1982). Computer algorithms are available to generate random variables drawn from gamma distributions (e.g., Bratley et al., 1987; Johnson, 1987), or together Example 4.15 and Section 4.7.5 can be used to simulate from the mixed exponential distribution, to produce synthetic precipitation amounts on days when the Markov chain calls for a wet day. The tacit assumption that precipitation amounts on consecutive wet days are independent has turned out to be a reasonable approximation in most instances where it has been investigated (e.g., Katz, 1977; Stern and Coe, 1984), but may not adequately simulate extreme multiday precipitation events that could arise, for example, from a slow-moving landfalling hurricane (Wilks, 2002a). Generally both the Markov chain transition probabilities and the parameters of the distributions describing precipitation amounts change through the year. These seasonal cycles can be handled by fitting separate sets of parameters for each of the 12 calendar months (e.g., Wilks, 1989), or by representing them using smoothly varying sine and cosine functions (Stern and Coe, 1984).

Properties of longer term precipitation quantities resulting from simulated daily series (e.g., the monthly frequency distributions of numbers of wet days in a month, or of total monthly precipitation) can be calculated from the parameters of the chain-dependent process that governs the daily precipitation series. Since observed monthly precipitation statistics are computed from individual daily values, it should not be surprising that the statistical characteristics of monthly precipitation quantities will depend directly on the statistical characteristics of daily precipitation occurrences and amounts. Katz (1977, 1985) gives equations specifying some of these relationships, which can be used in a variety of ways (e.g., Katz and Parlange, 1993; Wilks, 1992, 1999b; Wilks and Wilby, 1999).

Finally, another interesting perspective on the Markov chain model for daily precipitation occurrence is in relation to forecasting precipitation probabilities. Recall that forecast skill is assessed relative to a set of benchmark, or reference forecasts (Equation 9.4). Usually one of two reference forecasts is used: either the climatological probability of the forecast event, in this case π_1; or persistence forecasts specifying unit probability if precipitation occurred in the previous period, or zero probability if the event did not occur. Neither of these reference forecasting systems is particularly sophisticated, and both are relatively easy to improve upon, at least for short-range forecasts. A more challenging, yet still fairly simple alternative is to use the transition probabilities of a two-state Markov chain as the reference forecasts. If precipitation did not occur in the preceding period, the reference forecast would be p_{01}, and the conditional reference forecast probability for precipitation following a day with precipitation would be p_{11}. Note that for quantities exhibiting persistence, $0 < p_{01} < \pi_1 < p_{11} < 1$, so that reference forecasts consisting of Markov chain transition probabilities constitute a compromise between the persistence (either 0 or 1) and climatological (π_1) probabilities. Furthermore, the balance of this compromise depends on the strength of the persistence exhibited by the climatological data on which the estimated transition probabilities are based. A weakly persistent quantity would be characterized by transition probabilities differing little from π_1, whereas strong serial correlation will produce transition probabilities much closer to 0 and 1.

10.2.5. Multiple-State Markov Chains

Markov chains are also useful for representing the time correlation of discrete variables that can take on more than two values. For example, a three-state, first-order Markov chain is illustrated schematically in Figure 10.4. Here the three states are arbitrarily labeled 1, 2, and 3. At each time t, the random variable in the series can take on one of the three values $x_t = 1$, $x_t = 2$, or $x_t = 3$, and each of these values corresponds

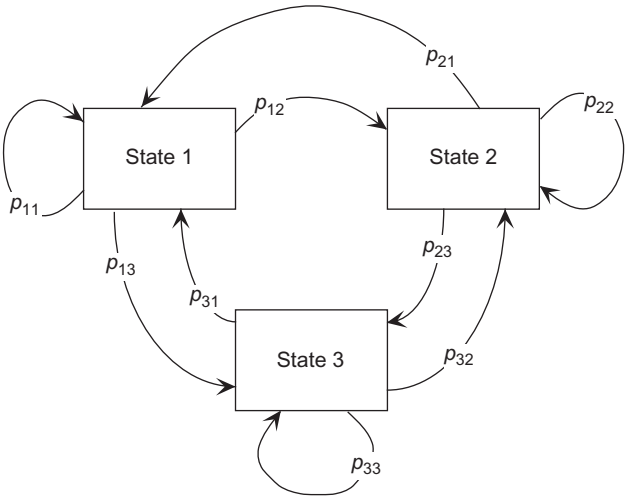

FIGURE 10.4 Schematic illustration of a three-state, first-order Markov chain. There are nine possible transitions among the three states, including the possibility that two consecutive points in the time series will be in the same state. First-order time dependence implies that the transition probabilities depend only on the current state of the system, or present value of the time series.

to a different state. First-order time dependence implies that the transition probabilities for x_{t+1} depend only on the state x_t, so that there are $3^2 =$ nine transition probabilities, p_{ij}. In general, for a first-order, s-state Markov chain, there are s^2 transition probabilities.

As is also the case for the two-state Markov chain, the transition probabilities for multiple-state Markov chains are conditional probabilities. For example, the transition probability p_{12} in Figure 10.4 is the conditional probability that state 2 will occur at time $t+1$, given that state 1 occurred at time t. Therefore in an s-state Markov chain it must be the case that the probabilities for the s transitions emanating from each state must sum to one, or $\Sigma_j\, p_{ij} = 1$ for each value of i.

Estimation of the transition probabilities for multiple-state Markov chains is a straightforward generalization of the formulas in Equations 10.3 for two-state chains. Each of these estimates is simply obtained from the conditional relative frequencies of the transition counts,

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i\bullet}}, \quad i,j = 1,\ldots,s. \tag{10.9}$$

As before, the dot indicates summation over all values of the replaced subscript so that, for example, $n_{1\bullet} = \Sigma_j\, n_{1j}$. For the $s = 3$-state Markov chain represented in Figure 10.4, for example, $\hat{p}_{12} = n_{12}/(n_{11}+n_{12}+n_{13})$. In general, a contingency table of transition counts, corresponding to Figure 10.2a for the $s = 2$-state case, will contain s^2 entries.

Testing whether the observed degree of serial correlation is significantly different from zero in a multiple-state situation can be done using the χ^2 test in Equation 10.8. Here the summations are over all s possible states and will include s^2 terms. As before, the expected numbers of transition counts e_{ij} are computed using Equation 10.7. Under the null hypothesis of no serial correlation, the distribution of the test statistic in Equation 10.8 is χ^2 with $\nu = (s-1)^2$ degrees of freedom.

Three-state Markov chains have been used to characterize transitions between below-normal, near-normal, and above-normal months, as defined by the U.S. Climate Prediction Center (see Example 4.9), by Preisendorfer and Mobley (1984) and Wilks (1989). Mo and Ghil (1987) used a five-state Markov chain to characterize transitions between persistent hemispheric 500-mb flow types.

10.2.6. Higher-Order Markov Chains

First-order Markov chains often provide good representations of daily precipitation occurrences, but it is not obvious just from inspection of the series in Table 10.1, for example, that this simple model will be adequate to capture the observed correlation structure. More generally, an mth order Markov chain is one where the transition probabilities depend on the states in the previous m time periods. Formally, the extension of the Markovian property expressed in Equation 10.1 to the mth order Markov chain is

$$\Pr\{X_{t+1} | X_t, X_{t-1}, X_{t-2}, \ldots, X_1\} = \Pr\{X_{t+1} | X_t, X_{t-1}, \ldots, X_{t-m}\}. \tag{10.10}$$

Consider, for example, a second-order Markov chain. Second-order time dependence means that the transition probabilities depend on the states (values of the time series) at lags of both one and two time periods. Notationally, then, the transition probabilities for a second-order Markov chain require three subscripts: the first denotes the state at time t–1, the second denotes the state in time t, and the third specifies the state at (the future) time $t+1$. The notation for the transition probabilities of a second-order Markov chain can be defined as

$$p_{hij} = \{X_{t+1} = j | X_t = i, X_{t-1} = h\}. \tag{10.11}$$

In general, the notation for an mth order Markov chain requires $m+1$ subscripts on the transition counts and transition probabilities. If Equation 10.11 is being applied to a binary time series such as that in Table 10.1, the model would be a two-state, second-order Markov chain, and the indices h, i, and j could take on either of the $s = 2$ values of the time series, say 0 and 1. However, Equation 10.11 is equally applicable to discrete time series with larger numbers ($s > 2$) of states.

As is the case for first-order Markov chains, transition probability estimates are obtained from relative frequencies of observed transition counts. However, since data values further back in time now need to be considered, the number of possible transitions increases exponentially with the order, m, of the Markov chain. In particular, for an s-state, mth order Markov chain, there are $s^{(m+1)}$ distinct transition counts and transition probabilities. The arrangement of the resulting transition counts, in the form of Figure 10.2a, is presented in Table 10.2 for an $s = 2$ state, $m =$ second-order Markov chain. The transition counts are determined from the observed data series by examining consecutive groups of $m+1$ data points. For example, the first three data points in Table 10.1 are $x_{t-1} = 0$, $x_t = 1$, $x_{t+1} = 1$, and this triplet would contribute one to the transition count n_{011}. Overall the data series in Table 10.1 exhibits three transitions of this kind, so the final transition count $n_{011} = 3$ for this data set. The second triplet in

TABLE 10.2 Arrangement of the $2^{(2+1)} = 8$ Transition Counts for a Two-State, Second-Order Markov Chain in a Table of the Form of Figure 10.2a

X_{t-1}	X_t	$X_{t+1} = 0$	$X_{t+1} = 1$	Marginal Totals
0	0	n_{000}	n_{001}	$n_{00\cdot} = n_{000} + n_{001}$
0	1	n_{010}	n_{011}	$n_{01\cdot} = n_{010} + n_{011}$
1	0	n_{100}	n_{101}	$n_{10\cdot} = n_{100} + n_{101}$
1	1	n_{110}	n_{111}	$n_{11\cdot} = n_{110} + n_{111}$

Determining these counts from an observed time series requires examination of successive triplets of data values.

the data set in Table 10.1 would contribute one count to n_{110}. There is only one other triplet in this data for which $x_{t-1} = 1$, $x_t = 1$, $x_{t+1} = 0$, yielding the final count for $n_{110} = 2$.

The transition probabilities for a second-order Markov chain are obtained from the conditional relative frequencies of the transition counts

$$\hat{p}_{hij} = \frac{n_{hij}}{n_{hi\bullet}}. \tag{10.12}$$

That is, given the value of the time series at time t–1 was $x_{t-1} = h$ and the value of the time series at time t was $x_t = i$, the probability that the next value of the time series $x_{t+1} = j$ is p_{hij}, and the sample estimate of this probability is given in Equation 10.12. Just as the two-state first-order Markov chain consists essentially of two conditional Bernoulli distributions, a two-state second-order Markov chain amounts to four conditional Bernoulli distributions, with parameters $p = p_{hi1}$, for each of the four distinct combinations of the indices h and i.

Note that the small data set in Table 10.1 is really too short to fit a second-order Markov chain. Since there are no consecutive triplets in this series for which $x_{t-1} = 1, x_t = 0$, and $x_{t+1} = 1$ (i.e., a single dry day following and followed by a wet day) the transition count $n_{101} = 0$. This zero transition count would lead to the sample estimate for the transition probability $\hat{p}_{101} = 0$, even though there is no physical reason why that particular sequence of wet and dry days could not or should not occur.

10.2.7. Deciding Among Alternative Orders of Markov Chains

How are we to know what order m for a Markov chain is appropriate to represent a particular data series? One approach is to use a hypothesis test. For example, the χ^2 test in Equation 10.8 can be used to assess the plausibility of a first-order Markov chain model versus a null zero-order, or binomial model. The mathematical structure of this test can be modified to investigate the suitability of, say, a first-order versus a second-order, or a second-order versus a third-order Markov chain, but the overall statistical significance of a collection of such tests would be difficult to evaluate. This difficulty arises in part because of the issue of test multiplicity. As discussed in Section 5.4, the level of the strongest of a collection of simultaneous, correlated tests is difficult if not impossible to evaluate.

Two criteria are in common use for choosing among alternative orders of Markov chain models. These are the *Akaike Information Criterion* (AIC) (Akaike, 1974; Tong, 1975) and the *Bayesian Information Criterion* (BIC) (Schwarz, 1978; Katz, 1981). Both are based on the log-likelihood functions. In the present context these log-likelihoods depend on the transition counts and the estimated transition probabilities. The log-likelihoods for s-state Markov chains of order 0, 1, 2, and 3 are

$$L_0 = \sum_{j=0}^{s-1} n_j \ln\left(\hat{p}_j\right) \tag{10.13a}$$

$$L_1 = \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} n_{ij} \ln\left(\hat{p}_{ij}\right) \tag{10.13b}$$

$$L_2 = \sum_{h=0}^{s-1}\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} n_{hij} \ln\left(\hat{p}_{hij}\right) \tag{10.13c}$$

and

$$L_3 = \sum_{g=0}^{s-1}\sum_{h=0}^{s-1}\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} n_{ghij} \ln\left(\hat{p}_{ghij}\right), \tag{10.13d}$$

with obvious extension for fourth and higher-order Markov chains. Here the summations are over all s states of the Markov chain, and so will include only two terms each for two-state (binary) time series. Equation 10.13a is simply the log-likelihood for the independent binomial model.

Example 10.2. Likelihood Ratio Test for the Order of a Markov Chain

To illustrate the application of Equations 10.13, consider a likelihood ratio test of first-order dependence of the binary time series in Table 10.1 versus the null hypothesis of zero serial correlation. The test involves computation of the log-likelihoods in Equations 10.13a and 10.13b. The resulting two log-likelihoods are compared using the test statistic given by Equation 5.20.

In the last 30 data points in Table 10.1, there are $n_0 = 15$ 0's and $n_1 = 15$ 1's, yielding the unconditional relative frequencies of rain and no rain $\hat{p}_0 = 15/30 = 0.500$ and $\hat{p}_1 = 15/30 = 0.500$, respectively. The last 30 points are used because the first-order Markov chain amounts to two conditional Bernoulli distributions, given the previous day's value, and the value for 31 December 1986 is not available in Table A.1. The log-likelihood in Equation 10.13a for these data is $L_0 = 15\ \ln(0.500) + 15\ \ln(0.500) = -20.79$. Values of n_{ij} and $\hat{p}_{ij}$ were computed previously and can be substituted into Equation 10.13b to yield $L_1 = 11\ \ln(0.688) + 5\ \ln(0.312) + 4\ \ln(0.286) + 10\ \ln(0.714) = -18.31$. Necessarily, $L_1 \geq L_0$ because the greater number of parameters in the more elaborate first-order Markov model provide more flexibility for a closer fit to the data at hand. The statistical significance of the difference in log-likelihoods can be assessed knowing that the null distribution of $\Lambda = 2\,(L_1 - L_0) = 4.96$ is χ^2, with $\nu = (s^{m(H_A)} s^{m(H_0)})\ (s-1)$ degrees of freedom. Since the time series being tested in binary, $s = 2$. The null hypothesis is that the time dependence is zero order, so $m(H_0) = 0$, and the alternative hypothesis is first-order serial dependence, or $m(H_A) = 1$. Thus $\nu = (2^1 - 2^0)(2-1) = 1$ degree of freedom. In general the appropriate degrees-of-freedom will be the difference in dimensionality (i.e., the parameter counts) between the competing models. This likelihood test result is consistent with the χ^2 goodness-of-fit test conducted in Example 10.1, where $\chi^2 = 4.82$, which is not surprising because the χ^2 test conducted there is an approximation to the likelihood ratio test. ◇

Both the AIC and BIC criteria attempt to find the most appropriate model order by striking a balance between goodness of fit, as reflected in the log-likelihoods, and a penalty that increases with the number of fitted parameters. The two approaches differ only in the form of the penalty function. The AIC and BIC statistics are computed for each trial order m, using

$$AIC(m) = -2L_m + 2s^m(s-1), \tag{10.14}$$

and

$$BIC(m) = -2L_m + s^m\ \ln(n), \tag{10.15}$$

respectively. The order m is chosen as appropriate that minimizes either Equation 10.14 or 10.15. The BIC criterion tends to be more conservative, generally picking lower orders than the AIC criterion when results of the two approaches differ. Use of the BIC statistic is generally preferable for sufficiently long time series, although "sufficiently long" may range from around $n = 100$ to over $n = 1000$, depending on the nature of the serial correlation (Katz, 1981).

10.3. TIME DOMAIN—II. CONTINUOUS DATA

10.3.1. First-Order Autoregression

The Markov chain models described in the previous section are not suitable for describing time series of data that are continuous, in the sense of the data being able to take on infinitely many values on all or part of the real line. As discussed in Chapter 4, atmospheric variables such as temperature, wind speed, geopotential height, and so on, are continuous variables in this sense. The correlation structure of such time series often can be represented successfully using a class of time series models known as *Box–Jenkins models*, after the classic text by Box and Jenkins (1976).

The *first-order autoregression*, or AR(1) model, is the simplest Box–Jenkins model. It is the continuous counterpart of the first-order Markov chain. As the name suggests, one way of viewing the AR(1) model is as a simple linear regression (Section 7.2.1), where the predictand is the value of the time series at time $t+1$, x_{t+1}, and the predictor is the current value of the time series, x_t. The AR(1) model can be written as

$$x_{t+1} - \mu = \phi(x_t - \mu) + \varepsilon_{t+1}, \tag{10.16}$$

where μ is the mean of the time series, ϕ is the autoregressive parameter, and ε_{t+1} is a random quantity corresponding to the residual in ordinary regression. The right-hand side of Equation 10.16 consists of a deterministic part in the first term and a random part in the second term. That is, the next value of the time series x_{t+1} is given by the function of x_t in the first term, plus the random shock or innovation ε_{t+1}.

The time series of x is assumed to be stationary, so that its mean μ is the same for each interval of time. The data series also exhibits a variance, σ_x^2, the sample counterpart of which is just the ordinary sample variance computed from the values of the time series by squaring Equation 3.6. The ε's are mutually independent random quantities having mean $\mu_\varepsilon = 0$ and variance σ_ε^2. Very often it is further assumed that the ε's follow a Gaussian distribution.

As illustrated in Figure 10.5, the autoregressive model in Equation 10.16 can represent the serial correlation of a time series. This is a scatterplot of minimum temperatures at Canandaigua, New York, during January 1987, from Table A.1. The first 30 data values, for 1–30 January, are plotted on the horizontal axis. The corresponding temperatures for the following days, 2–31 January, are plotted on the vertical axis. The serial correlation, or persistence, is evident from the appearance of the point cloud and from the positive slope of the regression line. Equation 10.16 can be viewed as a prediction equation for x_{t+1} using x_t as the predictor. Rearranging Equation 10.16 to more closely resemble the simple linear regression Equation 7.3 yields the intercept $a = \mu(1–\phi)$, and slope $b = \phi$.

Another way to look at Equation 10.16 is as an algorithm for generating synthetic time series of values of x, in the same sense as in Section 4.7. Beginning with an initial value, x_0, we would subtract the mean value (i.e., construct the corresponding anomaly), multiply by the autoregressive parameter ϕ, and then add a randomly generated variable ε_1 drawn from a Gaussian distribution (see Section 4.7.4) with mean zero and variance σ_ε^2. The first value of the time series, x_1, would then be produced by adding back the mean μ. The next time series value, x_2, would then be produced in a similar way, by operating on x_1 and adding a new random Gaussian quantity ε_2. For positive values of the parameter ϕ, synthetic time series constructed in this way will exhibit positive serial correlation because each newly generated data value x_{t+1} includes some information carried forward from the preceding value x_t. Since x_t was in turn generated in part from $x_{t–1}$, and so on, members of the time series separated by more than one time unit will be correlated, although this correlation becomes progressively weaker as the time separation increases.

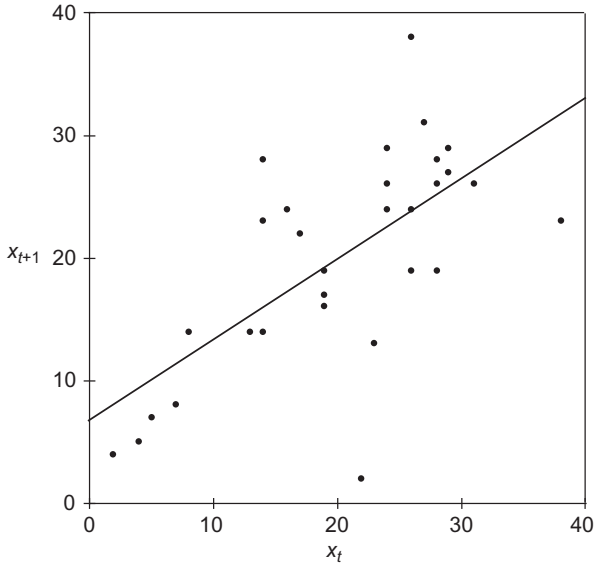

FIGURE 10.5 Scatterplot of January 1–30, 1987 minimum temperatures (°F) at Canandaigua, New York (x_t, horizontal) paired with minimum temperatures for the following days, January 2–31 (x_{t+1}, vertical). The data are from Table A.1. The regression line corresponding to the first term of the AR(1) time series model (Equation 10.16) is also shown.

The first-order autoregression is sometimes called the *Markov process* or Markov scheme. It shares with the first-order Markov chain the property that the full history of the time series prior to x_t provides no additional information regarding x_{t+1}, once x_t is known. This property can be expressed formally as

$$\Pr\{X_{t+1} \le x_{t+1} \,|\, X_t \le x_t, X_{t-1} \le x_{t-1}, \ldots, X_1 \le x_1\} = \Pr\{X_{t+1} \le x_{t+1} \,|\, X_t \le x_t\}. \tag{10.17}$$

Here the notation for continuous random variables has been used to express essentially the same idea as in Equation 10.1 for a sequence of discrete events. Again, Equation 10.17 does not imply that values of the time series separated by more than one time step are independent, but only that the influence of the prior history of the time series on its future values is fully contained in the current value x_t, regardless of the particular path by which the time series arrived at x_t.

Equation 10.16 is also sometimes known as a *red noise* process, because a positive value of the parameter ϕ averages or smooths out short-term fluctuations in the serially independent series of innovations, ε, while affecting the slower random variations much less strongly. The resulting time series is called red noise by analogy to visible light depleted in the shorter wavelengths, which appears reddish. This topic will be discussed more fully in Section 10.5, but the effect can be appreciated by looking at Figure 5.4. That figure compares a series of uncorrelated Gaussian values, ε_t (panel a), with an autocorrelated series generated from them using Equation 10.16 and the value $\phi = 0.6$ (panel b). It is evident that the most erratic point-to-point variations in the uncorrelated series have been smoothed out, but the slower random variations are essentially preserved. In the time domain this smoothing is expressed as positive serial correlation. From a frequency perspective, the resulting series is "reddened."

Parameter estimation for the first-order autoregressive model is straightforward. The estimated mean of the time series, μ, is simply the usual sample average (Equation 3.2) of the data set, provided that the series can be considered to be stationary. Nonstationary series must first be dealt with in one of the ways sketched in Section 10.1.1.

The estimated autoregressive parameter is simply equal to the sample lag-1 autocorrelation coefficient, Equation 3.36,

$$\hat{\phi} = r_1. \tag{10.18}$$

For the resulting probability model to be stationary, it is required that $-1 < \phi < 1$. As a practical matter this presents no problem for the first-order autoregression, because the correlation coefficient also is bounded by the same limits. For most atmospheric time series the parameter ϕ will be positive, reflecting persistence. Negative values of ϕ are possible, but correspond to very jagged (anticorrelated) time series with a tendency for alternating values above and below the mean. Because of the Markovian property, the full (theoretical, or population, or generating-process) autocorrelation function for a time series governed by a first-order autoregressive process can be written in terms of the autoregressive parameter as

$$\rho_k = \phi^k. \tag{10.19}$$

Equations 10.18 and 10.19 correspond directly to Equation 10.6 for the discrete first-order Markov chain. Thus the autocorrelation function for an AR(1) process with $\phi > 0$ decays exponentially from $\rho_0 = 1$, approaching zero as $k \rightarrow \infty$.

A series of truly independent data would have $\phi = 0$. However, a finite sample of independent data generally will exhibit a nonzero sample estimate for the autoregressive parameter. For a sufficiently long data series the sampling distribution of $\hat{\phi}$ is approximately Gaussian, with $\mu_\phi = \hat{\phi}$ and variance $\sigma_\phi^2 = (1-\phi^2)/n$. Therefore a test for the sample estimate of the autoregressive parameter, corresponding to Equation 5.3 with the null hypothesis that $\phi = 0$, can be carried out using the test statistic

$$z = \frac{\hat{\phi} - 0}{\left[\mathrm{Var}\left(\hat{\phi}\right)\right]^{1/2}} = \frac{\hat{\phi}}{[1/n]^{1/2}}, \tag{10.20}$$

because $\phi = 0$ under the null hypothesis. Statistical significance would be assessed approximately using standard Gaussian probabilities. This test is virtually identical to the t test for the slope of a regression line (Section 7.2.5).

The final parameter of the statistical model in Equation 10.16 is the residual variance, or innovation variance, σ_ε^2. This quantity is sometimes also known as the *white-noise variance*, for reasons that are explained in Section 10.5. This parameter expresses the variability or uncertainty in the time series not accounted for by the serial correlation or, put another way, the uncertainty in x_{t+1} given that x_t is known. The brute-force approach to estimating σ_ε^2 is to estimate ϕ using Equation 10.18, compute the time series ε_{t+1} from the data using a rearrangement of Equation 10.16, and then to compute the ordinary sample variance of these ε values. Since the variance of the data is often computed as a matter of course, another way to estimate the white-noise variance is to use the relationship between the variances of the data series and the innovation series in the AR(1) model,

$$\sigma_\varepsilon^2 = \left(1 - \phi^2\right) \sigma_x^2. \tag{10.21}$$

Equation 10.21 implies $\sigma_\varepsilon^2 \leq \sigma_x^2$, with equality only for independent data, for which $\phi = 0$. Equation 10.21 also implies that knowing the current value of an autocorrelated time series decreases uncertainty about the next value of the time series. In practical settings we work with sample estimates of the autoregressive parameter and of the variance of the data series, so that the corresponding sample estimator of the white-noise variance is

$$\hat{s}_\varepsilon^2 = \frac{1 - \hat{\phi}^2}{n-2} \sum_{t=1}^{n} (x_t - \bar{x})^2 = \frac{n-1}{n-2}\left(1 - \hat{\phi}^2\right) s_x^2. \tag{10.22}$$

The difference between Equations 10.22 and 10.21 is appreciable only if the data series is relatively short.

Example 10.3. A First-Order Autoregression

Consider fitting an AR(1) process to the series of January 1987 minimum temperatures from Canandaigua, in Table A.1. As indicated in the table, the average of these 31 values is 20.23°F, and this would be adopted as the estimated mean of the time series, assuming stationarity. The sample lag-1 autocorrelation coefficient, from Equation 3.29, is $r_1 = 0.67$, and this value would be adopted as the estimated autoregressive parameter according to Equation 10.18.

The scatterplot of these data against themselves lagged by one time unit in Figure 10.5 suggests the positive serial correlation typical of daily temperature data. A formal test of the estimated autoregressive parameter versus the null hypothesis that it is really zero would use the test statistic in Equation 10.20, $z = 0.67/[1/31]^{1/2} = 3.73$. This test provides strong evidence that the observed nonzero sample autocorrelation did not arise by chance from a sequence of 31 independent values.

The sample standard deviation of the 31 Canandaigua minimum temperatures in Table A1 is 8.81°F. Using Equation 10.22, the estimated white-noise variance for the fitted autoregression would be $s_\varepsilon^2 = (30/29)(1–0.67^2)(8.81^2) = 44.24°F^2$, corresponding to a standard deviation of 6.65°F. By comparison, the brute-force sample standard deviation of the series of sample residuals, each computed from the rearrangement of Equation 10.16 as $e_{t+1} = (x_{t+1}–\mu)–\phi(x–\mu)$ is 6.55°F.

The computations in this example have been conducted under the assumption that the time series being analyzed is stationary, which implies that the mean value does not change through time. This assumption is not exactly satisfied by these data, as illustrated in Figure 10.6. Here the time series of the January 1987 Canandaigua minimum temperature data is shown together with the climatological average temperatures for the period 1961–1990 (dashed line), and the linear trend fit to the 31 data points for 1987 (solid line).

Of course, the dashed line in Figure 10.6 is a better representation of the long-term (population) mean minimum temperatures at this location, and it indicates that early January is slightly warmer than late January on average. Strictly speaking, the data series is not stationary, since the underlying mean value for the time series is not constant through time. However, the change through the month represented by the dashed line is sufficiently minor (in comparison to the variability around this mean function) that

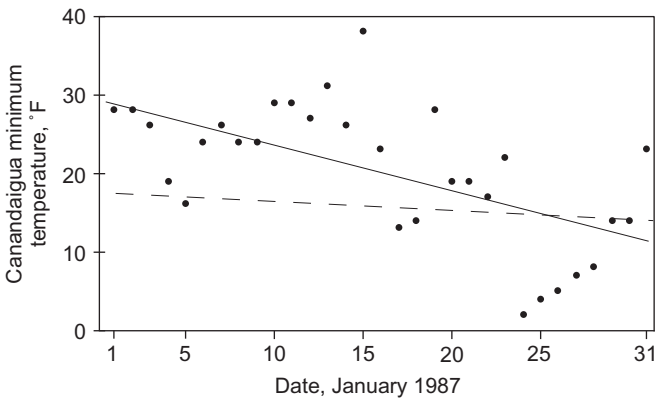

FIGURE 10.6 Time series of the January 1987 Canandaigua minimum temperature data. Solid line is the least-squares linear trend in the data, and the dashed line represents the climatological average minimum temperatures for the period 1961–1990.

generally we would be comfortable in pooling data from a collection of Januaries and assuming stationarity. In fact, the preceding results for the 1987 data sample are not very much different if the January 1987 mean minimum temperature of 20.23°F, or the long-term climatological temperatures represented by the dashed line, are assumed. In this latter case, we find $\phi = 0.64$, and $s_\varepsilon^2 = 6.49°\text{F}$.

Because the long-term climatological minimum temperature declines so slowly, it is clear that the rather steep negative slope of the solid line in Figure 10.6 results mainly from sampling variations in this short example data record. Normally an analysis of this kind would be carried out using a much longer data series. However, if no other information about the January minimum temperature climate of this location were available, it would be sensible to produce a stationary series before proceeding further, by subtracting the mean values represented by the solid line from the data points, provided the estimated slope is significantly different from zero (after accounting for the serial correlation in the data). The regression equation for this line is $\mu(t) = 29.6–0.584\ t$, where t is the date, and the slope is indeed nominally significant. Hypothetically, the autoregressive process in Equation 10.16 would then be fit using the time series of the anomalies $x'_t = x_t – \mu(t)$. For example, $x'_1 = 28°\text{F}–(29.6–0.584) = –1.02°$ F. Since the average residual from a least-squares regression line is zero (see Section 7.2.2), the mean of this series of anomalies x'_t will be zero. Fitting Equation 10.16 to this anomaly series yields $\hat{\phi} = 0.47$, and $s_\varepsilon^2 = 39.95°\text{F}^2$. ◇

10.3.2. Higher-Order Autoregressions

The first-order autoregression in Equation 10.16 generalizes readily to higher orders. That is, the regression equation predicting x_{t+1} can be expanded to include data values progressively further back in time as predictors. The general autoregressive model of order K, or AR(K) model is

$$x_{t+1} - \mu = \sum_{k=1}^{K} [\phi_k (x_{t-k+1} - \mu)] + \varepsilon_{t+1}. \tag{10.23}$$

Here the anomaly for the next time point, $x_{t+1} - \mu$, is a weighted sum of the previous K anomalies plus the random component ε_{t+1}, where the weights are the autoregressive coefficients ϕ_k. As before, the ε's are mutually independent, with zero mean and variance σ_ε^2. Stationarity of the process implies that μ and σ_ε^2 do not change through time. For $K = 1$, Equation 10.23 is identical to Equation 10.16.

Estimation of the K autoregressive parameters ϕ_k is perhaps most easily done using the set of equations relating them to the autocorrelation function, which are known as the *Yule–Walker equations*. These are

$$\begin{aligned}
r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 + \hat{\phi}_3 r_2 + \cdots + \hat{\phi}_K r_{K-1} \\
r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 + \hat{\phi}_3 r_1 + \cdots + \hat{\phi}_K r_{K-2} \\
r_3 &= \hat{\phi}_1 r_2 + \hat{\phi}_2 r_1 + \hat{\phi}_3 + \cdots + \hat{\phi}_K r_{K-3} \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \cdots \qquad\quad \vdots \\
r_K &= \hat{\phi}_1 r_{K-1} + \hat{\phi}_2 r_{K-2} + \hat{\phi}_3 r_{K-3} + \cdots + \hat{\phi}_K
\end{aligned} \tag{10.24}$$

Here $\phi_k = 0$ for $k > K$. The Yule–Walker equations arise from Equation 10.23, by multiplying by x_{t-k}, applying the expected value operator, and evaluating the result for different values of k (e.g., Box and

Jenkins, 1976). These equations can be solved simultaneously for the ϕ_k. Another approach to estimating the autoregressive parameters in Equation 10.23 is through maximum likelihood, assuming Gaussian-distributed random innovations ε_t (Miller, 1995). Alternatively, a method to use these equations recursively for parameter estimation—that is, to compute ϕ_1 and ϕ_2 to fit the AR(2) model knowing ϕ for the AR(1) model, and then to compute ϕ_1, ϕ_2, and ϕ_3 for the AR(3) model knowing ϕ_1 and ϕ_2 for the AR(2) model, and so on—is given in Box and Jenkins (1976) and Katz (1982). Constraints on the autoregressive parameters necessary for Equation 10.23 to describe a stationary process are given in Box and Jenkins (1976).

The theoretical autocorrelation function corresponding to a particular set of the ϕ_k's can be determined by solving Equation 10.24 for the first K autocorrelations and then recursively applying

$$\rho_m = \sum_{k=1}^{K} \phi_k \, \rho_{m-k}. \tag{10.25}$$

Equation 10.25 holds for lags $m \geq k$, with the understanding that $\rho_0 \equiv 1$. Finally, the generalization of Equation 10.21 for the relationship between the white-noise variance and the variance of the data values themselves is

$$\sigma_\varepsilon^2 = \left(1 - \sum_{k=1}^{K} \phi_k \rho_k\right) \sigma_x^2. \tag{10.26}$$

10.3.3. The AR(2) Model

The second-order autoregression, or AR(2) process is a common and important higher-order autoregressive model. It is reasonably simple, requiring the fitting of only two parameters in addition to the sample mean and variance of the series, yet it can describe a variety of qualitatively quite different behaviors of time series. The defining equation for AR(2) processes is

$$x_{t+1} - \mu = \phi_1 \, (x_t - \mu) + \phi_2 \, (x_{t-1} - \mu) + \varepsilon_{t+1}, \tag{10.27}$$

which is easily seen to be a special case of Equation 10.23. Using the first $K = 2$ of the Yule–Walker Equations (10.24),

$$r_1 = \hat{\phi}_1 + \hat{\phi}_2 r_1 \tag{10.28a}$$

and

$$r_2 = \hat{\phi}_1 r_1 + \hat{\phi}_2, \tag{10.28b}$$

the two autoregressive parameters can be estimated as

$$\hat{\phi}_1 = \frac{r_1 \, (1 - r_2)}{1 - r_1^2} \tag{10.29a}$$

and

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}, \tag{10.29b}$$

by solving Equations 10.28 for $\hat{\phi}_1$ and $\hat{\phi}_2$.

The white-noise variance for a fitted AR(2) model can be estimated in several ways. For very large samples, Equation 10.26 with $K = 2$ can be used with the sample variance of the time series, s_x^2. Alternatively, once the autoregressive parameters have been fit using Equations 10.29 or some other means, the corresponding estimated time series of the random innovations ε can be computed from a rearrangement of Equation 10.27 and their sample variance computed, as was done in Example 10.3 for the fitted AR(1) process. Another possibility is to use the recursive equation given by Katz (1982),

$$s_\varepsilon^2(m) = \left[1 - \hat{\phi}_m^2(m)\right] s_\varepsilon^2(m-1). \tag{10.30}$$

Here the autoregressive models AR(1), AR(2), … are fitted successively, $s_\varepsilon^2(m)$ is the estimated white-noise variance of the mth (i.e., current) autoregression, $s_\varepsilon^2(m-1)$ is the estimated white-noise variance for the previously fitted (one order smaller) model, and $\hat{\phi}_m(m)$ is the estimated autoregressive parameter for the highest lag in the current model. For the AR(2) model, Equation 10.30 can be used with the expression for $s_\varepsilon^2(1)$ in Equation 10.22 to yield

$$s_\varepsilon^2(2) = \left(1 - \hat{\phi}_2^2\right) \frac{n-1}{n-2} \left(1 - r_1^2\right) s_x^2, \tag{10.31}$$

since $\hat{\phi} = r_1$ for the AR(1) model.

For an AR(2) process to be stationary, its two parameters must satisfy the constraints

$$\left.\begin{array}{l} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 \end{array}\right\}, \tag{10.32}$$

which define the triangular region in the (ϕ_1, ϕ_2) plane shown in Figure 10.7. Note that substituting $\phi_2 = 0$ into Equation 10.32 yields the stationarity condition $-1 < \phi_1 < 1$ applicable to the AR(1) model. Figure 10.7 includes AR(1) models as special cases on the horizontal $\phi_2 = 0$ line, for which that stationarity condition applies.

The first two values of the theoretical autocorrelation function for a particular AR(2) process can be obtained by solving Equations 10.28 as

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \tag{10.33a}$$

and

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2}, \tag{10.33b}$$

and subsequent values of the autocorrelation function can be calculated using Equation 10.25. Figure 10.7 indicates that a wide range of types of autocorrelation functions, and thus a wide range of time correlation behaviors, can be represented by AR(2) processes. First, AR(2) models include the simpler AR(1) models as special cases. Two AR(1) autocorrelation functions are shown in Figure 10.7. The autocorrelation function for the model with $\phi_1 = 0.5$ and $\phi_2 = 0.0$ decays exponentially toward zero, following Equation 10.19. Autocorrelation functions for many atmospheric time series exhibit this kind of behavior, at least approximately. The other AR(1) model for which an autocorrelation function is shown is for $\phi_1 = -0.6$ and $\phi_2 = 0.0$. Because of the negative lag-one autocorrelation, the autocorrelation function exhibits oscillations around zero that are progressively damped at longer lags (again, compare Equation 10.19).

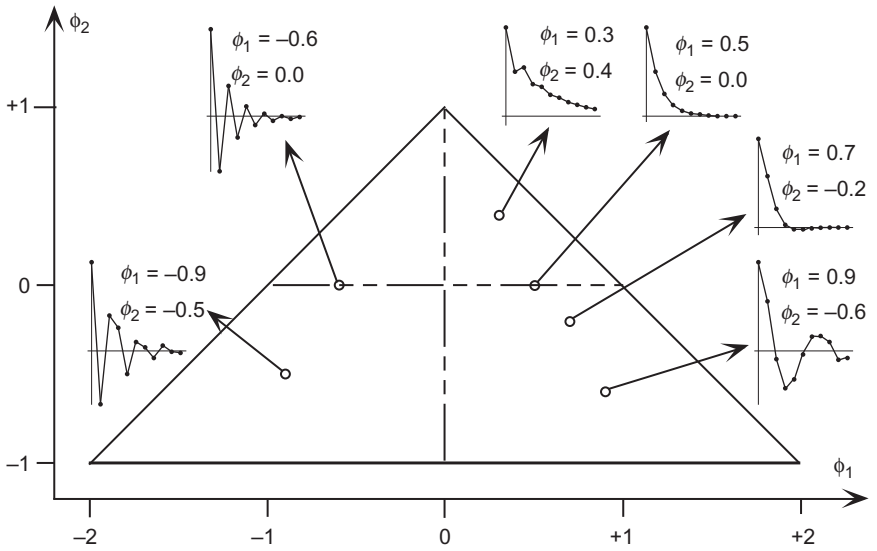

FIGURE 10.7 The allowable parameter space for stationary AR(2) processes, with insets showing autocorrelation functions for selected AR(2) models. The horizontal $\phi_2 = 0$ line locates the AR(1) models as special cases, and autocorrelation functions for two of these are shown. AR(2) models appropriate to atmospheric time series usually exhibit $\phi_1 > 0$.

That is, there is a tendency for the anomalies of consecutive data values to have opposite signs, so that data separated by even numbers of lags are positively correlated. This kind of behavior rarely is seen in atmospheric data series, and most AR(2) models for atmospheric data have $\phi_1 > 0$.

The second autoregressive parameter allows many other kinds of behaviors in the autocorrelation function. For example, the autocorrelation function for the AR(2) model with $\phi_1 = 0.3$ and $\phi_2 = 0.4$ exhibits a larger correlation at two lags than at one lag. For $\phi_1 = 0.7$ and $\phi_2 = -0.2$ the autocorrelation function declines very quickly and is almost zero for lags $k \geq 4$. The autocorrelation function for the AR(2) model with $\phi_1 = 0.9$ and $\phi_2 = -0.6$ is very interesting in that it exhibits a slow damped oscillation around zero. This characteristic reflects what are called *pseudoperiodicities* in the corresponding time series. That is, time series values separated by very few lags exhibit fairly strong positive correlation, those separated by a few more lags exhibit negative correlation, and values separated by a few more additional lags exhibit positive correlation again. The qualitative effect is for the corresponding time series to exhibit oscillations around the mean resembling an irregular cosine curve with an average period that is approximately equal to the number of lags at the first positive hump in the autocorrelation function. Thus AR(2) models can represent data that are approximately but not strictly periodic, such as barometric pressure variations resulting from the movement of midlatitude synoptic systems.

Some properties of autoregressive models are illustrated by the four example synthetic time series in Figure 10.8. Series (a) is simply a sequence of 50 independent Gaussian variates with $\mu = 0$. Series (b) is a realization of the AR(1) process generated using Equation 10.16 or, equivalently, Equation 10.27 with $\mu = 0$, $\phi_1 = 0.5$ and $\phi_2 = 0.0$. The apparent similarity between series (a) and (b) arises because series (a) has been used as the ε_{t+1} series forcing the autoregressive process in Equation 10.27. The effect of the parameter $\phi = \phi_1 > 0$ is to smooth out step-to-step variations in the white-noise series (a), and to give the resulting time series a bit of memory. The relationship of the series in these two panels is analogous to that in Figure 5.4, in which $\phi = 0.6$.

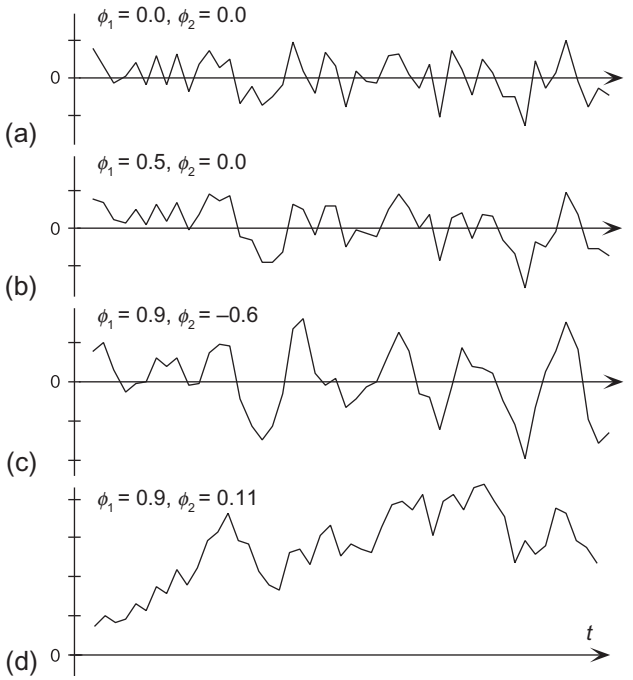

FIGURE 10.8 Four synthetic time series illustrating some properties of autoregressive models. Series (a) consists of independent Gaussian variables (white noise). Series (b) is a realization of the AR(1) process with $\phi_1 = 0.5$, and series (c) is a realization of the AR(2) process with $\phi_1 = 0.9$ and $\phi_2 = -0.6$, autocorrelation functions for both of which are shown in Figure 10.7. Series (d) is nonstationary because its parameters lie outside the triangle in Figure 10.7, and this nonstationarity can be seen as a drifting in the mean value. The series (b)–(d) were constructed using Equation 10.27 with $\mu = 0$ and the ε's from series (a).

Series (c) in Figure 10.8 is a realization of the AR(2) process with $\mu = 0$, $\phi_1 = 0.9$ and $\phi_2 = -0.6$. It resembles qualitatively some atmospheric series (e.g., midlatitude sea-level pressures), but has been generated using Equation 10.27 with series (a) as the forcing white noise. This series exhibits pseudoperiodicities. That is, peaks and troughs in this time series tend to recur with a period near six or seven time intervals, but these are not so regular that a cosine function or the sum of a few cosine functions would represent them very well. This feature is the expression in the data series of the positive hump in the autocorrelation function for this autoregressive model shown in the inset of Figure 10.7, which occurs at a lag interval of six or seven time periods. Similarly, the peak–trough pairs tend to be separated by perhaps three or four time intervals, corresponding to the minimum in the autocorrelation function at these lags shown in the inset in Figure 10.7.

The autoregressive parameters $\phi_1 = 0.9$ and $\phi_2 = 0.11$ for series (d) in Figure 10.8 fall outside the triangular region in Figure 10.7 that defines the limits of stationary AR(2) processes. The series is therefore not stationary, and the nonstationarity can be seen as a drifting of the mean value in the realization of this process shown in Figure 10.8.

Finally, series (a) through (c) in Figure 10.8 illustrate the nature of the relationship between the variance of the time series, σ_x^2, and the white-noise variance, σ_ε^2, of an autoregressive process. Series (a) consists simply of independent Gaussian variates, or white noise. Formally, it can be viewed as a special case of an autoregressive process, with all the ϕ_k's $= 0$. Using Equation 10.26 it is clear that

$\sigma_x^2 = \sigma_\varepsilon^2$ for this series. Since series (b) and (c) were generated using series (a) as the white-noise forcing ε_{t+1}, σ_ε^2 for all three of these series are equal. Time series (c) gives the visual impression of being more variable than series (b), which in turn appears to exhibit more variability than series (a). Using Equations 10.33 with Equation 10.26 it is easy to compute that σ_x^2 for series (b) is 1.33 times larger than the common σ_ε^2, and for series (c) it is 2.29 times larger. The equations on which these computations are based pertain only to stationary autoregressive series, and so cannot be applied meaningfully to the nonstationary series (d).

10.3.4. Order Selection Criteria

The Yule–Walker equations (10.24) can be used to fit autoregressive models to essentially arbitrarily high order. At some point, however, expanding the complexity of the model will not appreciably improve its representation of the data. Adding more terms in Equation 10.23 will eventually result in the model being overfit or excessively tuned to the data used for parameter estimation.

The BIC (Schwarz, 1978) and AIC (Akaike, 1974) statistics, applied to Markov chains in Section 10.2, are also often used to decide among potential orders of autoregressive models. Both statistics involve the log-likelihood plus a penalty for the number of parameters, with the two criteria differing only in the form of the penalty function. Here the likelihood function involves the estimated (assumed Gaussian) white-noise variance.

For each candidate order m, the order selection statistics

$$BIC(m) = n\ \ln\left[\frac{n}{n-m-1}s_\varepsilon^2(m)\right] + (m+1)\ \ln(n), \tag{10.34}$$

or

$$AIC(m) = n\ \ln\left[\frac{n}{n-m-1}s_\varepsilon^2(m)\right] + 2(m+1), \tag{10.35}$$

are computed, using $s_\varepsilon^2(m)$ from Equation 10.30. Better fitting models will exhibit smaller white-noise variance, implying less residual uncertainty. Arbitrarily adding more parameters (fitting higher- and higher-order autoregressive models) will not increase the white-noise variance estimated from the data sample, but neither will the estimated white-noise variance decrease much if the extra parameters are not effective in improving the description of the behavior of the data. Thus the penalty functions serve to guard against overfitting. That order m is chosen as appropriate, which minimizes either Equation 10.34 or 10.35.

Example 10.4. Order Selection among Autoregressive Models

Table 10.3 summarizes the results of fitting successively higher-order autoregressive models to the January 1987 Canandaigua minimum temperature data, assuming that they are stationary without removal of a trend. The second column shows the sample autocorrelation function up to seven lags. The estimated white-noise variance for autoregressions of orders one through seven, computed using the Yule–Walker equations and Equation 10.30, is shown in the third column. Notice that $s_\varepsilon^2(0)$ is simply the sample variance of the time series itself, or s_x^2. The estimated white-noise variances decrease progressively as more terms are added to Equation 10.23, but toward the bottom of the table adding yet more terms has little further effect.

The BIC and AIC statistics for each candidate autoregression are shown in the last two columns. Both indicate that the AR(1) model is most appropriate for this data, as $m = 1$ produces the minimum in both order selection statistics. Similar results also are obtained for the other three temperature series in

TABLE 10.3 Illustration of Order Selection for Autoregressive Models to Represent the January 1987 Canandaigua Minimum Temperature Series, Assuming Stationarity

Lag, m	r_m	$s_\varepsilon^2(m)$	BIC(m)	AIC(m)
0	1.000	77.58	138.32	136.89
1	0.672	42.55	125.20	122.34
2	0.507	42.11	129.41	125.11
3	0.397	42.04	133.91	128.18
4	0.432	39.72	136.76	129.59
5	0.198	34.39	136.94	128.34
6	0.183	33.03	140.39	130.35
7	0.161	33.02	145.14	133.66

Presented are the autocorrelation function for the first seven lags m, the estimated white-noise variance for each AR(m) model, and the BIC and AIC statistics for each trial order. For $m = 0$ the autocorrelation function is 1.00, and the white-noise variance is equal to the sample variance of the series. The AR(1) model is selected by both the BIC and AIC criteria.

Table A.1. Note, however, that with a larger sample size, higher-order autoregressions could be chosen by both criteria. For the estimated residual variances presented in Table 10.3, using the AIC statistic would lead to the choice of the AR(2) model for n greater than about 290, and the AR(2) model would minimize the BIC statistic for n larger than about 430. ◇

10.3.5. The Variance of a Time Average

Estimation of the sampling distribution of the average of a correlated time series is an important application of time series models in atmospheric data analysis. Recall that a sampling distribution characterizes the batch-to-batch variability of a statistic computed from a finite data sample. If the data values making up a sample average are independent, the variance of the sampling distribution of that average is given by the variance of the data, s_x^2, divided by the sample size (Equation 5.4).

When the underlying data are positively autocorrelated, using Equation 5.4 to calculate the variance of (the sampling distribution of) their time average leads to an underestimate. This discrepancy is a consequence of the tendency for nearby values of correlated time series to be similar, leading to less batch-to-batch consistency of the sample average. The phenomenon is illustrated in Figure 5.4. As discussed in Chapter 5, underestimating the variance of the sampling distribution of the mean can cause serious problems for statistical inference, leading for example, to unwarranted rejection of valid null hypotheses.

The effect of serial correlation on the variance of a time average over a sufficiently large sample can be accounted for through a variance inflation factor, V, modifying Equation 5.4:

$$\mathrm{Var}[\bar{x}] = \frac{V\sigma_x^2}{n}. \tag{10.36}$$

If the data series is uncorrelated, $V = 1$ and Equation 10.36 corresponds to Equation 5.4. If the data exhibit positive serial correlation, $V > 1$ and the variance of the time average is inflated above what would be implied for independent data. Note, however, that even if the underlying data are correlated,

the mean of the sampling distribution of the time average is the same as the underlying mean of the data being averaged,

$$E[\overline{x}] = \mu_{\overline{x}} = E[x_t] = \mu_x. \tag{10.37}$$

For arbitrarily large sample sizes, the variance inflation factor depends on the autocorrelation function according to

$$V = 1 + 2\sum_{k=1}^{\infty} \rho_k. \tag{10.38}$$

However, the variance inflation factor can be estimated with greater ease and precision if a data series is well represented by an autoregressive model. In terms of the parameters of an AR(K) model, the large-sample variance inflation factor in Equation 10.38 is well approximated by

$$V = \frac{1 - \sum_{k=1}^{K} \phi_k \rho_k}{\left[1 - \sum_{k=1}^{K} \phi_k\right]^2}. \tag{10.39}$$

Note that the theoretical autocorrelations ρ_k in Equation 10.39 can be expressed in terms of the autoregressive parameters by solving the Yule–Walker Equations (10.24) for the correlations. In the special case of an AR(1) model being appropriate for a time series of interest, Equation 10.39 reduces to

$$V = \frac{1 + \phi}{1 - \phi}, \tag{10.40}$$

which was used to estimate the effective sample size in Equation 5.12 and the variance of the sampling distribution of a sample mean in Equation 5.13. Equations 10.39 and 10.40 are convenient large-sample approximations to the formula for the variance inflation factor based on sample autocorrelation estimates

$$V = 1 + 2\sum_{k=1}^{n} \left(1 - \frac{k}{n}\right) r_k, \tag{10.41}$$

which more closely follows Equation 10.38. Equation 10.41 approaches Equations 10.39 and 10.40 for large sample size n, when the autocorrelations r_k are expressed in terms of the autoregressive parameters (Equation 10.24). Usually either Equation 10.39 or 10.40, as appropriate, would be used to compute the variance inflation factor, although the Diebold–Mariano test (Diebold and Mariano, 1995) uses Equation 10.41 (see Section 5.2.4).

Example 10.5. Variances of Time Averages of Different Lengths

The relationship between the variance of a time average and the variance of the individual elements of a time series in Equation 10.36 can be useful in surprising ways. Consider, for example, the average winter (December–February) geopotential heights, and the standard deviations of those averages, for the northern hemisphere shown in Figures 10.9a and 10.9b, respectively. Figure 10.9a shows the average

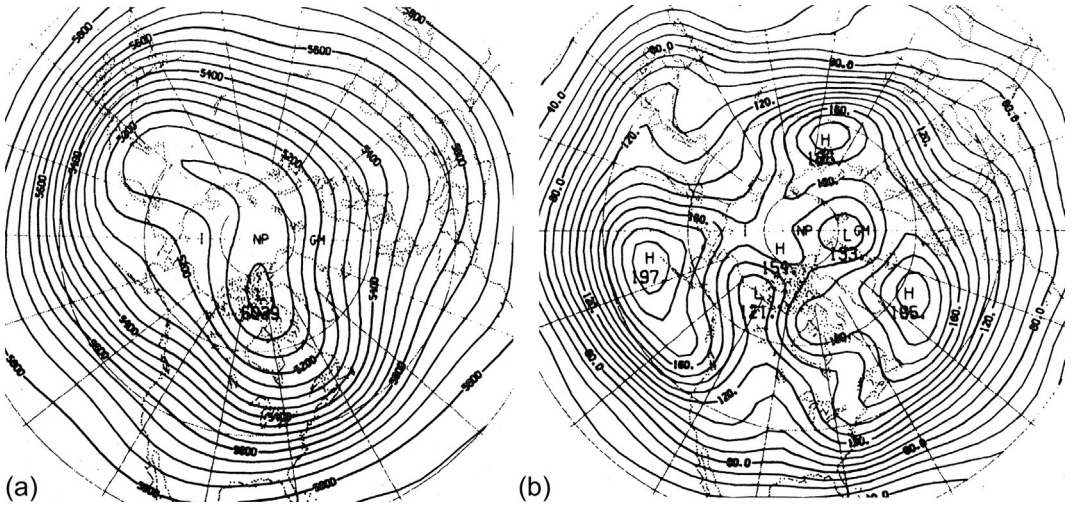

FIGURE 10.9 Average 500 mb height field for the northern hemisphere winter (a), and the field of standard deviations of that average, reflecting winter-to-winter variations (b). *From Blackmon (1976). © American Meteorological Society. Used with permission.*

field (Equation 10.37), and Figure 10.9b shows the standard deviation of 90-day averages of winter 500 mb heights, representing the interannual variability. That is, Figure 10.9b shows the square root of Equation 10.36, with s_x^2 being the variance of the daily 500 mb height measurements and $n = 90$. The maps in Figure 10.9 summarize statistics that have been computed for a large number of gridpoints in the hemispheric domain.

Suppose, however, that the sampling distribution of 500 mb heights averaged over a different length of time were needed. We might be interested in the variance of 10-day averages of 500 mb heights at selected locations, for use as a climatological reference when calculating the skill of forecasts of 10-day averages of this quantity, using Equation 9.37. (The variance of the climatological distribution is exactly the mean-squared error of the climatological reference forecast.) Assuming that time series of winter 500 mb heights are stationary, the variance of an average over some different time period can be approximated without explicitly knowing the variance inflation factor in either Equations 10.38 or 10.39, and therefore without necessarily having the daily data. The ratio of the variances of 10-day and 90-day averages can be constructed from Equation 10.36,

$$\frac{\mathrm{Var}[\bar{x}_{10}]}{\mathrm{Var}[\bar{x}_{90}]} = \frac{V\, s_x^2/10}{V\, s_x^2/90}, \tag{10.42a}$$

leading to

$$\mathrm{Var}[\bar{x}_{10}] = \frac{90}{10}\mathrm{Var}[\bar{x}_{90}]. \tag{10.42b}$$

Regardless of the averaging period, the variance inflation factor V and the variance of the daily observations s_x^2 are the same because they are characteristics of the underlying daily time series. Thus the variance of a 10-day average is approximately nine times larger than the variance of a 90-day average, and a map of hemispheric 10-day standard deviations of winter 500 mb heights would be qualitatively very similar to Figure 10.9b, but exhibiting magnitudes about $\sqrt{9} = 3$ times larger. ◇

10.3.6. Autoregressive-Moving Average Models

Autoregressions actually constitute a subset of a broader class of the Box–Jenkins time-domain models, also known as *autoregressive-moving average*, or *ARMA, models*. The general ARMA(K, M) model has K autoregressive terms, as in the AR(K) process in Equation 10.23, and in addition contains M moving average terms that compose a weighted average of the M previous values of the ε's. The ARMA(K, M) model thus contains K autoregressive parameters ϕ_k and M moving average parameters θ_m that affect the time series according to

$$x_{t+1} - \mu = \sum_{k=1}^{K} \phi_k \, (x_{t-k+1} - \mu) + \varepsilon_{t+1} - \sum_{m=1}^{M} \theta_m \, \varepsilon_{t-m+1}. \tag{10.43}$$

The AR(K) process in Equation 10.23 is a special case of the ARMA(K, M) model in Equation 10.43, with all the $\theta_m = 0$. Similarly, a pure *moving average process* of order M, or MA(M) process, would be a special case of Equation 10.43, with all the $\phi_k = 0$. Parameter estimation and derivation of the autocorrelation function for the general ARMA(K, M) process is more difficult than for the simpler AR(K) models. Parameter estimation methods are given in Box and Jenkins (1976), and many time-series computer packages will fit ARMA models.

The ARMA(1,1) is an important and common ARMA model,

$$x_{t+1} - \mu = \phi_1 \, (x_t - \mu) + \varepsilon_{t+1} - \theta_1 \, \varepsilon_t. \tag{10.44}$$

Computing parameter estimates even for this simple ARMA model is somewhat complicated, although Box and Jenkins (1976) present an easy graphical technique that allows estimation of ϕ_1 and θ_1 using the first two sample lagged correlations r_1 and r_2. The autocorrelation function for the ARMA (1,1) process can be calculated from the parameters using

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \tag{10.45a}$$

and

$$\rho_k = \phi_1 \, \rho_{k-1}, \quad k > 1. \tag{10.45b}$$

The autocorrelation function of an ARMA(1, 1) process decays exponentially from its value at ρ_1, which depends on both ϕ_1 and θ_1. This differs from the autocorrelation function for an AR(1) process, which decays exponentially from $\rho_0 = 1$. The relationship between the time-series variance and the white-noise variance of an ARMA(1,1) process is

$$\sigma_\varepsilon^2 = \frac{1 - \phi_1^2}{1 + \theta_1^2 + 2\phi_1\theta_1} \sigma_x^2. \tag{10.46}$$

Equations 10.45 and 10.46 also apply to the simpler AR(1) and MA(1) processes, for which $\theta_1 = 0$ or $\phi_1 = 0$, respectively.

10.3.7. Simulation and Forecasting with Continuous Time-Domain Models

An important application of time-domain models is in the simulation of synthetic (i.e., random number, as in Section 4.7) series having statistical characteristics that are similar to observed data series. Such

Monte Carlo simulations are useful for investigating impacts of atmospheric variability in situations where the record length of the observed data is known or suspected to be insufficient to include representative sequences of the relevant variable(s). Here it is necessary to choose the type and order of time-series model carefully, so that the simulated time series will well represent the variability of the real generating process.

Once an appropriate time series model has been identified and its parameters estimated, its defining equation can be used as an algorithm to generate synthetic time series. For example, if an AR(2) model is representative of the data, Equation 10.27 would be used, whereas Equation 10.44 would be used as the generation algorithm for ARMA(1,1) models. The simulation method is similar to that described earlier for sequences of binary variables generated using the Markov chain model. Here, however, the noise or innovation series, ε_{t+1}, usually is assumed to consist of independent Gaussian variates with $\mu_\varepsilon = 0$ and variance σ_ε^2, which is estimated from the data as described earlier.

At each time step, a new Gaussian ε_{t+1} is chosen (see Section 4.7.4) and substituted into the defining equation. The next value of the synthetic time series x_{t+1} is then computed using the previous K values of x (for AR models), the previous M values of ε (for MA models), or both (for ARMA models). The only real difficulty in implementing the process is at the beginning of each synthetic series, where there are no prior values of x and/or ε that can be used. A simple solution to this problem is to substitute the corresponding averages (expected values) for the unknown previous values. That is, $(x_t - \mu) = 0$ and $\varepsilon_t = 0$ for $t \leq 0$ might be assumed.

A better procedure is to generate the first values in a way that is consistent with the structure of the time-series model. For example, with an AR(1) model we could choose x_0 from a Gaussian distribution with variance $\sigma_x^2 = \sigma_\varepsilon^2/(1-\phi^2)$ (cf. Equation 10.21). Another very workable solution is to begin with $(x_t - \mu) = 0$ and $\varepsilon_t = 0$, but generate a longer time series than needed. The first few members of the resulting time series, which are most influenced by the initial values, are then discarded.

Example 10.6. Statistical Simulation with an Autoregressive Model

The time series in Figures 10.8b–d were produced according to the procedure just described, using the independent Gaussian series in Figure 10.8a as the series of ε's. The first and last few values of this independent series, and of the two series plotted in Figures 10.8b and c, are presented in Table 10.4. For all three series, $\mu = 0$ and $\sigma_\varepsilon^2 = 1$. Equation 10.16 has been used to generate the values of the AR(1) series, with $\phi_1 = 0.5$, and Equation 10.27 was used to generate the AR(2) series, with $\phi_1 = 0.9$ and $\phi_2 = -0.6$.

Consider the more difficult case of generating the AR(2) series. Calculating x_1 and x_2 in order to begin the series presents an immediate problem, because x_0 and x_{-1} do not exist. This simulation was initialized by assuming the expected values $\mathrm{E}[x_0] = \mathrm{E}[x_{-1}] = \mu = 0$. Thus since $\mu = 0$, $x_1 = \phi_1 x_0 + \phi_2 x_{-1} + \varepsilon_1 = (0.9)(0) - (0.6)(0) + 1.526 = 1.562$. Having generated x_1 in this way, it is then used to obtain $x_2 = \phi_1 x_1 + \phi_2 x_0 + \varepsilon_2 = (0.9)(1.526) - (0.6)(0) + 0.623 = 1.996$. For values of the AR(2) series at times $t = 3$ and larger, the computation is a straightforward application of Equation 10.27. For example, $x_3 = \phi_1 x_2 + \phi_2 x_1 + \varepsilon_3 = (0.9)(1.996) - (0.6)(1.526) - 0.272 = 0.609$. Similarly, $x_4 = \phi_1 x_3 + \phi_2 x_2 + \varepsilon_4 = (0.9)(0.609) - (0.6)(1.996) + 0.092 = -0.558$. If this synthetic series were to be used as part of a larger simulation, the first portion would generally be discarded, so that the retained values would have negligible memory of the initial condition $x_{-1} = x_0 = 0$. ◇

Purely statistical forecasts of the future evolution of time series can be produced using time-domain models. These are accomplished by simply extrapolating the most recently observed value(s) into the future using the appropriate defining equation, on the basis of parameter estimates fitted from the

TABLE 10.4 Values of the Time Series Plotted in Figure 10.8a–c

t	Independent Gaussian Series, ε_t, (Figure 10.8a)	AR(1) Series, x_t, (Figure 10.8b)	AR(2) Series, x_t, (Figure 10.8c)
1	1.526	1.526	1.526
2	0.623	1.387	1.996
3	−0.272	0.421	0.609
4	0.092	0.302	−0.558
5	0.823	0.974	−0.045
⋮	⋮	⋮	⋮
49	−0.505	−1.073	−3.172
50	−0.927	−1.463	−2.648

The AR(1) and AR(2) series have been generated from the independent Gaussian series using Equations 10.16 and 10.27, respectively, as the algorithms.

previous history of the series. Since the future values of the ε's cannot be known, the extrapolations are usually made using their expected values, that is, $E[\varepsilon] = 0$. Probability bounds on these extrapolations can be calculated as well.

The nature of this kind of forecast is most easily appreciated for the AR(1) model, the defining equation for which is Equation 10.16. Assume that the mean μ and the autoregressive parameter ϕ have been estimated from past data, the most recent of which is x_t. A nonprobabilistic forecast for x_{t+1} could be made by setting the unknown future ε_{t+1} to zero, and rearranging Equation 10.16 to yield $x_{t+1} = \mu + \phi(x_t - \mu)$. Note that, in common with the forecasting of a binary time series using a Markov chain model, this forecast is a compromise between persistence ($x_{t+1} = x_t$, which would result if $\phi = 1$) and climatology ($x_{t+1} = \mu$, which would result if $\phi = 0$). Further projections into the future would be obtained by extrapolating the previously forecast values, for example, $x_{t+2} = \mu + \phi(x_{t+1} - \mu)$, and $x_{t+3} = \mu + \phi(x_{t+2} - \mu)$. For the AR(1) model and $\phi > 0$, this series of forecasts would exponentially approach $x_\infty = \mu$.

The same procedure is used for higher-order autoregressions, except that the most recent K values of the time series are needed to extrapolate an AR(K) process (Equation 10.23). Forecasts derived from an AR(2) model, for example, would be made using the previous two observations of the time series, or $x_{t+1} = \mu + \phi_1(x_t - \mu) + \phi_2(x_{t-1} - \mu)$. Forecasts using ARMA models are only slightly more difficult, requiring that the last M values of the ε series be backcalculated before the projections begin.

Forecasts made using time-series models are of course uncertain, and the forecast uncertainty increases for longer lead times into the future. This uncertainty also depends on the nature of the appropriate time-series model (e.g., the order of an autoregression and its parameter values) and on the intrinsic uncertainty in the random noise series that is quantified by the white-noise variance σ_ε^2. The variance of a forecast made only one time step into the future is simply equal to the white-noise variance. Assuming the ε's follow a Gaussian distribution, a 95% probability interval on a forecast of x_{t+1} is then approximately $x_{t+1} \pm 2\sigma_\varepsilon$. For very long extrapolations, the variance of the forecasts approaches the variance of the time series itself, σ_x^2, which for AR models can be computed from the white-noise variance using Equation 10.26.

For intermediate lead times, calculation of forecast uncertainty is more complicated. For a forecast j time units into the future, the variance of the forecast is given by

$$\sigma^2\left(x_{t+j}\right) = \sigma_\varepsilon^2\left[1 + \sum_{i=1}^{j-1}\psi_i^2\right]. \tag{10.47}$$

Here the weights ψ_i depend on the parameters of the time series model, so that Equation 10.47 indicates that the variance of the forecast increases with both the white-noise variance and the lead time, and that the increase in uncertainty at increasing lead times depends on the specific nature of the time-series model. For the $j = 1$ time step forecast, there are no terms in the summation in Equation 10.47, and the forecast variance is equal to the white-noise variance, as noted earlier.

For AR(1) models, the ψ weights are simply

$$\psi_i = \phi^i, \quad i > 0, \tag{10.48}$$

so that, for example, $\psi_1 = \phi$, $\psi_2 = \phi^2$, and so on. More generally, for AR(K) models, the ψ weights are computed recursively, using

$$\psi_i = \sum_{k=1}^{K}\phi_k\psi_{i-k}, \tag{10.49}$$

where it is understood that $\psi_0 = 1$ and $\psi_i = 0$ for $i < 0$. For AR(2) models, for example, $\psi_1 = \phi_1$, $\psi_2 = \phi_1^2 + \phi_2$, $\psi_3 = \phi_1\,(\phi_1^2 + \phi_2) + \phi_2\,\phi_1$, and so on. Equations that can be used to compute the ψ weights for MA and ARMA models are given in Box and Jenkins (1976).

Example 10.7. Forecasting with an Autoregressive Model

Figure 10.10 illustrates forecasts using the AR(2) model with $\phi_1 = 0.9$ and $\phi_2 = -0.6$. The first six points in the time series, shown by the circles connected by heavy lines, are the same as the final six points in the time series shown in Figure 10.8c. The extrapolation of this series into the future, using Equation 10.27 with all $\varepsilon_{t+1} = 0$, is shown by the continuation of the heavy line connecting the x's. Note that only the

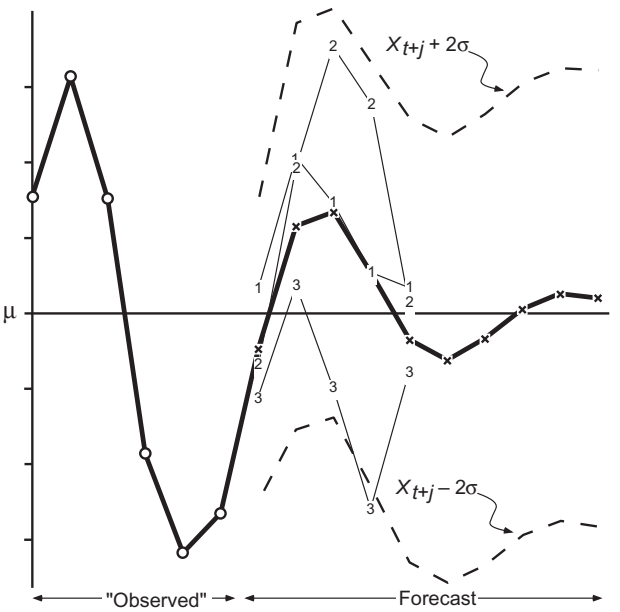

FIGURE 10.10 The final six points of the AR(2) time series in Figure 10.8c (heavy line, with circles), and its forecast evolution (heavy line, with x's) extrapolated using Equation 10.27 with all the $\varepsilon = 0$. The $\pm 2\sigma$ limits describing the uncertainty of the forecasts are shown with dashed lines. These standard deviations depend on the forecast lead time. For the 1-step ahead forecast, the width of the confidence interval is a function simply of the white-noise variance, $\pm 2\sigma_\varepsilon$. For very long lead times, the forecast series converges to the mean, μ, of the process, and the width of the confidence interval increases to $\pm 2\sigma_x$. Three example realizations of the first five points of the future evolution of the time series, simulated using Equation 10.27 and particular random ε values, are also shown (thin lines connecting numbered points).

final two observed values are used to extrapolate the series. The forecast series continues to show the pseudoperiodic behavior characteristic of this AR(2) model, but its oscillations damp out at longer lead times as the forecast series approaches the mean μ.

The approximate 95% confidence intervals for the forecast time series values, given by $\pm 2\sigma(x_{t+j})$ as computed from Equation 10.47 are shown by the dashed lines in Figure 10.10. For the particular values of the autoregressive parameters $\phi_1 = 0.9$ and $\phi_2 = -0.6$, Equation 10.49 yields $\psi_1 = 0.90$, $\psi_2 = 0.21$, $\psi_3 = -0.35$, $\psi_4 = -0.44$, and so on. Note that the confidence band follows the oscillations of the forecast series, and broadens from $\pm 2\sigma_\varepsilon$ at a lead of one time unit to nearly $\pm 2\sigma_x$ at the longer lead times.

Finally, Figure 10.10 shows the relationship between the forecast time series mean, and the first five points of three realizations of this AR(2) process, shown by the thin lines connecting points labeled "1," "2," and "3." (In effect these represent a 3-member ensemble forecast for the future evolution of this time series.) Each of these three series was computed using Equation 10.27, starting from $x_t = -2.648$ and $x_{t-1} = -3.172$, but using different sequences of independent Gaussian ε's. For the first two or three projections these remain reasonably close to the (mean) forecasts. Subsequently the three series begin to diverge as the influence of the final two points from Figure 10.8c diminishes and the accumulated influence of the new (and different) random ε's increases. For clarity these series have not been plotted more than five time units into the future, although doing so would have shown each to oscillate irregularly, with progressively less relationship to the forecast mean series. ◇

10.4. FREQUENCY DOMAIN—I. HARMONIC ANALYSIS

Analysis in the frequency domain involves representing data series in terms of contributions made at different timescales. For example, a time series of hourly temperature data from a midlatitude location usually will exhibit strong variations both at the daily timescale (corresponding to the diurnal cycle of solar heating) and at the annual timescale (reflecting the march of the seasons). In the time domain, these cycles would appear as large positive values in the autocorrelation function for lags at and near 24 hours for the diurnal cycle, and 24 x 365 = 8760 hours for the annual cycle. Thinking about the same time series in the frequency domain, we speak of large contributions to the total variability of the time series at periods of 24 and 8760 h, or at frequencies of $1/24 = 0.0417\ \text{h}^{-1}$ and $1/8760 = 0.000114\ \text{h}^{-1}$.

Harmonic analysis consists of representing the fluctuations or variations in a time series as having arisen from the adding together of a series of sine and cosine functions. These trigonometric functions are "harmonic" in the sense that they are chosen to have frequencies exhibiting integer multiples of the fundamental frequency determined by the sample size (i.e., length) of the data series. A common physical analogy is the musical sound produced by a vibrating string, where the pitch is determined by the fundamental frequency, but the esthetic quality of the sound depends also on the relative contributions of the higher harmonics.

10.4.1. Cosine and Sine Functions

It is worthwhile to review briefly the nature of the cosine function $\cos(\alpha)$, and the sine function $\sin(\alpha)$. The argument in both is a quantity α, measured in angular units, which can be either degrees or radians. Figure 10.11 shows portions of the cosine (solid) and sine (dashed) functions, on the angular interval 0 to $5\pi/2$ radians (0° to 450°).

The cosine and sine functions extend through indefinitely large negative and positive angles. The same wave pattern repeats every 2π radians or 360°, so that

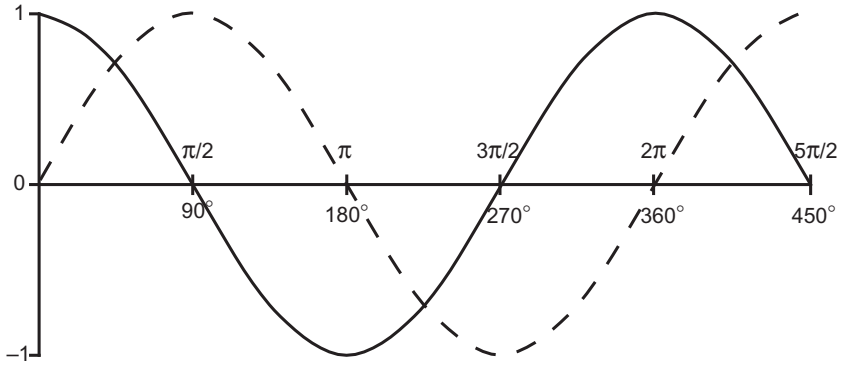

FIGURE 10.11 Portions of the cosine (solid) and sine (dashed) functions on the interval 0° to 450° or, equivalently, 0 to $5\pi/2$ radians. Each executes a full cycle every 360°, or 2π radians, and extends left to $-\infty$ and right to $+\infty$.

$$\cos(2\pi k + \alpha) = \cos(\alpha), \tag{10.50}$$

where k is any integer. An analogous equation holds for the sine function. That is, both the cosine and sine functions are periodic. Both functions oscillate around their average value of zero and attain maximum values of +1 and minimum values of –1. The cosine function is maximized at 0°, 360°, and so on, and the sine function is maximized at 90°, 450°, and so on.

These two functions have exactly the same shape but are offset from each other by 90°. Sliding the cosine function to the right by 90° produces the sine function, and sliding the sine function to the left by 90° produces the cosine function. That is,

$$\cos\left(\alpha - \frac{\pi}{2}\right) = \sin(\alpha) \tag{10.51a}$$

and

$$\sin\left(\alpha + \frac{\pi}{2}\right) = \cos(\alpha). \tag{10.51b}$$

10.4.2. Representing a Simple Time Series with a Harmonic Function

Even in the simple situation of time series having a sinusoidal character and executing a single cycle over the course of n observations, three small difficulties must be overcome in order to use a sine or cosine function to represent it. These are:

(1) The argument of a trigonometric function is an angle, whereas the data series is a function of time.
(2) Cosine and sine functions fluctuate between +1 and –1, but the data will generally fluctuate between different limits.
(3) The cosine function is at its maximum value for $\alpha = 0$ and $\alpha = 2\pi$, and the sine function is at its mean value for $\alpha = 0$ and $\alpha = 2\pi$. Both the sine and cosine may thus be positioned arbitrarily in the horizontal with respect to the data.

The solution to the first problem comes through regarding the length of the data record, n, as constituting a full cycle, or the fundamental period. Since the full cycle corresponds to 360° or 2π radians in angular measure, it is easy to proportionally rescale time to angular measure, using

$$\alpha = \left(\frac{360^\circ}{\text{cycle}}\right)\left(\frac{t\,\text{time units}}{n\,\text{time units/cycle}}\right) = \frac{t}{n}360^\circ \tag{10.52a}$$

or

$$\alpha = \left(\frac{2\pi}{\text{cycle}}\right)\left(\frac{t\,\text{time units}}{n\,\text{time units/cycle}}\right) = 2\pi\frac{t}{n}. \tag{10.52b}$$

These equations can be viewed as specifying the angle that subtends proportionally the same part of the angular distance between 0 and 2π, as the point t is located in time between 0 and n. The quantity

$$\omega_1 = \frac{2\pi}{n} \tag{10.53}$$

is called the *fundamental frequency*. This quantity is an angular frequency, having physical dimensions of radians per unit time. The fundamental frequency specifies the fraction of the full cycle, spanning n time units, that is executed during a single time unit. The subscript "1" indicates that ω_1 pertains to the wave that executes one full cycle over the whole data series.

The second problem is overcome by shifting a cosine or sine function up or down to the general level of the data, and then stretching or compressing it vertically until its range corresponds to that of the data. Since the mean of a pure cosine or sine wave is zero, simply adding the mean value of the data series to the cosine function assures that it will fluctuate around that mean value. The stretching or shrinking is accomplished by multiplying the cosine function by a constant, C_1, known as the *amplitude*. Again, the subscript indicates that this is the amplitude of the fundamental harmonic. Since the maximum and minimum values of a cosine function are ± 1, the maximum and minimum values of the function $C_1 \cos(\alpha)$ will be $\pm C_1$. Combining the solutions to these first two problems for a data series (call it y) yields

$$y_t = \bar{y} + C_1 \cos\left(\frac{2\pi t}{n}\right). \tag{10.54}$$

This function is plotted as the lighter curve in Figure 10.12. In this figure the horizontal axis indicates the equivalence of angular and time measure, through Equation 10.52, and the vertical shifting and stretching has produced a function fluctuating around the mean, with a range of $\pm C_1$.

Finally, it is usually necessary to shift a harmonic function laterally in order to have it match the peaks and troughs of a data series. This time shifting is most conveniently accomplished when the cosine function is used, because its maximum value is achieved when the angle on which it operates is zero. Shifting the cosine function to the right by the angle ϕ_1 results in a new function that is maximized at ϕ_1,

$$y_t = \bar{y} + C_1 \cos\left(\frac{2\pi t}{n} - \phi_1\right). \tag{10.55}$$

The angle ϕ_1 is called the *phase angle* or *phase shift*. Shifting the cosine function to the right by this amount requires subtracting ϕ_1, so that the argument of the cosine function is zero when $(2\pi\, t/n) = \phi_1$. Notice that by using Equation 10.51 it would be possible to rewrite Equation 10.55 using the sine function. However, the cosine usually is used as in Equation 10.55, because the phase angle can then be easily interpreted as corresponding to the time of the maximum of the harmonic function. That is, the function in Equation 10.55 is maximized at time $t = \phi_1 n/2\pi$.

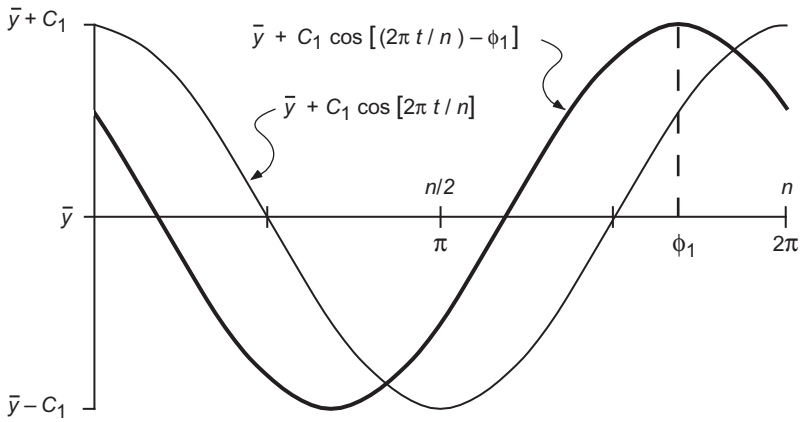

FIGURE 10.12 Transformation of a simple cosine function defined on 0 to 2π radians to a function representing a data series on the interval 0 to n time units. After changing from time to angular units, multiplying the cosine function by the amplitude C_1 stretches it so that it fluctuates through a range of $2C_1$. Adding the mean of the time series then shifts it to the proper vertical level, producing the lighter curve. The function can then be shifted laterally by subtracting the phase angle ϕ_1 that corresponds to the time of the maximum in the function (heavier curve).

Example 10.8. Transforming a Cosine Wave to Represent an Annual Cycle

Figure 10.13 illustrates the foregoing procedure using the 12 mean monthly temperatures (°F) for 1943–1989 at Ithaca, New York. Figure 10.13a is simply a plot of the 12 data points, with $t = 1$ indicating January, $t = 2$ indicating February, and so on. The overall annual average temperature of 46.1°F is located by the dashed horizontal line. These data appear to be at least approximately sinusoidal, executing a single full cycle over the course of the 12 months. The warmest mean temperature is 68.8°F in July and the coldest is 22.2°F in January.

The light curve at the bottom of Figure 10.13b is simply a cosine function with the argument transformed so that it executes one full cycle in 12 months. It is obviously a poor representation of the data. The dashed curve in Figure 10.13b shows this function lifted to the level of the average annual temperature, and stretched so that its range is similar to that of the data series (Equation 10.54). The stretching has been done only approximately, by choosing the amplitude C_1 to be half the difference between the July and January temperatures.

Finally, the cosine curve needs to be shifted to the right to line up well with the data. The maximum in the curve can be made to occur at $t = 7$ months (July) by introducing the phase shift, using Equation 10.52, of $\phi_1 = (7)(2\pi)/12 = 7\pi/6$. The result is the heavy curve in Figure 10.13b, which is of the form of Equation 10.55. This function lines up with the data points, albeit somewhat roughly. The correspondence between the curve and the data can be improved by using better estimators for the amplitude and phase of the cosine wave, as explained in the next section. ◇

10.4.3. Estimation of the Amplitude and Phase of a Single Harmonic

The heavy curve in Figure 10.13 represents the associated temperature data reasonably well, but the correspondence will be improved with more refined choices for C_1 and ϕ_1. The easiest way to do this is to use the trigonometric identity

$$\cos(\alpha - \phi_1) = \cos(\phi_1)\cos(\alpha) + \sin(\phi_1)\sin(\alpha). \tag{10.56}$$

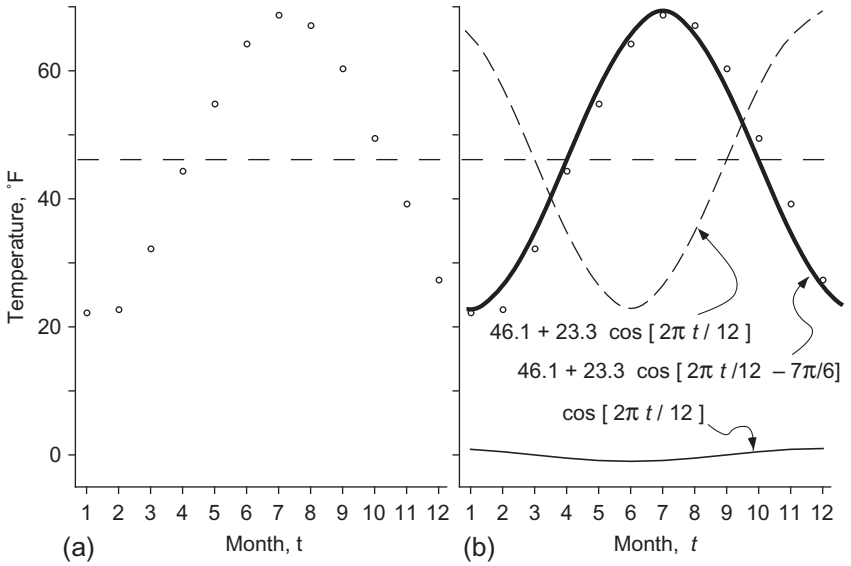

FIGURE 10.13 Illustration of the approximate matching of a cosine function to a data series. (a) Average monthly temperatures (° F) for Ithaca, New York for the years 1943–1989 (the data values are given in Table 10.5). The annual cycle of average temperature is evidently approximately sinusoidal. (b) Three cosine functions illustrating transformation from time to angular measure (light curve at bottom), vertical positioning and stretching (dashed curve), and lateral shifting (heavy curve) yielding finally the function matching the data approximately. The horizontal dashed lines indicate the average of the 12 data points, 46.1°F.

Substituting $\alpha = 2\pi\ t/n$ from Equation 10.52 and multiplying both sides by the amplitude C_1 yields

$$\begin{aligned} C_1 \cos\left(\frac{2\pi t}{n} - \phi_1\right) &= C_1 \cos(\phi_1) \cos\left(\frac{2\pi t}{n}\right) + C_1 \sin(\phi_1) \sin\left(\frac{2\pi t}{n}\right) \\ &= A_1 \cos\left(\frac{2\pi t}{n}\right) + B_1 \sin\left(\frac{2\pi t}{n}\right), \end{aligned} \tag{10.57}$$

where

$$A_1 = C_1 \cos(\phi_1). \tag{10.58a}$$

and

$$B_1 = C_1 \sin(\phi_1). \tag{10.58b}$$

Equation 10.57 says that it is mathematically equivalent to represent a harmonic wave either as a cosine function with amplitude C_1 and phase ϕ_1, or as the sum of an unshifted cosine and unshifted sine wave with amplitudes A_1 and B_1.

For the purpose of estimating one or the other of these pairs of parameters from a set of data, the advantage of representing the wave using the second line of Equation 10.57 rather than Equation 10.55 derives from the fact that the former is a linear function of the parameters A_1 and B_1. Notice that making the variable transformations $x_1 = \cos(2\pi\ t/n)$ and $x_2 = \sin(2\pi\ t/n)$, and substituting these into the second line of Equation 10.57, produces what looks like a two-predictor regression

equation with $A_1 = b_1$ and $B_1 = b_2$. In fact, given a data series y_t we can use this transformation together with ordinary regression software to find least-squares estimates of the parameters A_1 and B_1, with y_t as the predictand. Furthermore, the regression package will also produce the average of the predictand values as the intercept, b_0. Subsequently, the more convenient form of Equation 10.55 can be recovered by inverting Equations 10.58 to yield

$$C_1 = \left[A_1^2 + B_1^2\right]^{1/2} \tag{10.59a}$$

and

$$\phi_1 = \begin{cases} \tan^{-1}(B_1/A_1), & A_1 > 0 \\ \tan^{-1}(B_1/A_1) \pm \pi, \quad \text{or} \pm 180^\circ, & A_1 < 0. \\ \pi/2, \text{ or } 90^\circ, & A_1 = 0 \end{cases} \tag{10.59b}$$

Notice that since the trigonometric functions are periodic, effectively the same phase angle is produced by adding or subtracting a half-circle of angular measure if $A_1 < 0$. The alternative that yields $0 < \phi_1 < 2\pi$ is usually selected.

Finding the parameters A_1 and B_1 in Equation 10.57 using least-squares regression will work in the general case. For the special (although not too unusual) situation where the data are equally spaced in time with no missing values, the properties of the sine and cosine functions allow the same least-squares parameter values to be obtained more easily and efficiently using

$$A_1 = \frac{2}{n}\sum_{t=1}^{n} y_t \cos\left(\frac{2\pi t}{n}\right) \tag{10.60a}$$

and

$$B_1 = \frac{2}{n}\sum_{t=1}^{n} y_t \sin\left(\frac{2\pi t}{n}\right). \tag{10.60b}$$

Example 10.9. Harmonic Analysis of Average Monthly Temperatures

Table 10.5 presents the calculations necessary to obtain least-squares estimates for the parameters of the annual harmonic representing the Ithaca mean monthly temperatures plotted in Figure 10.13a, using Equations 10.60. The temperature data are shown in the column labeled y_t, and their average is easily computed as 552.9/12 = 46.1°F. The $n = 12$ terms of the sums in Equations 10.60a and 10.60b are shown in the last two columns. Applying Equations 10.60 to these yields $A_1 = (2/12)(-110.329) = -18.39$, and $B_1 = (2/12)(-86.417) = -14.40$.

Equation 10.59 transforms these two amplitudes to the parameters of the amplitude-phase form of Equation 10.55. This transformation allows easier comparison to the heavy curve plotted in Figure 10.13b. The amplitude is $C_1 = [-18.39^2 - 14.40^2]^{1/2} = 23.36$°F, and the phase angle is $\phi_1 = \tan^{-1}(-14.40/-18.39) + 180° = 218°$. Here 180° has been added rather than subtracted, so that $0° < \phi_1 < 360°$. The least-squares amplitude of $C_1 = 23.36$°F is quite close to the one used to draw Figure 10.13b, and the phase angle is 8° greater than the (7)(360°)/12 = 210° angle that was eyeballed on the basis of the July mean being the warmest of the 12 months. The value of $\phi_1 = 218°$ is a better estimate and implies a somewhat later (than mid-July) date for the time of the climatologically warmest temperature at this location. In fact, since there are very nearly as many degrees in a full cycle as there are days in one year, the results from Table 10.5 indicate that the heavy curve in Figure 10.13b should be

TABLE 10.5 Illustration of the Mechanics of Using Equations 10.60 to Estimate the Parameters of a Fundamental Harmonic

	y_t	cos $(2\pi\ t/12)$	sin $(2\pi\ t/12)$	y_t cos $(2\pi\ t/12)$	y_t sin $(2\pi\ t/12)$
1	22.2	0.866	0.500	19.225	11.100
2	22.7	0.500	0.866	11.350	19.658
3	32.2	0.000	1.000	0.000	32.200
4	44.4	−0.500	0.866	−22.200	38.450
5	54.8	−0.866	0.500	−47.457	27.400
6	64.3	−1.000	0.000	−64.300	0.000
7	68.8	−0.866	−0.500	−59.581	−34.400
8	67.1	−0.500	−0.866	−33.550	−58.109
9	60.2	0.000	−1.000	0.000	−60.200
10	49.5	0.500	−0.866	24.750	−42.867
11	39.3	0.866	−0.500	34.034	−19.650
12	27.4	1.000	0.000	27.400	0.000
Sums:	552.9	0.000	0.000	−110.329	−86.417

The data series y_t are the mean monthly temperatures at Ithaca for month t plotted in Figure 10.13a. Each of the 12 terms in Equations 10.60a and 10.60b, respectively, is shown in the last two columns.

shifted to the right by about one week. It is apparent that the result would be an improved correspondence with the data points. ◇

Example 10.10. Interpolation of the Annual Cycle to Average Daily Values

The calculations in Example 10.9 result in a smoothly varying representation of the annual cycle of mean temperature at Ithaca, based on the monthly values. Particularly if this were a location for which daily data were not available, it might be valuable to be able to use a function like this to represent the climatological average temperatures on a day-by-day basis. In order to employ the cosine curve in Equation 10.55 with time t in days, it would be necessary to use $n = 365$ days rather than $n = 12$ months. The amplitude can be left unchanged, although Epstein (1991) suggests a method to adjust this parameter that will produce a somewhat better representation of the annual cycle of daily values. In any case, however, it is necessary to make an adjustment to the phase angle.

Consider that the time $t = 1$ month represents all of January, and thus might be reasonably assigned to the middle of the month, perhaps the 15th. Thus the $t = 0$ months point of this function corresponds to the middle of December. Therefore when using $n = 365$ rather than $n = 12$, simply substituting the day number (1 January = 1, 2 January = 2, …, 1 February = 32, etc.) for the time variable will result in a curve that is shifted too far left by about two weeks. What is required is a new phase angle, say ϕ_1', consistent with a time variable t' in days, that will position the cosine function correctly.

On 15 December, the two time variables are $t = 0$ months, and $t' = -15$ days. On 31 December, they are $t = 0.5$ month = 15 days, and $t' = 0$ days. Thus in consistent units, $t' = t - 15$ days, or $t = t' + 15$ days. Substituting $n = 365$ days and $t = t' + 15$ into Equation 10.55 yields

$$\begin{aligned} y_t &= \bar{y} + C_1 \cos\left[\frac{2\pi t}{12} - \phi_1\right] = \bar{y} + C_1 \cos\left[\frac{2\pi\,(t'+15)}{365} - \phi_1\right] \\ &= \bar{y} + C_1 \cos\left[\frac{2\pi t'}{365} + 2\pi\frac{15}{365} - \phi_1\right] \\ &= \bar{y} + C_1 \cos\left[\frac{2\pi t'}{365} - \left(\phi_1 - 2\pi\frac{15}{365}\right)\right] \\ &= \bar{y} + C_1 \cos\left[\frac{2\pi t'}{365} - \phi_1'\right]. \end{aligned} \tag{10.61}$$

That is, the required new phase angle is $\phi_1' = \phi_1 - (2\pi)(15)/365$. ◇

10.4.4. Higher Harmonics

The computations in Example 10.9 produced a single cosine function passing quite close to the 12 monthly mean temperature values. This very good fit results because the shape of the annual cycle of temperature at this location is approximately sinusoidal, with a single full cycle being executed over the $n = 12$ points of the data series. We do not expect that a single harmonic wave will represent every time series this well. However, just as adding more predictors to a multiple regression will improve the fit to a set of training data, adding more cosine waves to a harmonic analysis will improve the fit to any time series.

Any data series consisting of n points can be represented exactly, meaning that a function can be found that passes through each of the points, by adding together a series of $n/2$ harmonic functions,

$$y_t = \bar{y} + \sum_{k=1}^{n/2} \left\{ C_k \cos\left[\frac{2\pi kt}{n} - \phi_k\right] \right\} \tag{10.62a}$$

$$= \bar{y} + \sum_{k=1}^{n/2} \left\{ A_k \cos\left[\frac{2\pi kt}{n}\right] + B_k \sin\left[\frac{2\pi kt}{n}\right] \right\}, \tag{10.62b}$$

which is sometimes called the *synthesis equation*. Equation 10.62b emphasizes that Equation 10.57 holds for any cosine wave, regardless of its frequency. The cosine wave that is the $k = 1$ term of Equation 10.62a is simply the fundamental, or first harmonic, that was the subject of the previous section. The other $n/2–1$ terms in the summation of Equation 10.62 are *higher harmonics*, or cosine waves with frequencies

$$\omega_k = \frac{2\pi k}{n} \tag{10.63}$$

that are integer multiples of the fundamental frequency ω_1.

For example, the second harmonic is that cosine function that completes exactly two full cycles over the n points of the data series. It has its own amplitude C_2 and phase angle ϕ_2. Notice that the factor k inside the cosine and sine functions in Equation 10.62a is of critical importance. When $k = 1$, the angle $\alpha = 2\pi\, kt/n$ varies through a single full cycle of 0 to 2π radians as the time index increased from $t = 0$ to $t = n$, as described earlier. In the case of the second harmonic where $k = 2$, $\alpha = 2\pi\, kt/n$ executes one full cycle as t increases from 0 to $n/2$, and then executes a second full cycle between $t = n/2$ and

$t = n$. Similarly, the third harmonic is defined by the amplitude C_3 and the phase angle ϕ_3, and varies through three cycles as t increases from 0 to n.

Equation 10.62b suggests that the coefficients A_k and B_k corresponding to particular data series y_t can be found using multiple regression methods, after the data transformations $x_1 = \cos(2\pi\, t/n)$, $x_2 = \sin(2\pi\, t/n)$, $x_3 = \cos(2\pi\, 2t/n)$, $x_4 = \sin(2\pi\, 2t/n)$, $x_5 = \cos(2\pi\, 3t/n)$, and so on. This is, in fact, the case in general, but if the data series is equally spaced in time and contains no missing values, Equation 10.60 generalizes to

$$A_k = \frac{2}{n}\sum_{t=1}^{n} y_t \cos\left(\frac{2\pi\, kt}{n}\right) \tag{10.64a}$$

and

$$B_k = \frac{2}{n}\sum_{t=1}^{n} y_t \sin\left(\frac{2\pi\, kt}{n}\right), \tag{10.64b}$$

which are sometimes called *analysis equations*. To compute a particular A_k, for example, these equations indicate than an n-term sum is formed, consisting of the products of the data series y_t with values of a cosine function executing k full cycles during the n time units. For relatively short data series these equations can be easily programmed or evaluated using spreadsheet software. For larger data series the A_k and B_k coefficients usually are computed using a more efficient method that will be mentioned in Section 10.5.3. Having computed these coefficients, the amplitude-phase form of the first line of Equation 10.62 can be arrived at by computing, separately for each harmonic,

$$C_k = \left[A_k^2 + B_k^2\right]^{1/2} \tag{10.65a}$$

and

$$\phi_k = \begin{cases} \tan^{-1}(B_k/A_k), & A_k > 0 \\ \tan^{-1}(B_k/A_k) \pm \pi, \text{ or} \pm 180^\circ, & A_k < 0 \\ \pi/2 \text{ , or } 90^\circ, & A_k = 0 \end{cases}. \tag{10.65b}$$

Recall that a multiple regression function will pass through all the developmental data points, and exhibit $R^2 = 100\%$, if there are as many predictor values as data points. The series of cosine terms in Equation 10.62a is an instance of this overfitting principle, because there are two parameters (the amplitude and phase) for each harmonic term. Thus the $n/2$ harmonics in Equation 10.62 consist of n predictor variables, and any set of data, regardless of how untrigonometric it may look, can be represented exactly using Equation 10.62.

Since the sample mean in Equation 10.62 is effectively one of the estimated parameters, corresponding to the regression intercept b_0, an adjustment to Equation 10.62 is required if n is odd. In this case a summation over only $(n–1)/2$ harmonics is required to completely represent the function. That is, $(n–1)/2$ amplitudes plus $(n–1)/2$ phase angles plus the sample average of the data equals n. If n is even, there are $n/2$ terms in the summation, but the phase angle for the final and highest harmonic, $\phi_{n/2}$, is zero.

We may or may not want to use all $n/2$ harmonics indicated in Equation 10.62, depending on the context. Often for defining, say, an annual cycle of a climatological quantity, the first few harmonics may give a quite adequate representation from a practical standpoint, and will typically be more accurate than simpler sample averages (e.g., 12 discrete monthly mean values), in terms of representing future data values not used in the fitting (Narapusetty et al., 2009). If the goal is to find a function passing exactly through each of the data points, then all $n/2$ harmonics would be used. Recall that Section 7.4 warned against overfitting in the context of developing forecast equations, because the artificial skill exhibited on the developmental data does not carry forward when the equation is used to forecast future independent data. In this latter case the goal would not be to forecast but rather to represent the data, so that the overfitting ensures that Equation 10.62 reproduces a particular data series exactly.

Example 10.11. A More Complicated Annual Cycle

Figure 10.14 illustrates the use of a small number of harmonics to smoothly represent the annual cycle of a climatological quantity. Here the quantity is the probability (expressed as a percentage) of five consecutive days without measurable precipitation, for El Paso, Texas. The irregular curve is a plot of the individual daily relative frequencies, computed using data for the years 1948–1983. These execute a regular but asymmetric annual cycle, with the wettest time of year being summer, and with dry springs and falls separated by a somewhat less dry winter. The figure also shows irregular, short-term fluctuations that have probably arisen mainly from sampling variations particular to the specific years analyzed. If a different sample of El Paso precipitation data had been used to compute the relative frequencies (say, 1984–2018), the same broad pattern would be evident, but the details of the individual "wiggles" would be different.

The annual cycle in Figure 10.14 is quite evident, yet it does not resemble a simple cosine wave. However, this cycle is reasonably well represented by the smooth curve, which is the sum of the first three harmonics. That is, the smooth curve is a plot of Equation 10.62 with three, rather than $n/2$, terms in the summation. The mean value for these data is 61.4%, and the parameters for the first two of these

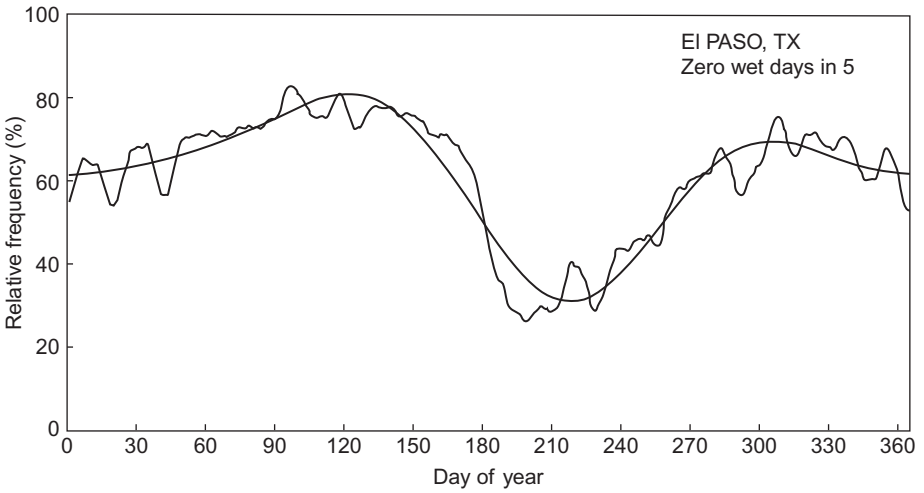

FIGURE 10.14 The annual cycle of the climatological probability that no measurable precipitation will fall during the five-day period centered on the date on the horizontal axis, for El Paso, Texas. Irregular line is the plot of the daily relative frequencies, and the smooth curve is a three-harmonic fit to the data. *From Epstein and Barnston (1988).*

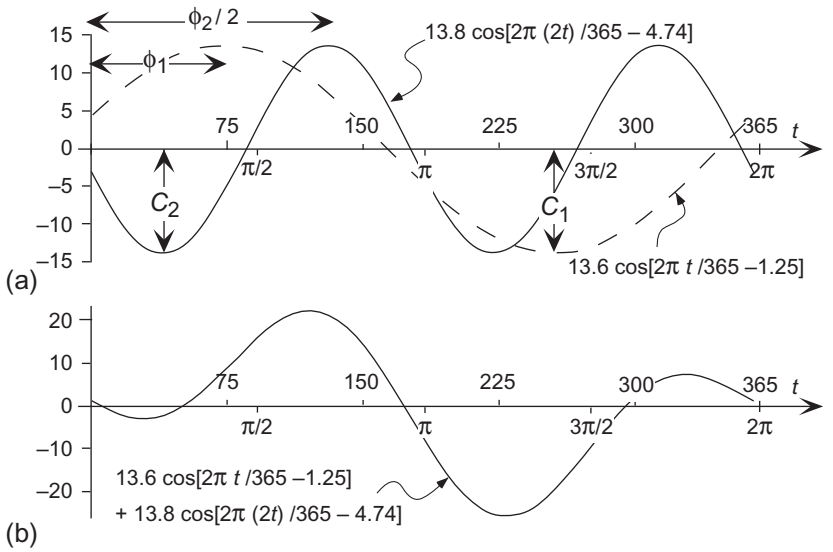

FIGURE 10.15 Illustration of the construction of the smooth curve in Figure 10.14. (a) The first (dashed) and second (solid) harmonics of the annual cycle plotted separately. These are defined by $C_1 = 13.6\%$, $\phi_1 = 72° = 0.4\ \pi$, $C_2 = 13.8\%$, and $\phi_2 = 272° = 1.51\ \pi$. The horizontal axis is labeled both in days and radians. (b) The smoothed representation of the annual cycle is produced by adding the values of the two functions in panel (a) for each time point. Subsequently adding the annual mean value of 61.4% produces a curve very similar to that in Figure 10.14. The small differences are accounted for by the third harmonic. Note that panels (a) and (b) have different vertical scales.

harmonics are $C_1 = 13.6\%$, $\phi_1 = 72° = 0.4\ \pi$, $C_2 = 13.8\%$, and $\phi_2 = 272° = 1.51\ \pi$. These values were computed from the underlying data using Equations 10.64 and 10.65. Computing and plotting the sum of all possible (365–1)/2 = 182 harmonics would result in a function identical to the irregular curve in Figure 10.14.

Figure 10.15 illustrates the construction of the smooth curve representing the annual cycle in Figure 10.14. Panel (a) shows the first (dashed) and second (solid) harmonics plotted separately, both as a function of time (t) in days and as a function of the corresponding angular measure in radians. Also indicated are the magnitudes of the amplitudes C_k in the vertical, and the correspondence of the phase angles ϕ_k to the maxima of the two functions. Note that since the second harmonic executes two cycles during the full 365 days of the year, there are two times of maximum, located at $\phi_2/2$ and $\pi + \phi_2/2$. (The maxima for the third harmonic would occur at $\phi_3/3$, $2\pi/3 + \phi_3/3$, and $4\pi/3 + \phi_3/3$, with a similar pattern holding for the higher harmonics.)

The curve in Figure 10.15b has been constructed by simply adding the values for the two functions in Figure 10.15a at each time point. (Note that the two panels in Figure 10.15 have been plotted using different vertical scales.) During times of the year where the two harmonics are of opposite sign but comparable magnitude, their sum is near zero. The maximum and minimum of the function in Figure 10.15b are achieved when its two components have relatively large magnitudes of the same sign. Adding the annual mean value of 61.4% to the lower curve results in a close approximation to the smooth curve in Figure 10.14, with the small differences between the two attributable to the third harmonic. ◇

10.5. FREQUENCY DOMAIN—II. SPECTRAL ANALYSIS

10.5.1. The Harmonic Functions as Uncorrelated Regression Predictors

Equation 10.62b suggests the use of multiple regression to find best-fitting harmonics for a given data series y_t. But for equally spaced data with no missing values Equations 10.64b will produce the same least-squares estimates for the coefficients A_k and B_k as will multiple regression software. Notice, however, that Equations 10.64 do not depend on any harmonic other than the one whose coefficients are being computed. That is, these equations depend on the current value of k, but not $k-1$, or $k-2$, or any other harmonic index. This fact implies that the coefficients A_k and B_k for any particular harmonic can be computed independently of those for any other harmonic.

Recall that usually regression parameters need to be recomputed each time a new predictor variable is entered into a multiple regression equation, and each time a predictor variable is removed from a regression equation. As noted in Section 7.3, this recomputation is necessary in the general case of sets of predictor variables that are mutually correlated, because correlated predictors carry redundant information to a greater or lesser extent. It is a remarkable property of the harmonic functions that (for equally spaced and complete data) they are uncorrelated so, for example, the parameters (amplitude and phase) for the first or second harmonic are the same whether or not they will be used in an equation with the third, fourth, or any other harmonics.

This attribute of the harmonic functions is a consequence of what is called the *orthogonality* property of the sine and cosine functions. That is, for integer harmonic indices k and j,

$$\sum_{t=1}^{n} \cos\left(\frac{2\pi kt}{n}\right) \sin\left(\frac{2\pi jt}{n}\right) = 0, \text{ for any integer values of } k \text{ and } j; \tag{10.66a}$$

and

$$\sum_{t=1}^{n} \cos\left(\frac{2\pi kt}{n}\right) \cos\left(\frac{2\pi jt}{n}\right) = \sum_{t=1}^{n} \sin\left(\frac{2\pi kt}{n}\right) \sin\left(\frac{2\pi jt}{n}\right) = 0, \text{ for } k \neq j. \tag{10.66b}$$

To illustrate, consider the two transformed predictor variables $x_1 = \cos[2\pi t/n]$ and $x_3 = \cos[2\pi (2t)/n]$. The Pearson correlation between these derived variables is given by

$$r_{x_1x_3} = \frac{\sum_{t=1}^{n} (x_1 - \bar{x}_1)(x_3 - \bar{x}_3)}{\left[\sum_{t=1}^{n} (x_1 - \bar{x}_1)^2 \sum_{t=1}^{n} (x_3 - \bar{x}_3)^2\right]}, \tag{10.67a}$$

and since the averages $\bar{x}_1$ and $\bar{x}_3$ of cosine functions over integer numbers of cycles are zero,

$$r_{x_1x_3} = \frac{\sum_{t=1}^{n} \cos\left(\frac{2\pi t}{n}\right) \cos\left(\frac{2\pi 2t}{n}\right)}{\left[\sum_{t=1}^{n} \cos^2\left(\frac{2\pi t}{n}\right) \cos^2\left(\frac{2\pi 2t}{n}\right)\right]^{1/2}} = 0, \tag{10.67b}$$

because the numerator is zero by Equation 10.66b.

Since the relationships between harmonic predictor variables and the data series y_t do not depend on what other harmonic functions are also being used to represent the series, the proportion of the variance of y_t accounted for by each harmonic is also fixed. Expressing this proportion as the R^2 statistic commonly computed in regression, the R^2 for the kth harmonic is simply

$$R_k^2 = \frac{(n/2)\,C_k^2}{(n-1)\,s_y^2}. \tag{10.68}$$

In terms of the regression ANOVA table, the numerator of Equation 10.68 is the regression sum of squares for the kth harmonic. The factor s_y^2 is simply the sample variance of the data series, so the denominator of Equation 10.68 is the total sum of squares, SST. Notice that the strength of the relationship between the kth harmonic and the data series can be expressed entirely in terms of the ratio C_k/s_y. The phase angle ϕ_k is necessary only to determine the positioning of the cosine curve in time. Furthermore, since each harmonic provides independent information about the data series, the joint R^2 exhibited by a regression equation with only harmonic predictors is simply the sum of the R_k^2 values for each of the harmonics,

$$R^2 = \sum_{k \text{ in the equation}} R_k^2. \tag{10.69}$$

If all the $n/2$ possible harmonics are used as predictors (Equation 10.62), then the total R^2 in Equation 10.69 will be exactly 1.

Equation 10.62 says that a data series y_t of length n can be specified completely in terms of the n parameters of $n/2$ harmonic functions. Equivalently, we can take the view that the data y_t are transformed into new set of quantities A_k and B_k according to Equations 10.64. For this reason, Equations 10.64 are called the *discrete Fourier transform*. The data series can be represented as the n quantities C_k and ϕ_k, obtained from the A_k's and B_k's using the transformations in Equations 10.65. According to Equations 10.68 and 10.69, this data transformation accounts for all of the variation in the series y_t. Another perspective on Equations 10.68 and 10.69 is that the variance of the time-series variable y_t can be apportioned among the $n/2$ harmonic functions, each of which represents a different timescale of variation.

10.5.2. The Periodogram or Fourier Line Spectrum

The foregoing suggests that a different way to look at a time series is as a collection of Fourier coefficients A_k and B_k that are a function of frequency ω_k (Equation 10.63), rather than as a collection of data points y_t measured as a function of time. The advantage of this new perspective is that it allows us to see separately the contributions to a time series that are made by processes varying at different speeds, that is, by processes operating at a spectrum of different frequencies. Panofsky and Brier (1958) illustrate this distinction with an analogy: "An optical spectrum shows the contributions of different wave lengths or frequencies to the energy of a given light source. The spectrum of a time series shows the contributions of oscillations with various frequencies to the variance of a time series." Even if the underlying physical basis for a data series y_t is not really well represented by a series of cosine waves, often much can still be learned about the data by viewing it from this perspective.

The characteristics of a time series that has been Fourier transformed into the frequency domain are most often examined graphically, using a plot known as the *periodogram*, or *Fourier line spectrum*. This plot sometimes is called the *power spectrum*, or simply the spectrum, of the data series. In simplest form, the plot of a spectrum consists of the squared amplitudes C_k^2 as a function of the frequencies ω_k. Note that

information contained in the phase angles ϕ_k is not portrayed in the spectrum. Therefore the spectrum conveys the proportion of variation in the original data series accounted for by oscillations at the harmonic frequencies, but does not supply information about when in time these oscillations are expressed. Fisher (2006) suggests that this characteristic is similar to what would be achieved by representing the frequencies of use of the various pitches in a piece of music with a histogram. Such a histogram might well identify the musical key, but not the piece of music itself. A spectrum thus does not provide a full picture of the behavior of the time series from which it has been calculated and is not sufficient to reconstruct the time series.

The vertical axis of a plotted spectrum is sometimes numerically rescaled, in which case the plotted points are proportional to the squared amplitudes. One choice for this proportional rescaling is that in Equation 10.68. It is also common for the vertical axis of a spectrum to be plotted on a logarithmic scale. Plotting the vertical axis logarithmically is particularly useful if the variations in the time series are dominated by harmonics of only a few frequencies. In this case a linear plot might result in the remaining spectral components being invisibly small. A logarithmic vertical axis also regularizes the representation of confidence limits for the spectral estimates (Section 10.5.6).

The horizontal axis of the line spectrum consists of $n/2$ frequencies ω_k if n is even, and $(n–1)/2$ frequencies if n is odd. The smallest of these will be the lowest frequency $\omega_1 = 2\pi/n$ (the fundamental frequency), and this corresponds to the cosine wave that executes a single cycle over the n time points. The highest frequency, $\omega_{n/2} = \pi$, is called the *Nyquist frequency*. It is the frequency of the cosine wave that executes a full cycle over only two time intervals, and which executes $n/2$ cycles over the full data record. The Nyquist frequency depends on the time resolution of the original data series y_t and imposes an important limitation on the information available from a spectral analysis.

The horizontal axis of a plotted spectrum is often simply the angular frequency, ω, with units of radians/time. A common alternative is to use the frequencies

$$f_k = \frac{k}{n} = \frac{\omega_k}{2\pi}, \tag{10.70}$$

which have dimensions of time^{-1}. Under this alternative convention, the allowable frequencies range from $f_1 = 1/n$ for the fundamental to $f_{n/2} = 1/2$ for the Nyquist frequency. The horizontal axis of a spectrum can also be scaled according to the reciprocal of the frequency, or the period of the kth harmonic,

$$\tau_k = \frac{n}{k} = \frac{2\pi}{\omega_k} = \frac{1}{f_k}. \tag{10.71}$$

The period τ_k specifies the length of time required for a cycle of frequency ω_k to be completed. Associating periods with the periodogram estimates can help visualize the timescales on which the important variations in the data are occurring.

Example 10.12. Discrete Fourier Transform of a Small Data Set

Table 10.6 presents a simple data set and its discrete Fourier transform. The leftmost columns contain the observed average monthly temperatures at Ithaca, New York, for the two years 1987 and 1988. This is such a familiar type of data that, even without doing a spectral analysis, we know in advance that the primary feature will be the annual cycle of cold winters and warm summers. This expectation is validated by the plot of the data in Figure 10.16a, which shows these temperatures as a function of time. The overall impression is of a data series that is approximately sinusoidal with a period of 12 months, but that a single cosine wave with this period would not pass exactly through all the points.

TABLE 10.6 Average Monthly Temperatures, °F, at Ithaca, New York, for 1987–1988, and Their Discrete Fourier Transform

Month	1987	1988	k	τ_k, months	A_k	B_k	C_k
1	21.4	20.6	1	24.00	−0.14	0.44	0.46
2	17.9	22.5	2	12.00	−23.76	−2.20	23.86
3	35.9	32.9	3	8.00	−0.99	0.39	1.06
4	47.7	43.6	4	6.00	−0.46	−1.25	1.33
5	56.4	56.5	5	4.80	−0.02	−0.43	0.43
6	66.3	61.9	6	4.00	−1.49	−2.15	2.62
7	70.9	71.6	7	3.43	−0.53	−0.07	0.53
8	65.8	69.9	8	3.00	−0.34	−0.21	0.40
9	60.1	57.9	9	2.67	1.56	0.07	1.56
10	45.4	45.2	10	2.40	0.13	0.22	0.26
11	39.5	40.5	11	2.18	0.52	0.11	0.53
12	31.3	26.7	12	2.00	0.79	—	0.79

Columns 4 to 8 of Table 10.6 show the same data after being subjected to the discrete Fourier transform. Since $n = 24$ is an even number, the data are completely represented by $n/2 = 12$ harmonics. These are indicated by the rows labeled by the harmonic index, k. Column 5 of Table 10.6 indicates the period (Equation 10.71) of each of the 12 harmonics used to represent the data. The period of the fundamental frequency, $\tau_1 = 24$ months, is equal to the length of the data record. Since there are two annual cycles in the $n = 24$ month record, it is the $k = 2^{\text{nd}}$ harmonic with period $\tau_2 = 24/2 = 12$ months that is expected to be most important. The Nyquist frequency is $\omega_{12} = \pi$ radians/month, or $f_{12} = 0.5$ month^{-1}, corresponding to the period $\tau_{12} = 2$ months.

The coefficients A_k and B_k that could be used in Equation 10.62 to reconstruct the original data are shown in the next columns of the table. These constitute the discrete Fourier transform of the data series of temperatures. Notice that there are only 23 Fourier coefficients, because 24 independent pieces of information are necessary to fully represent the $n = 24$ data points, including the sample mean of 46.1°F. To use the synthesis Equation 10.62 to reconstitute the data, we would substitute $B_{12} = 0$.

Column 8 in Table 10.6 shows the amplitudes C_k, computed according to Equation 10.65a. The phase angles could also be computed, using Equation 10.65b, but these are not needed to plot the spectrum. Figure 10.16b shows the spectrum for these temperature data, plotted in the form of a histogram. The vertical axis consists of the squared amplitudes C_k^2, normalized according to Equation 10.68 to show the R^2 attributable to each harmonic. The horizontal axis is linear in frequency, but the corresponding periods are also shown, to aid the interpretation. Clearly most of the variation in the data is described by the second harmonic, the R^2 for which is 97.5%. As expected, the variations of the annual cycle dominate these data, but the fact that the amplitudes of the other harmonics are not zero indicates that the data do not consist of a pure cosine wave with a frequency of $f_2 = 1$ year $^{-1}$. Notice, however, that the logarithmic vertical axis tends to deemphasize the smallness of these other harmonics. If the vertical axis had been scaled linearly, the plot would consist of a spike at $k = 2$ and a small bump at $k = 6$, with the rest of the points being essentially indistinguishable from the horizontal axis. ◇

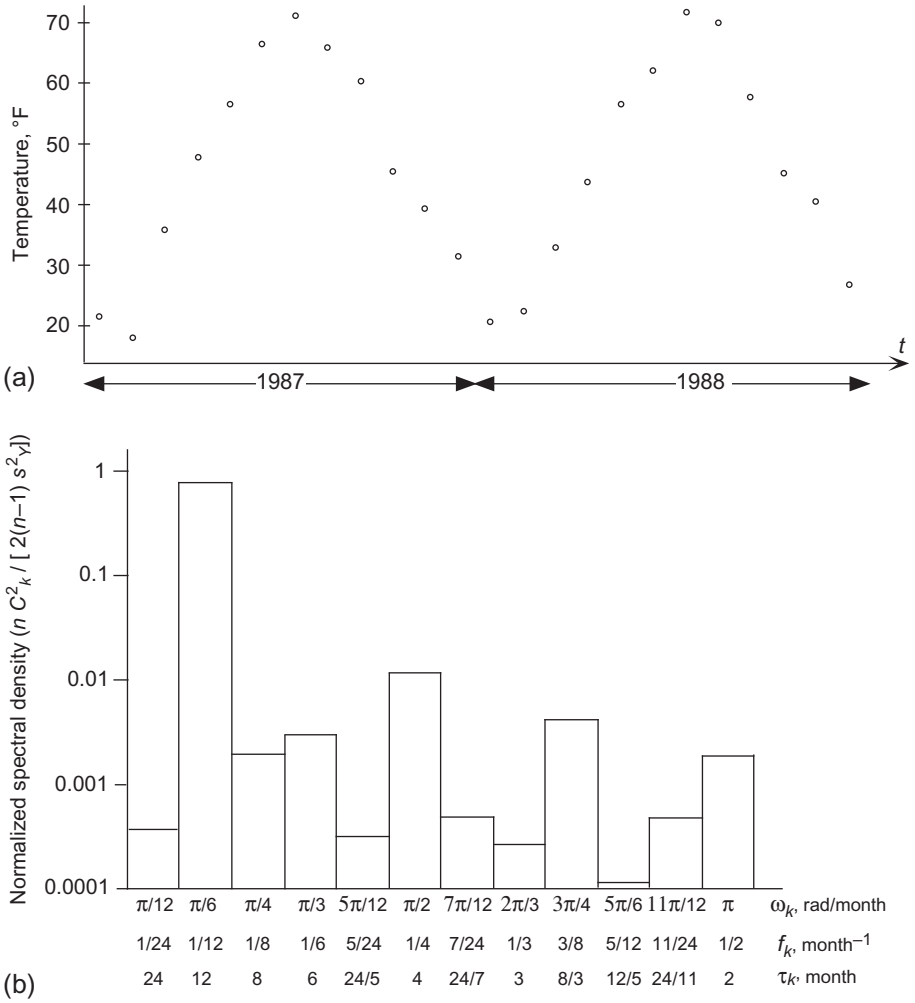

FIGURE 10.16 Illustration of the relationship between a simple time series and its spectrum. (a) Average monthly temperatures at Ithaca, New York for 1987–1988, from Table 10.6. The data are approximately sinusoidal, with a period of 12 months. (b) The spectrum of the data series in panel (a), plotted in the form of a histogram, and expressed in the normalized form of Equation 10.68. Clearly the most important variations in the data series are represented by the second harmonic, with period $\tau_2 = 12$ months, which is the annual cycle. Note that the vertical scale is logarithmic, so that the next most important harmonic accounts for barely more than 1% of the total variation. The horizontal scale is linear in frequency.

Example 10.13. Another Sample Spectrum

A less trivial example spectrum is shown in Figure 10.17. This is a portion of the (smoothed) spectrum of the monthly Tahiti minus Darwin sea-level pressure time series for 1951–1979. That time series resembles the (normalized) SOI index shown in Figure 3.16, including the tendency for quasiperiodic behavior. That the variations in the time series are not strictly periodic is evident from the irregular variations in Figure 3.16, and from the broad (i.e., spread over many frequencies) maximum in the spectrum.

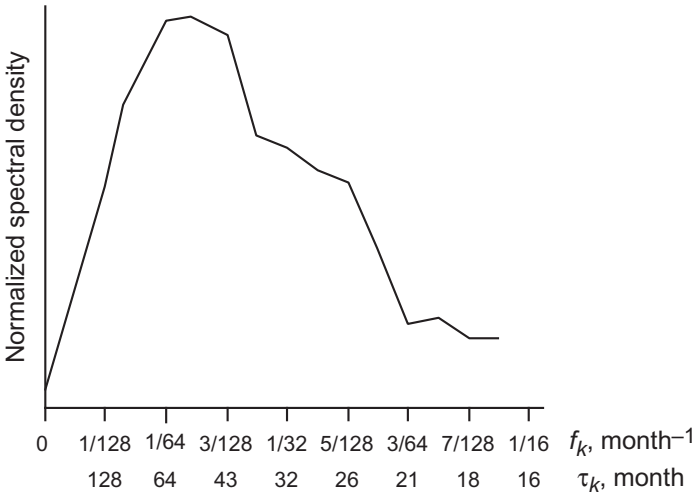

FIGURE 10.17 The low-frequency portion of the smoothed spectrum for the monthly time series of Tahiti minus Darwin sea-level pressures, 1951–1979. The underlying time series resembles that in Figure 3.16, and the tendency for oscillations to occur in roughly three- to seven-year cycles is reflected in the broad maximum of the spectrum in this range. *After Chen (1982). © American Meteorological Society. Used with permission.*

Figure 10.17 indicates that the typical length of one of these pseudocycles (corresponding to typical times between El Niño events) is something between $\tau = [(1/36)\ \text{mo}^{-1}]^{-1} = 3$ years and $\tau = [(1/84)\ \text{mo}^{-1}]^{-1} = 7$ years.

The vertical axis in Figure 10.17 has been plotted on a linear scale, but units have been omitted because they do not contribute to a qualitative interpretation of the plot. The horizontal axis is linear in frequency, and therefore nonlinear in period. Notice also that the labeling of the horizontal axis indicates that the full spectrum of the underlying data series has not been presented in the figure. Since the data series consists of monthly values, the Nyquist frequency must be 0.5 month^{-1}, corresponding to a period of two months. Only the left-most one-eighth of the spectrum has been shown because it is these lower frequencies that reflect the physical phenomenon of interest, namely, the ENSO cycle. The estimated spectral density function for the omitted higher frequencies would show only a long, irregular, and generally uninformative right tail exhibiting small spectral density estimates. ◇

10.5.3. Computing Spectra

One way to compute the spectrum of a data series is simply to apply Equations 10.64 and then to find the amplitudes using Equation 10.65a. This is a reasonable approach for relatively short data series and can be programmed easily using, for example, spreadsheet software. These equations would be implemented only for $k = 1, 2, \ldots, (n/2–1)$. Because we want exactly n Fourier coefficients (A_k and B_k) to represent the n points in the data series, the computation for the highest harmonic, $k = n/2$, is done using

$$A_{n/2} = \begin{cases} \left(\frac{1}{2}\right)\left(\frac{2}{n}\right)\sum_{t=1}^{n} y_t \cos\left[\frac{2\pi\,(n/2)t}{n}\right] = \left(\frac{1}{n}\right)\sum_{t=1}^{n} y_t \cos[\pi t], & n \text{ even} \\ 0, & n \text{ odd} \end{cases} \tag{10.72a}$$

and

$$B_{n/2} = 0, \quad n \text{ even or odd.} \tag{10.72b}$$

Although straightforward notationally, this method of computing the discrete Fourier transform is quite inefficient computationally. In particular, many of the calculations called for by Equation 10.64 are redundant. Consider, for example, the data for April 1987 in Table 10.6. The term for $t = 4$ in the summation in Equation 10.64b is (47.7°F) sin[(2π)(1)(4)/24] = (47.7°F) (0.866) = 41.31°F. However, the term involving this same data point for $k = 2$ is exactly the same: (47.7°F) sin[(2π)(2)(4)/24] = (47.7°F)(0.866) = 41.31°F. There are many other such redundancies in the computation of discrete Fourier transforms using Equations 10.64. These can be avoided through the use of clever algorithms known as *Fast Fourier Transforms* (*FFT*s) (Cooley and Tukey, 1965). Most scientific software packages include one or more FFT routines, which give very substantial speed improvements, especially as the length of the data series increases. In comparison to computation of the Fourier coefficients using a regression approach, an FFT is approximately $n/\log_2(n)$ times faster; or about 15 times faster for $n = 100$, and about 750 times faster for $n = 10000$.

It is worth noting that FFTs usually are documented and implemented in terms of the *Euler complex exponential notation*,

$$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t), \tag{10.73}$$

where i is the unit imaginary number satisfying $i = \sqrt{-1}$, and $i^2 = -1$. Complex exponentials are used rather than sines and cosines purely as a notational convenience that makes some of the manipulations less cumbersome. The mathematics are still entirely the same. In terms of complex exponentials, Equation 10.62 becomes

$$y_t = \overline{y} + \sum_{k=1}^{n/2} H_k \, e^{i\,[2\pi\, k/n]t}, \tag{10.74}$$

where H_k is the complex Fourier coefficient

$$H_k = A_k + i B_k. \tag{10.75}$$

That is, the real part of H_k is the coefficient A_k, and the imaginary part of H_k is the coefficient B_k.

10.5.4. Aliasing

Aliasing is a hazard inherent in spectral analysis of discrete data. It arises because of the limits imposed by the sampling interval, or the time between consecutive pairs of data points. Since a minimum of two points are required to even sketch a cosine wave—one point for the peak and one point for the trough—the highest representable frequency is the Nyquist frequency, with $\omega_{n/2} = \pi$, or $f_{n/2} = 0.5$. A wave of this frequency executes one cycle every two data points, and thus a discrete data set can represent explicitly variations that occur no faster than this speed.

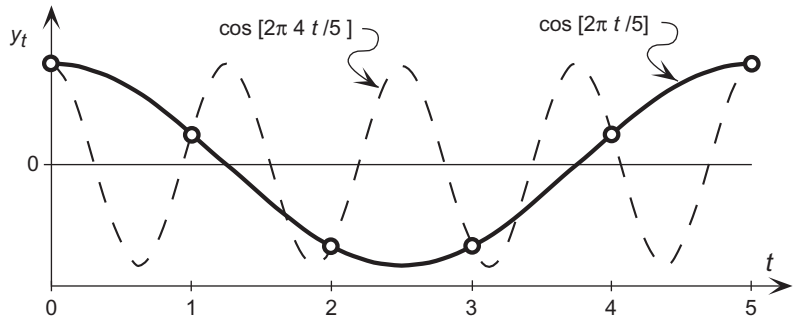

FIGURE 10.18 Illustration of the basis of aliasing. Heavy circles represent data points in a time series y_t. Fitting a harmonic function to them results in the heavy curve. However, if the data series actually had been produced by the process indicated by the light dashed curve, the fitted heavy curve would present the misleading impression that the source of the data was actually fluctuating at the lower frequency. The lighter curve has not been sampled densely enough because its frequency, $\omega = 8\pi/5$ (or $f = 4/5$), is higher than the Nyquist frequency of $\omega = \pi$ (or $f = 1/2$). Variations at the frequency of the dashed curve are said to be aliased to the frequency of the heavier curve.

It is worth wondering what happens to the spectrum of a data series if it includes an important physical process that varies faster than the Nyquist frequency. If so, the data series is said to be *undersampled*, which means that the points in the time series are spaced too far apart to properly capture these fast variations. However, variations that occur at frequencies higher than the Nyquist frequency do not disappear. Rather, their contributions are spuriously attributed to some lower but representable frequency, between ω_1 and $\omega_{n/2}$. These high-frequency variations are said to be aliased.

Figure 10.18 illustrates the meaning of aliasing. Imagine that the physical data-generating process is represented by the dashed cosine curve. The data series y_t is produced by sampling this process at integer values of the time index t, resulting in the points indicated by the circles. However, the frequency of the dashed curve ($\omega = 8\pi/5$, or $f = 4/5$) is higher than the Nyquist frequency ($\omega = \pi$, or $f = 1/2$), meaning that it oscillates too quickly to be adequately sampled at this time resolution. If only the information in the discrete time points is available, these data look like the heavy cosine function, the frequency of which ($\omega = 2\pi/5$, or $f = 1/5$) is lower than the Nyquist frequency, and is therefore representable. Note that because the cosine functions are orthogonal, this same effect will occur whether or not variations of different frequencies are also present in the data.

The effect of aliasing on spectral analysis is that any energy (squared amplitudes) attributable to processes varying at frequencies higher than the Nyquist frequency will be erroneously added to that of one of the $n/2$ frequencies that are represented by the discrete Fourier spectrum. A frequency $f_A > 1/2$ will be aliased into one of the representable frequencies f (with $0 < f \leq 1/2$) if it differs by an integer multiple of 1 time^{-1}, that is, if

$$f_A = j \pm f, j \text{ any positive integer.} \tag{10.76a}$$

In terms of angular frequency, variations at the aliased frequency ω_A appear to occur at the representable frequency ω if

$$\omega_A = 2\pi j \pm \omega, j \text{ any positive integer.} \tag{10.76b}$$

These equations imply that the squared amplitudes for frequencies higher than the Nyquist frequency will be added to the representable frequencies in an accordion-like pattern, with each "fold" of the

accordion occurring at an integer multiple of the Nyquist frequency. For this reason the Nyquist frequency is sometimes called the *folding frequency*. An aliased frequency f_A that is just slightly higher than the Nyquist frequency of $f_{n/2} = 1/2$ is aliased to a frequency slightly lower than 1/2. Frequencies only slightly lower than twice the Nyquist frequency are aliased to frequencies only slightly higher than zero. The pattern then reverses for $2f_{n/2} < f_A < 3f_{n/2}$. That is, frequencies just higher than $2f_{n/2}$ are aliased to very low frequencies, and frequencies almost as high as $3f_{n/2}$ are aliased to frequencies near $f_{n/2}$.

Figure 10.19 illustrates the effects of aliasing on a hypothetical spectrum. The gray line represents the true spectrum, which exhibits a concentration of power at low frequencies, but also has a sharp peak at $f = 5/8$ and a broader peak at $f = 19/16$. These second two peaks occur at frequencies higher than the Nyquist frequency of $f = 1/2$, which means that the physical process that generated the data was not sampled often enough to resolve them explicitly. The variations actually occurring at the frequency $f_A = 5/8$ are aliased to (i.e., appear to occur at) the frequency $f = 3/8$. That is, according to Equation 10.76a, $f_A = 1 - f = 1 - 3/8 = 5/8$. In the spectrum, the squared amplitude for $f_A = 5/8$ is added to the (genuine) squared amplitude at $f = 3/8$ in the true spectrum. Similarly, the variations represented by the broader hump centered at $f_A = 19/16$ in the true spectrum are aliased to frequencies at and around $f = 3/16$ ($f_A = 1 + f = 1 + 3/16 = 19/16$). The dashed line in Figure 10.19 indicates the portions of the aliased spectral energy (the total area between the gray and black lines) contributed by frequencies between $f = 1/2$ and $f = 1$ (the area below the dashed line), and by frequencies between $f = 1$ and $f = 3/2$ (the area above the dashed line); emphasizing the fan-folded nature of the aliased spectral density.

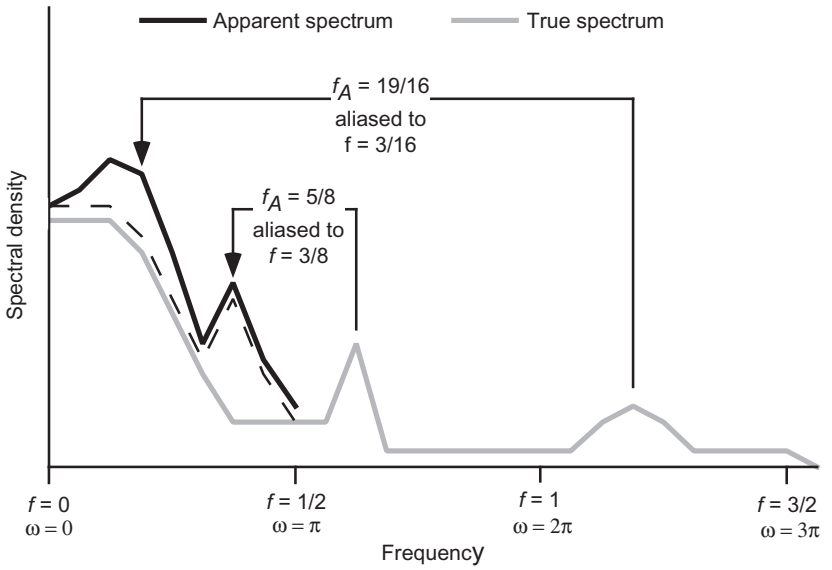

FIGURE 10.19 Illustration of aliasing in a hypothetical spectrum. The true spectrum (gray line) exhibits a sharp peak at $f = 5/8$, and a broader peak at $f = 19/16$. Since both of these frequencies are higher than the Nyquist frequency $f = 1/2$, they are aliased in the spectrum (erroneously attributed) to the frequencies indicated. The aliasing follows an accordion-like pattern, with the area between the gray line and the dashed line contributed by frequencies from $f = 1$ to $f = 1/2$, and the area between the dashed line and the heavy line contributed by frequencies between $f = 1$ and $f = 3/2$. The resulting apparent spectrum (heavy line) includes both the true spectral density values for frequencies between zero and 1/2, as well as the contributions from the aliased frequencies.

Aliasing can be particularly severe when isolated segments of a time series are averaged and then analyzed, for example, a time series of average January values in each of n years. This problem has been studied by Madden and Jones (2001), who conclude that badly aliased spectra are expected to result unless the averaging time is at least as large as the sampling interval. For example, a spectrum for January averages is expected to be heavily aliased since the one-month averaging period is much shorter than the annual sampling interval.

Unfortunately, once a data series has been collected, there is no way to "de-alias" its spectrum. That is, it is not possible to tell from the data values alone whether appreciable contributions to the spectrum have been made by frequencies higher than $f_{n/2}$, or how large these contributions might be. In practice, it is desirable to have an understanding of the physical basis of the processes generating the data series, so that it can be seen in advance that the sampling rate is adequate. Of course in an exploratory setting this advice is of no help, since the point of an exploratory analysis is exactly to gain a better understanding of an unknown or a poorly known generating process. In this latter situation, we would like to see the spectrum approach zero for frequencies near $f_{n/2}$, which would give some hope that the contributions from higher frequencies are minimal. A spectrum such as the heavy line in Figure 10.19 would lead us to expect that aliasing might be a problem, since its not being essentially zero at the Nyquist frequency could well mean that the true spectrum is nonzero at higher frequencies as well.

10.5.5. The Spectra of Autoregressive Models

Another perspective on the time-domain autoregressive models described in Section 10.3 is provided by their spectra. The types of time dependence produced by different autoregressive models produce characteristic spectral signatures that can be related to the autoregressive parameters.

The simplest case is the AR(1) process, Equation 10.16. Here positive values of the single autoregressive parameter ϕ induce a memory into the time series that tends to smooth over short-term (high-frequency) variations in the ε series, and emphasize the slower (low-frequency) variations. In terms of the spectrum, these effects lead to more density at lower frequencies and less density at higher frequencies. Furthermore, these effects are progressively stronger for ϕ closer to 1.

These ideas are quantified by the spectral density function for AR(1) processes,

$$S(f) = \frac{4\sigma_\varepsilon^2/n}{1 + \phi^2 - 2\phi\cos(2\pi f)}, 0 \le f \le 1/2. \tag{10.77}$$

This is a function that associates a spectral density with all frequencies in the range $0 \le f \le 1/2$. The shape of the function is determined by the denominator, and the numerator contains scaling constants that give the function numerical values that are comparable to the empirical spectrum of squared amplitudes, C_k^2. This equation also applies for negative values of the autoregressive parameter, which produce time series tending to oscillate quickly around the mean, and for which the spectral density is greatest at the high frequencies.

Note that, for zero frequency, Equation 10.77 is proportional to the variance of a time average. This can be appreciated by substituting $f = 0$, and Equations 10.21 and 10.39 into Equation 10.77, and comparing to Equation 10.36. Thus the extrapolation of the spectrum to zero frequency has been used to estimate variances of time averages (e.g., Madden and Shea, 1978).

Figure 10.20 shows spectra for the AR(1) processes having $\phi = 0.5$, 0.3, 0.0, and –0.6. The autocorrelation functions for the first and last of these are shown as insets in Figure 10.7. As might have been

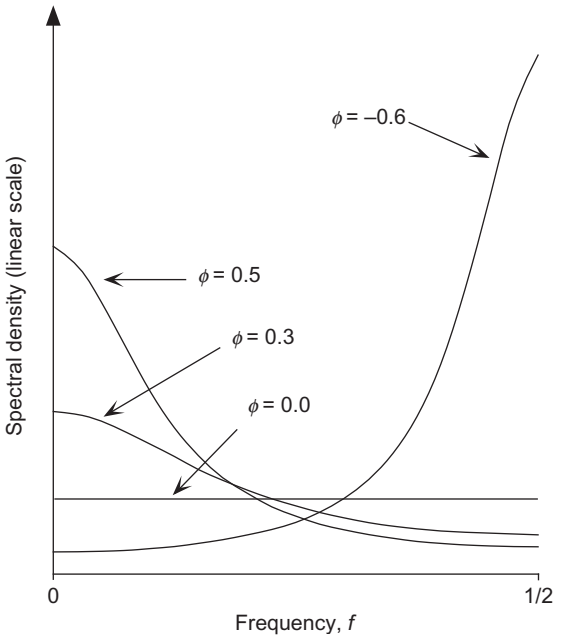

FIGURE 10.20 Spectral density functions for four AR(1) processes, computed using Equation 10.77. Autoregressions with $\phi > 0$ are red-noise processes, since their spectra are enriched at the lower frequencies and depleted at the higher frequencies. The spectrum for the $\phi = 0$ process (i.e., serially independent data) is flat, exhibiting no tendency to emphasize either high- or low-frequency variations. This is a white-noise process. The autoregression with $\phi = -0.6$ tends to exhibit rapid variations around its mean, which results in a spectrum enriched in the high frequencies, or a blue-noise process. Autocorrelation functions for the $\phi = 0.5$ and $\phi = -0.6$ processes are shown as insets in Figure 10.7.

anticipated, the two processes with $\phi > 0$ show enrichment of the spectral densities in the lower frequencies and depletion in the higher frequencies, and these characteristics are stronger for the process with the larger autoregressive parameter. By analogy to the properties of visible light, AR(1) processes with $\phi > 0$ are sometimes referred to as *red-noise* processes.

AR(1) processes with $\phi = 0$ consist of series of temporally uncorrelated data values $x_{t+1} = \mu + \varepsilon_{t+1}$ (compare Equation 10.16). These exhibit no tendency to emphasize either low-frequency or high-frequency variations, so their spectrum is constant or flat. Again by analogy to visible light, this is called *white noise* because of the equal mixture of all frequencies. This analogy is the basis of the independent series of ε's being called the white-noise forcing, and of the parameter σ_ε^2 being known as the white-noise variance.

Finally, the AR(1) process with $\phi = -0.6$ tends to produce erratic short-term variations in the time series, resulting in negative correlations at odd lags and positive correlations at even lags. (This kind of correlation structure is rare in atmospheric time series.) The spectrum for this process is thus enriched at the high frequencies and depleted at the low frequencies, as indicated in Figure 10.20. Such series are accordingly known as *blue-noise* processes.

Expressions for the spectra of other autoregressive processes, and for ARMA processes as well, are given in Box and Jenkins (1976). Of particular importance is the spectrum for the AR(2) process,

$$S(f) = \frac{4\sigma_\varepsilon^2/n}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1-\phi_2)\cos(2\pi f) - 2\phi_2\cos(4\pi f)}, 0 \le f \le 1/2. \qquad (10.78)$$

This equation reduces to Equation 10.77 for $\phi_2 = 0$, since an AR(2) process (Equation 10.27) with $\phi_2 = 0$ is simply an AR(1) process.

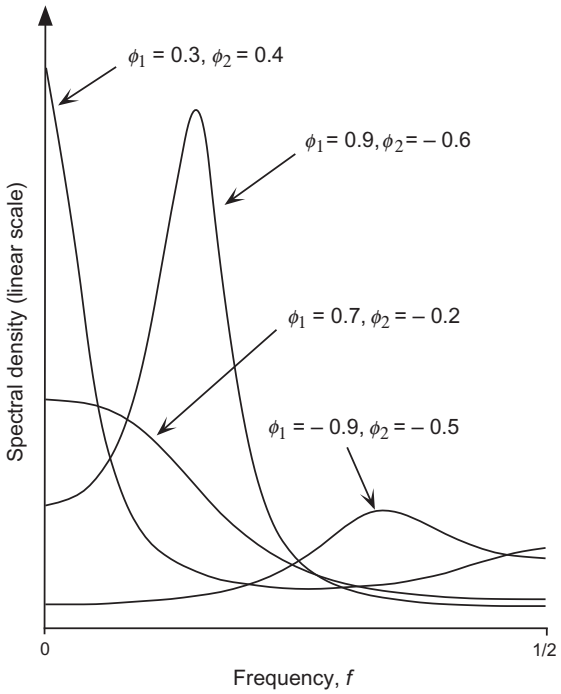

FIGURE 10.21 Spectral density functions for four AR(2) processes, computed using Equation 10.78. The diversity of the forms of the of spectra in this figure illustrates the flexibility of the AR(2) model. The autocorrelation functions for these autoregressions are shown as insets in Figure 10.7.

The AR(2) processes are particularly interesting because of their capacity to exhibit a wide variety of behaviors, including pseudoperiodicities. This diversity is reflected in the various forms of the spectra that are included in Equation 10.78. Figure 10.21 illustrates a few of these, corresponding to the AR(2) processes whose autocorrelation functions are shown as insets in Figure 10.7. The processes with $\phi_1 = 0.9$, $\phi_2 = -0.6$, and $\phi_1 = -0.9$, $\phi_2 = -0.5$, exhibit pseudoperiodicities, as indicated by the broad humps in their spectra at intermediate frequencies. The process with $\phi_1 = 0.3$, $\phi_2 = 0.4$ exhibits most of its variation at low frequencies, but also shows a smaller maximum at high frequencies. The spectrum for the process with $\phi_1 = 0.7$, $\phi_2 = -0.2$ resembles the red-noise spectra in Figure 10.20, although with a broader and flatter low-frequency maximum.

Example 10.14 Smoothing a Sample Spectrum Using an Autoregressive Model

The equations for spectra of autoregressive models can be useful in interpreting sample spectra from data series. The erratic sampling properties of the individual periodogram estimates, as described in Section 10.5.6, can make it difficult to discern features of the true spectrum that underlie a particular sample spectrum. However, if a well-fitting time-domain model can be estimated from the same data series, its spectrum can be used to guide the eye through the sample spectrum. Autoregressive models are sometimes fitted to time series for the sole purpose of obtaining smooth spectra. Chu and Katz (1989) show that the spectrum corresponding to a time-domain model fit using the SOI time series (see Figure 10.17) corresponds well to the spectrum computed directly from the data.

Consider the data series in Figure 10.8c, which was generated according to the AR(2) process with $\phi_1 = 0.9$ and $\phi_2 = -0.6$. The sample spectrum for this particular batch of 50 points is shown as the solid

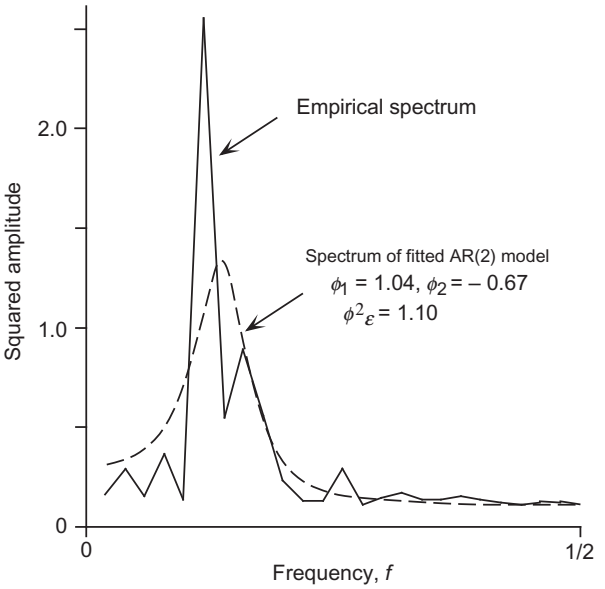

FIGURE 10.22 Illustration of the use of the spectrum of a fitted autoregressive model to guide the eye in interpreting a sample spectrum. The solid curve is the sample spectrum for the $n = 50$ data points shown in Figure 10.8c, generated by the AR(2) process with $\phi_1 = 0.9$, $\phi_2 = -0.6$, and $\sigma_\varepsilon^2 = 1.0$. A fuller perspective on this spectrum is provided by the dashed line, which is the spectrum of the AR(2) process fitted to this same series of 50 data points.

function in Figure 10.22. Apparently the series exhibits pseudoperiodicities in the frequency range around $f = 0.12$ through $f = 0.16$, but sampling variability makes the interpretation somewhat difficult. Although the empirical spectrum in Figure 10.22 somewhat resembles the theoretical spectrum for this AR(2) model shown in Figure 10.21, its nature might not be obvious from the empirical spectrum alone.

A fuller perspective on the spectrum in Figure 10.22 is gained when the dashed curve is provided to guide the eye. This is the spectrum for an AR(2) model fitted to the same data points from which the empirical spectrum was computed. The first two sample autocorrelations for these data are $r_1 = 0.624$ and $r_2 = -0.019$, which are near the true generating-process values that would be obtained from Equation 10.33. Using Equation 10.29, the corresponding estimated autoregressive parameters are $\phi_1 = 1.04$ and $\phi_2 = -0.67$. The sample variance of the $n = 50$ data points is 1.69, which leads through Equation 10.30 to the estimated white-noise variance $\sigma_\varepsilon^2 = 1.10$. The resulting spectrum, according to Equation 10.78, is plotted as the dashed curve. ◇

10.5.6. Sampling Properties of Spectral Estimates

Since the data from which empirical spectra are computed are subject to sampling fluctuations, Fourier coefficients computed from these data will exhibit random batch-to-batch variations as well. That is, different data batches of size n from the same source will transform to somewhat different C_k^2 values, resulting in somewhat different sample spectra.

Each squared amplitude is an unbiased estimator of the true spectral density, which means that averaged over many batches the mean of the many C_k^2 values at any frequency would closely approximate their true generating-process counterpart. Another favorable property of raw sample spectra is that the periodogram estimates at different frequencies are uncorrelated with each other. Unfortunately, the sampling distribution for an individual C_k^2 is rather broad. In particular, the sampling distribution of suitably scaled squared amplitudes is the χ^2 distribution with $\nu = 2$ degrees of freedom, which is an exponential distribution, or a gamma distribution having $\alpha = 1$ (compare Figure 4.9).

The particular scaling of the raw spectral estimates that has this χ^2 sampling distribution is

$$\frac{\nu C_k^2}{S(f_k)} \sim \chi_\nu^2, \tag{10.79}$$

where $S(f_k)$ is the (unknown, true) spectral density being estimated by C_k^2, and $\nu = 2$ degrees of freedom for a single spectral estimate C_k^2. Note that the various choices that can be made for multiplicative scaling of periodogram estimates will cancel in the ratio on the left-hand side of Equation 10.79. One way of appreciating the appropriateness of the χ^2 sampling distribution is to realize that the Fourier amplitudes in Equation 10.64 will be approximately Gaussian-distributed according to the Central Limit Theorem, because they are each derived from sums of n terms. Each squared amplitude C_k^2 is the sum of the squares of its respective pair of amplitudes A_k^2 and B_k^2, and the χ^2 is the distribution of the sum of ν squared independent standard Gaussian variates (cf. Section 4.4.5). Because the sampling distributions of the squared Fourier amplitudes in Equation 10.65a are not standard Gaussian, the scaling constants in Equation 10.79 are necessary to produce a χ^2 distribution.

Because the sampling distribution of an individual periodogram estimate is exponential, these estimates are strongly positively skewed, and their standard errors (standard deviation of the sampling distribution) are equal to their means. An unhappy consequence of these properties is that the individual C_k^2 estimates represent the true spectrum rather poorly. This very erratic nature of raw spectral estimates is illustrated by the two sample spectra shown in Figure 10.23. The heavy and light lines are two sample spectra computed from different batches of $n = 30$ independent Gaussian random variables. Each of the two sample spectra varies rather wildly around the true spectrum, which is shown by the dashed horizontal line. In a real application, the true spectrum is, of course, not known in advance, and Figure 10.23 shows that the poor sampling properties of the individual spectral estimates can make it very difficult to discern much about the true spectrum if only a single sample spectrum is available.

Confidence limits for the underlying population quantities corresponding to raw spectral estimates are rather broad. Equation 10.79 implies that confidence interval widths are proportional to the raw periodogram estimates themselves, so that

$$\Pr\left[\frac{\nu C_k^2}{\chi_\nu^2\,(1-\alpha/2)} < S(f_k) \le \frac{\nu C_k^2}{\chi_\nu^2\,(\alpha/2)}\right] = 1-\alpha, \tag{10.80}$$

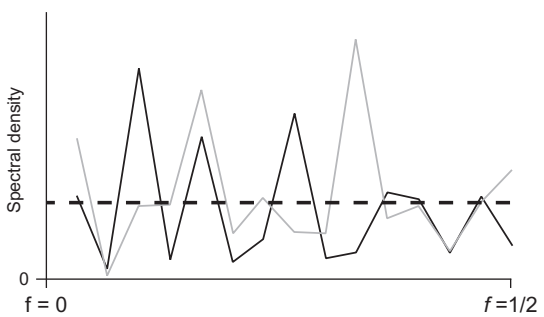

FIGURE 10.23 Illustration of the erratic sampling characteristics of estimated spectra. The solid and gray curves are two sample spectra, each computed using different batches of $n = 30$ independent Gaussian random variables. Both are quite erratic, with points of relative agreement being more fortuitous than meaningful. The true spectrum for the underlying serially independent data is shown by the horizontal dashed line. The vertical axis is linear.

where again $\nu = 2$ for a single raw periodogram estimate, and $\chi^2_\nu(\alpha)$ is the α quantile of the appropriate χ^2 distribution. For example, $\alpha = 0.05$ for a 95% confidence interval. The form of Equation 10.80 suggests one reason that it can be convenient to plot spectra on a logarithmic scale, since in that case the widths of the (1–α)•100% confidence intervals are constant across frequencies, regardless of the magnitudes of the estimated C_k^2.

The usual remedy in statistics for an unacceptably broad sampling distribution is to increase the sample size. For spectra, however, simply increasing the sample size does not give more precise information about any of the individual frequencies, but rather results in equally imprecise information about more frequencies. For example, the spectra in Figure 10.23 were computed from $n = 30$ data points, and thus consist of $n/2 = 15$ squared amplitudes. Doubling the sample size to $n = 60$ data values would result in a spectrum at $n/2 = 30$ frequencies, each point of which would exhibit the same large sampling variations as the individual C_k^2 values in Figure 10.23.

It is possible, however, to use larger data samples to obtain sample spectra that are more representative of the underlying population spectra. One approach is to compute replicate spectra from separate sections of the time series and then to average the resulting squared amplitudes. In the context of Figure 10.23, for example, a time series of $n = 60$ could be split into two series of length $n = 30$. The two spectra in Figure 10.23 might be viewed as having resulted from such a process. Here averaging each of the $n/2 = 15$ pairs of C_k^2 values would result in a less erratic spectrum, that somewhat more faithfully represents the true spectrum. In fact the sampling distributions of each of these $n/2$ average spectral values would be proportional (Equation 10.79) to the χ^2 distribution with $\nu = 4$, or a gamma distribution with $\alpha = 2$, as each would be proportional to the sum of four squared Fourier amplitudes. This distribution is substantially less variable and less strongly skewed than the exponential distribution, having standard deviation of $1/\sqrt{2}$ of the averaged estimates, or about 70% of the previous individual ($\nu = 2$) estimates. If we had a data series with $n = 300$ points, 10 sample spectra could be computed whose average would smooth out a large fraction of the sampling variability evident in Figure 10.23. The sampling distribution for the averaged squared amplitudes in that case would have $\nu = 20$. The standard deviations for these averages would be smaller by the factor $1/\sqrt{10}$, or about one-third of the magnitudes of those for single squared amplitudes. Since the confidence interval widths are still proportional to the estimated squared amplitudes, a logarithmic vertical scale again results in plotted confidence interval widths not depending on frequency. A drawback to this procedure is that the lowest frequencies are no longer resolved.

An essentially equivalent approach to obtaining a smoother and more representative spectrum using more data begins with computation of the discrete Fourier transform for the longer data series. Although this results at first in more spectral estimates that are equally variable, their sampling variability can be smoothed out by adding (not averaging) the squared amplitudes for groups of adjacent frequencies. The spectrum shown in Figure 10.17 has been smoothed in this way. For example, if we wanted to estimate the spectrum at the 15 frequencies plotted in Figure 10.23, these could be obtained by summing consecutive pairs of the 30 squared amplitudes obtained from the spectrum of a data record that was $n = 60$ observations long. If $n = 300$ observations were available, the spectrum at these same 15 frequencies could be estimated by adding the squared amplitudes for groups of 10 of the $n/2 = 150$ original frequencies. Here again the sampling distribution is χ^2 with ν equal to twice the number of pooled frequencies, or gamma with α equal to the number of pooled frequencies.

A variety of more sophisticated smoothing functions are commonly applied to sample spectra (e.g., Ghil et al., 2002; Jenkins and Watts, 1968; Von Storch and Zwiers, 1999). For example, a common approach to smoothing the sampling variability in a raw spectrum is to construct running weighted averages of the raw values,

$$\overline{C}_k^2 = \sum_{j=k-(m-1)/2}^{k+(m-1)/2} w_{j-k} C_j^2. \tag{10.81}$$

Here the averaging window covers an odd number m of adjacent frequencies; and the weights w, the indices for which run from $-(m-1)/2$ to $(m-1)/2$ are nonnegative, symmetric about the zero index, and sum to one. For example, running means of successive triplets would be achieved using Equation 10.81 with $m = 3$, and $w_{-1} = w_0 = w_1 = 1/3$. A $m = 3$-point triangular filter could be constructed using $w_{-1} = 1/4$, $w_0 = 1/2$, and $w_1 = 1/4$. Equation 10.80 for the confidence interval of one of these smoothed spectral estimates again applies, with degrees of freedom given by

$$\nu = \frac{2}{\sum_{j=-(m-1)/2}^{(m-1)/2} w_j^2}, \tag{10.82}$$

so, for example, $\nu = 6$ for the three-point running mean with equal weights.

Unfortunately, spectra smoothed using overlapping averaging windows yields individual spectral estimates that are no longer statistically independent of each other. Note also that, regardless of the specific form of the smoothing procedure, the increased smoothness and representativeness of the resulting spectra come at the expense of decreased frequency resolution and introduction of bias. Essentially, stability of the sampling distributions of the spectral estimates is obtained by smearing spectral information from a range of frequencies across a frequency band. Smoothing across broader bands produces less erratic estimates, but hides sharp contributions that may be made at particular frequencies. In practice, there is always a compromise to be made between sampling stability and frequency resolution, which is resolved as a matter of subjective judgment.

It is sometimes of interest to investigate whether the largest C_k^2 among K such squared amplitudes is significantly different from a hypothesized population value. That is, has the largest periodogram estimate plausibly resulted from sampling variations in the Fourier transform of data arising from a purely random process, or does it reflect a real periodicity that may be partially hidden by random noise in the time series? Addressing this question is complicated by two issues: choosing a null spectrum that is appropriate to the data series, and accounting for test multiplicity if the frequency f_k corresponding to the largest C_k^2 is chosen according to the test data rather than on the basis of external, prior information.

Initially we might adopt the white-noise spectrum (Equation 10.77, with $\phi = 0$) to define the null hypothesis. This could be an appropriate choice if there is little or no prior information about the nature of the data series, or if we expect in advance that the possible periodic signal is embedded in uncorrelated noise. However, most atmospheric time series are positively autocorrelated, and usually a null spectrum reflecting this tendency is a preferable null reference function (Gilman et al., 1963). Commonly it is the AR(1) spectrum (Equation 10.77) that is chosen for the purpose, with ϕ and σ_ε^2 fit to the data whose spectrum is being investigated. Using Equation 10.79, the null hypothesis that the squared amplitude C_k^2 at frequency f_k is significantly larger than the null (possibly red-noise) spectrum at that frequency, $S_0(f_k)$, would be rejected at the α level if

$$C_k^2 \geq \frac{S_0(f_k)}{\nu} \chi_\nu^2 (1-\alpha), \tag{10.83}$$

where $\chi_\nu^2 (1-\alpha)$ denotes right-tail quantiles of the appropriate chi-square distribution, given in Table B.3. The parameter ν may be greater than 2 if spectral smoothing has been employed.

The rejection rule given in Equation 10.83 is appropriate if the frequency f_k being tested has been chosen on the basis of prior, or external information, so that the choice is in no way dependent on the data used to calculate the C_k^2. When such prior information is lacking, testing the statistical significance of the largest squared amplitude is complicated by the problem of test multiplicity. When, in effect, K independent hypothesis tests are conducted in the search for the most significant squared amplitude, direct application of Equation 10.83 results in a test that is substantially less stringent than the nominal level, α. If the K spectral estimates being tested are uncorrelated, dealing with this multiplicity problem is reasonably straightforward, and involves choosing a nominal test level small enough that Equation 10.83 specifies the correct rejection rule when applied to the *largest* of the K squared amplitudes. Walker (1914) derived the appropriate level for computing the exact values,

$$\alpha = 1 - (1 - \alpha*)^{1/K}, \tag{10.84}$$

use of which is called the *Walker test* (Katz, 2002 provides more historical context). The derivation of Equation 10.84 is based on the sampling distribution of the smallest of K independent p values (Wilks, 2006a). The resulting individual test levels, α, to be used in Equation 10.83 to yield a true probability α^* of falsely rejecting the null hypothesis that the largest of K periodogram estimates is significantly larger than the null spectral density at that frequency, are closely approximated by those calculated using the Bonferroni method (Section 12.5.3),

$$\alpha = \alpha^*/K. \tag{10.85}$$

In order to account for the test multiplicity, Equation 10.85 chooses a nominal test level α that is smaller than the actual test level α^*, and that reduction is proportional to the number of frequencies (i.e., independent tests) being considered. Alternatively, the FDR approach (Section 5.4.2) could be used for this purpose. In either case the result is that a relatively large C_k^2 is required in order to reject the null hypothesis in the properly reformulated test.

Example 10.15. Statistical Significance of the Largest Spectral Peak Relative to a Red-Noise H_0

Imagine a hypothetical time series of length $n = 200$ for which the sample estimates of the lag-1 autocorrelation and white-noise variance are $r_1 = 0.6$ and $s_e^2 = 1$, respectively. A reasonable candidate to describe the behavior of these data as a purely random series could be the AR(1) process with these two parameters. Substituting these values into Equation 10.77 yields the spectrum for this process, shown as the heavy curve in Figure 10.24. A sample spectrum, C_k^2, $k = 1, \ldots, 100$, could also be computed from this series. This spectrum would include squared amplitudes at $K = 100$ frequencies because $n = 200$ data points have been Fourier transformed. The sample spectrum will be rather erratic whether or not the series also contains one or more periodic components, and it may be of interest to calculate how large the largest C_k^2 must be in order to infer that it is significantly different from the null red spectrum at that frequency. Equation 10.83 provides the decision criterion.

Because $K = 100$ frequencies are being searched for the largest squared amplitude, the standard of proof must be much more stringent than if a particular single frequency had been chosen for testing in advance of seeing the data. In particular, Equation 10.83 and Equation 10.85 both show that a test at the $\alpha^* = 0.10$ level requires that the largest of the 100 squared amplitudes trigger a test rejection at the nominal $\alpha = 0.10/100 = 0.001$ level, and a test at the $\alpha^* = 0.01$ level requires the nominal test level $\alpha = 0.01/100 = 0.0001$. Each squared amplitude in the unsmoothed sample spectrum follows a χ^2 distribution with $\nu = 2$ degrees of freedom, so the relevant right-tail quantiles $\chi_2^2(1–\alpha)$ from the second line of Table B.3 are $\chi_2^2(0.999) = 13.816$ and $\chi_2^2(0.9999) = 18.421$, respectively. (Because $\nu = 2$ these limits

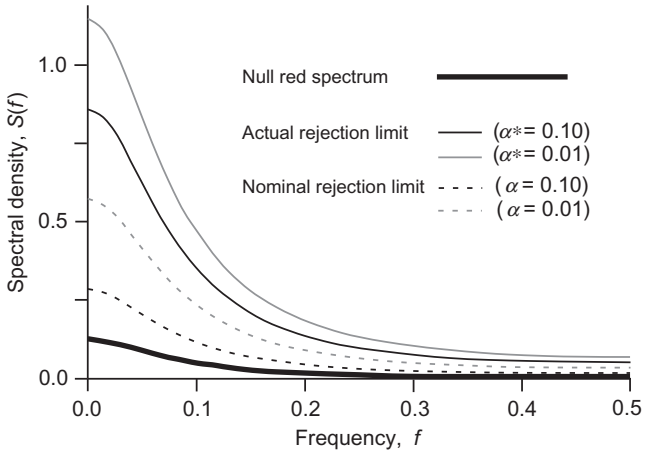

FIGURE 10.24 Red spectrum for $\phi_1 = 0.6$, $\sigma_\varepsilon^2 = 1.0$, and $n = 200$ (heavy curve) with minimum values necessary to conclude that the largest of $K = 100$ periodogram estimates is significantly larger (lighter solid curves) at the 0.10 (black) and 0.01 (gray) levels. Dashed curves show erroneous minimum values resulting when test multiplicity is not accounted for.

can also be calculated using the quantile function for the exponential distribution, Equation 4.94, with $\beta = 2$.) Substituting these values into Equation 10.83, and using Equation 10.77 with $\phi_1 = 0.6$ and $\sigma_\varepsilon^2 = 1$ to define $S_0(f_k)$, yields the two light solid lines in Figure 10.24. If the largest of the $K = 100$ C_k^2 values does not rise above these curves, the null hypothesis that the series arose from a purely random AR(1) process cannot be rejected at the specified α^* levels.

The dashed curves in Figure 10.24 are the rejection limits computed in the same way as the solid curves, except that the nominal test levels α have been taken to be equal to the overall test levels α^*, so that $\chi^2_2(0.90) = 4.605$ and $\chi^2_2(0.99) = 9.210$ have been used in Equation 10.83. These dashed curves would be appropriate thresholds for rejecting the null hypothesis that the estimated spectrum, at a single frequency that had been chosen in advance without reference to the data being tested, had resulted from sampling variations in the null red-noise process. If these thresholds were to be used to evaluate the largest among $K = 100$ squared amplitudes, the probabilities according to Equation 10.84 of falsely rejecting the null hypothesis if it were true would be $\alpha^* = 0.634$ and $\alpha^* = 0.99997$ (i.e., virtual certainty), at the nominal $\alpha = 0.01$ and $\alpha = 0.10$ levels, respectively.

Choice of the null spectrum can also have a large effect on the test results. If instead a white spectrum—Equation 10.77, with $\phi = 0$, implying $\sigma_x^2 = 1.5625$ (cf. Equation 10.21)—had been chosen as the baseline against which to judge potentially significant squared amplitudes, the null spectrum in Equation 10.83 would have been $S_0(f_k) = 0.031$ for all frequencies. In that case, the rejection limits would be parallel horizontal lines with magnitudes comparable to those at $f = 0.15$ in Figure 10.24. ◇

10.6. TIME-FREQUENCY ANALYSES

Although spectral analysis can be very useful for characterizing the important timescales of variability of a data series, the fact that it does not use the phase information from the Fourier transform of the data implies that the locations of these variations in time cannot be represented. For example, consider the two artificial time series in Figure 10.25. Figure 10.25a shows the function $[\cos(2\pi\ 11t/n) + \cos(2\pi\ 19t/n)]/2$, which is superposition of the 11th and 19th harmonics for the full data record. Figure 10.25b plots $\cos(2\pi$

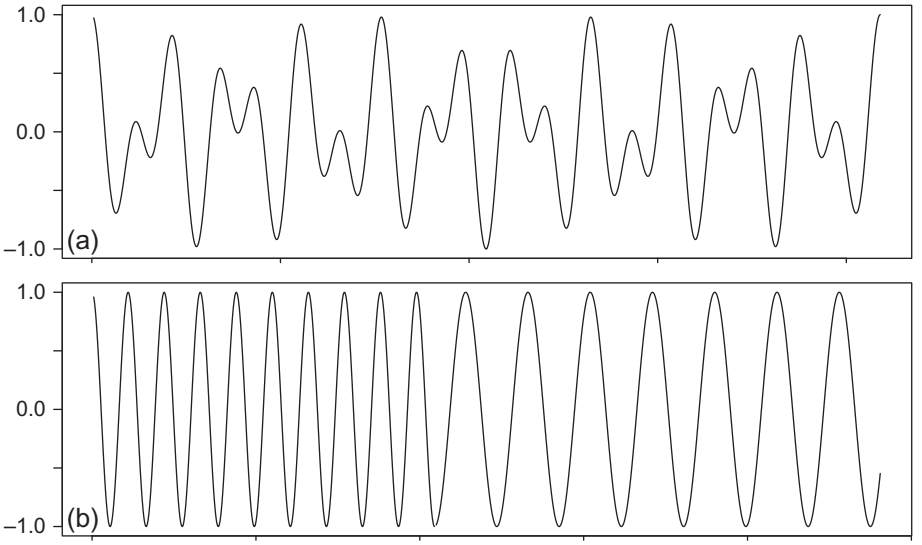

FIGURE 10.25 (a) The function $[\cos(2\pi\ 11t/n) + \cos(2\pi\ 19t/n)]/2$, and (b) $\cos(2\pi\ 19t/n)$] for the first half of the time series and $\cos(2\pi\ 11t/n)$ for the last half. Fourier spectra computed from the two series would be identical.

$19t/n$) for the first $n/2$ points of the time series and $\cos(2\pi\ 11t/n)$ for the last $n/2$. The power spectra corresponding to these two series are identical.

10.6.1. Windowed Fourier Transforms

Windowed Fourier transforms offer one approach to time localization of spectral features. This method involves choosing a segment length $l < n$, and computing spectra for consecutive time-series subsets of length l as the "window" is moved along the series. That is, the first spectrum computed would be for the time points $t = 1$ to l, the second would be for $t = 2$ to $l + 1$, and so on. To construct the time-frequency plot, each of these $n - l + 1$ spectra is rotated 90° counterclockwise so that the low-frequency amplitudes are at the bottom and the high-frequency amplitudes are at the top, and the individual spectra are placed consecutively in time order. The result is a two-dimensional plot in which the vertical axis is frequency and the horizontal axis is time. Horizontal nonuniformity in the plot will indicate where different frequency components are more or less important during different portions of the time series, and the overlapping nature of the consecutive windowed time series will generally yield smooth variations in the horizontal.

Figure 10.26a illustrates the result for the SOI series for the years 1876–1995, a different segment of which is shown in Figure 3.16. Here the window length l is 21 years, the initial and final years of which have been tapered to zero using a cosine shape. Consistent with the usual understanding of ENSO, maximum power is exhibited in the range of about 0.14 to 0.33 cycles/per year (periods of about 3–7 years, as indicated also in Figures 10.7 and 10.26c). However, Figure 10.26a indicates that these strongest SOI variations do not occur uniformly through the record, but rather are stronger in the first and last quarters of the 20th century, with a more quiescent period in the middle of the century.

The time resolution provided by windowed Fourier transforms comes at the cost that the method is unable to see longer-period variations. In particular, it cannot resolve variations with periods longer than the window length l. Choosing a larger l improves the low-frequency resolution, but degrades the time resolution, so that a balance between these two desirable attributes must be struck.

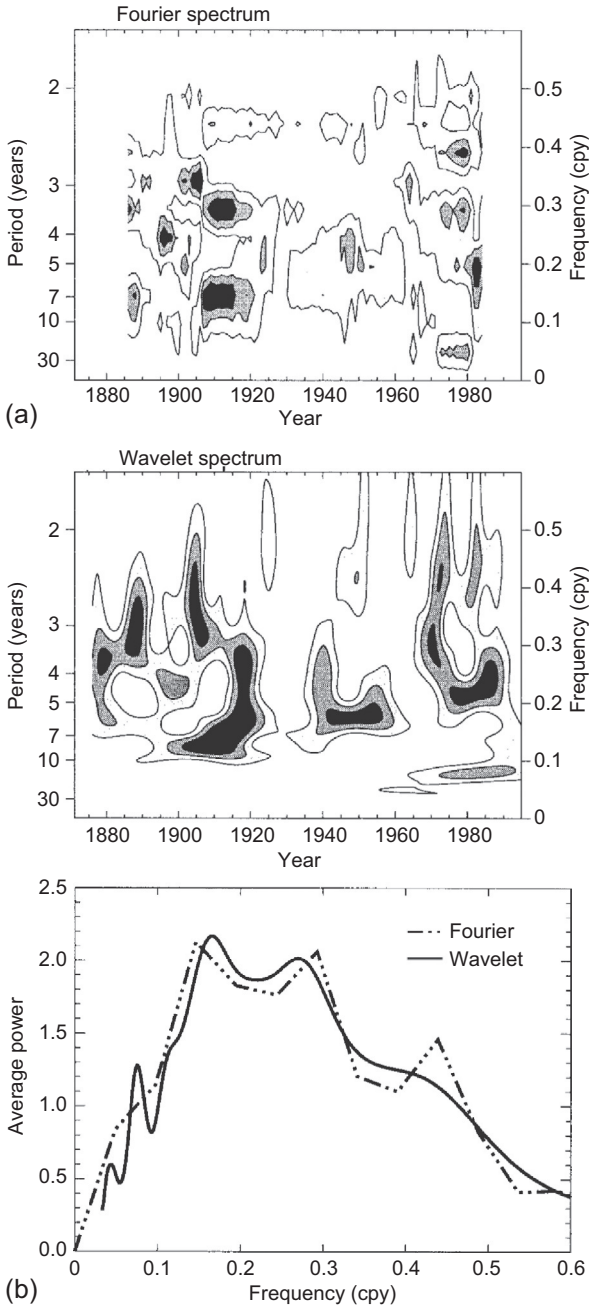

FIGURE 10.26 (a) Windowed Fourier transform, and (b) wavelet spectrum for the SOI series, 1876–1995. Contours indicate 20%, 40%, and 60% of maximum spectral power. (c) Time-integrated and standardized spectra from panel (a) (dash-dot) and (b) (solid). *From Kestin et al. (1998). © American Meteorological Society. Used with permission.*

10.6.2. Wavelets

Wavelet analysis is an alternative to windowed Fourier transforms that also yields a two-dimensional plot showing strengths of variations as a function of both period (or frequency) and time. Unlike Fourier analysis, which characterizes similarities between time series and trigonometric functions of infinite extent, wavelet analysis addresses similarities, over limited portions of the time series, to waves of limited time extent called wavelets.

A variety of choices for the basic wavelet shape are available (e.g., Torrance and Compo 1998), but all have zero mean and are nonzero over a limited portion of the real line. Often the Morlet wavelet is chosen in geophysical settings because of its similarity in shape to typical variations in those kinds of data. The Morlet wavelet consists of a sinusoidal wave that is damped for increasing absolute values of its arguments using a Gaussian shape,

$$\psi(t) = \pi^{-1/4} e^{i\omega_0 t} e^{-\left(t^2/2\right)}, \tag{10.86}$$

where often the constant $\omega_0 = 6$ is chosen. Figure 10.27 shows the shape of this wavelet function. It is nonzero for absolute values of its argument smaller than about 3.

The time localization in wavelet analysis is achieved through multiplication of the time series with a wavelet shape ψ that is centered at various points t' in the series, in effect sliding the wavelet shape along the time series and computing its similarity to that portion of the time series over which the wavelet is nonzero. Different frequencies are resolved by repeating this process for different dilations s, $s > 1$ of the basic wavelet shape, because larger values of s expand the portion of the time series for which the scaled wavelet function is nonzero, and its peaks and troughs are more fully separated. The result is a two-dimensional array of wavelet coefficients

$$W\left(s, t'\right) = \sum_{t=1}^{n} y_t \frac{1}{\sqrt{s}} \psi^* \left(\frac{t - t'}{s}\right), \tag{10.87}$$

where the asterisk denotes complex conjugate. Although the necessary computations can be carried out using Equation 10.87, more efficient algorithms are available (e.g., Torrence and Compo, 1998). Figure 10.26b shows the resulting wavelet spectrum for the 1876-1995 SOI series, using the Morlet wavelet. Here the coefficients $W(s, t')$ have been plotted in two dimensions and smoothed. The basic features portrayed by the windowed Fourier transform of these data in Figure 10.26a are also present in their wavelet spectrum.

Wavelet analysis can be generalized to representing the variations in two-dimensional images in terms of a spectrum of spatial scales. One application of two-dimensional wavelets is in spatial field verification (Section 9.8.5). Another is in in the JPEG format for photographic image compression.

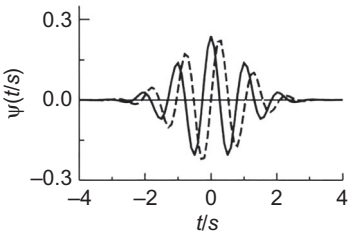

FIGURE 10.27 Basic shape of the Morlet wavelet, with real part solid and imaginary part dashed. *From Torrence and Compo (1998). © American Meteorological Society. Used with permission.*

Both the windowed Fourier transforms and the wavelet spectra would be different for the two time series in Figure 10.25, and both analysis methods would allow the different characteristics of the two time series to be diagnosed. For the superposition of the two underlying waves in Figure 10.25a, both approaches would yield horizontal bands at the frequencies $\omega_{11} = 2\pi\ 11/n$ and $\omega_{19} = 2\pi\ 19/n$, that would extend across the full widths of the diagrams. For the series in Figure 10.25b, both would exhibit horizontal bands at $\omega_{19} = 2\pi\ 19/n$ across the left halves of the diagrams, and separate horizontal bands at $\omega_{11} = 2\pi\ 11/n$ extending across the right halves.

10.7. EXERCISES

10.1. Using the January 1987 precipitation data for Canandaigua in Table A.1,
   a. Fit a two-state, first-order Markov chain to represent daily precipitation occurrence.
   b. Test whether this Markov model provides a significantly better representation of the data than does the assumption of independence.
   c. Compare the theoretical stationary probability, π_1 with the empirical relative frequency.
   d. Graph the theoretical autocorrelation function, for the first three lags.
   e. Compute the probability according to the Markov model that a sequence of consecutive wet days will last at least three days.

10.2. Graph the autocorrelation functions up to five lags for
   a. The AR(1) process with $\phi = 0.4$.
   b. The AR(2) process with $\phi_1 = 0.7$ and $\phi_2 = -0.7$.

10.3. Computing sample lag correlations for a time series with $n = 100$ values, whose variance is 100, yields $r_1 = 0.80$, $r_2 = 0.60$, and $r_3 = 0.50$.
   a. Use the Yule–Walker equations to fit AR(1), AR(2), and AR(3) models to the data. Assume the sample size is large enough that Equation 10.26 provides a good estimate for the white-noise variance.
   b. Select the best autoregressive model for the series according to the BIC statistic.
   c. Select the best autoregressive model for the series according to the AIC statistic.

10.4. Given that the mean of the time series in Exercise 10.3 is 50, use the fitted AR(2) model to forecast the future values of the time series x_1, x_2, and x_3; assuming the current value is $x_0 = 76$ and the previous value is $x_{-1} = 65$.

10.5. The variance of a time series governed by the AR(1) model with $\phi = 0.8$, is 25. Compute the variances of the sampling distributions of averages of consecutive values of this time series, with lengths
   a. $n = 5$,
   b. $n = 10$,
   c. $n = 50$.

10.6. A square-root transformed time series of hourly wind speeds is well described using an AR(2) model with parameters $\phi_1 = 1.09$ and $\phi_2 = -0.21$, and having Gaussian residuals with variance $\sigma_\varepsilon^2 = 0.12\,\text{m/s}$. The mean of the transformed series is 2.3 $(\text{m/s})^{1/2}$.
   a. Calculate the variance of the transformed wind speeds.
   b. If the wind speed in the current hour is 15 m/s and the wind speed for the previous hour was 10 m/s, calculate a 50% central prediction interval for the wind speed (m/s) in the coming hour.

10.7. For the temperature data in Table 10.7,
   a. Calculate the first two harmonics.
   b. Plot each of the two harmonics separately.
   c. Plot the function representing the annual cycle defined by the first two harmonics. Also include the original data points in this plot, and visually compare the goodness of fit.

TABLE 10.7 Average Monthly Temperature Data for New Delhi, India

Month	J	F	M	A	M	J	J	A	S	O	N	D
Average Temperature, °F	57	62	73	82	92	94	88	86	84	79	68	59

10.8. Use the two-harmonic equation for the annual cycle from Exercise 10.7 to estimate the mean daily temperatures for
   a. April 10.
   b. October 27.

10.9. The amplitudes of the third, fourth, fifth, and sixth harmonics, respectively, of the data in Table 10.7 are 1.4907, 0.5773, 0.6311, and 0.0001°F.
   a. Plot a periodogram for these data. Explain what it shows.
   b. What proportion of the variation in the monthly average temperature data is described by the first two harmonics?
   c. Which of the harmonics have squared amplitudes that are significant, relative to the white-noise null hypothesis, not accounting for test multiplicity, with $\alpha = 0.05$?
   d. Which of the harmonics have squared amplitudes that are significant, relative to the white-noise null hypothesis, using Bonferroni criterion, with $\alpha^* = 0.05$?
   e. Which of the harmonics have squared amplitudes that are significant harmonics, relative to the white-noise hull hypothesis, using FDR criterion, with $\alpha_{\text{global}} = 0.05$?

10.10. How many tick-marks for frequency are missing from the horizontal axis of Figure 10.17?

10.11. Suppose the minor peak in Figure 10.17 at $f = 13/256 = 0.0508$ mo^{-1} resulted in part from aliasing.
   a. Compute a frequency that could have produced this spurious signal in the spectrum.
   b. How often would the underlying sea-level pressure data need to be recorded and processed in order to resolve this frequency explicitly?

10.12. Plot the spectra for the two autoregressive processes in Exercise 10.2, assuming unit white-noise variance, and $n = 100$.

10.13. The largest squared amplitude in Figure 10.23 is $C_{11}^2 = 0.413$ (in the gray spectrum). Compute a 95% confidence interval for the value of the underlying spectral density at this frequency.