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 to characterize the nature of the ordering. Time-series methods are thus appropriate when the ordering of the data values in time are important to a given application.

Regarding an observed time series as having been generated by a theoretical (model) process is convenient because it allows characteristics of future, yet unobserved, values of a time series to be inferred from the inevitably limited data at hand. That is, characteristics of an observed time series are summarized by the parameters of the 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. The properties of the model-generating process can be used to infer characteristics of yet unobserved values of the series.

8.1.3 Time-Domain versus Frequency-Domain Approaches

There are two fundamental approaches to time-series analysis: time-domain analysis and frequency-domain analysis. These two approaches proceed very differently and may seem quite distinct, but they are not independent. Rather, they are complementary methods that are linked mathematically.

Time-domain methods seek to characterize data series in the same terms in which they are observed and reported. A primary tool for characterization of relationships between data values in the time-domain approach is the autocorrelation function. Mathematically, time-domain analyses operate in the same space as the data values. Separate sections in this chapter describes 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, whereas continuous random variables may take on any of the infinitely many real values within their range.

Frequency-domain analysis represents data series in terms of contributions occurring at different time scales, or characteristic frequencies. Each time scale is represented by a pair of sine and cosine functions. 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, frequency-domain analysis involves 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 is sometimes difficult to grasp. However, frequency-domain methods are very commonly applied to atmospheric time series, and important insights can be gained from frequency-domain analyses.

8.2 Time Domain. I. Discrete Data

8.2.1 Markov Chains

Recall that a discrete random variable is one that can take on values only from among a defined, finite 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 comprising 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 random variable in the time series, whereas 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 likelihoods for the system being in each of its possible states during the next time period. The most common 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, knowledge of 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

$$\Pr\{X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \ldots, X_1\} = \Pr\{X_{t+1} \mid X_t\}. \tag{8.1}$$

The probabilities of future states depend on the current state, but they do not depend on the particular way that the model system arrived at the current 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, but also that knowing yesterday's data value provides no additional information.

The transition probabilities of a Markov chain are conditional probabilities. For a first-order Markov chain, there is a conditional probability distribution pertaining to each possible current state of the system, 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 ability of a Markov chain to represent the serial correlation, or persistence, often exhibited by atmospheric variables. If the probability of the future state is 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 is not 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 tend to be higher than the probabilities of arriving at that state from other states.

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 achieve this is to specify that the probabilities vary according to some smooth periodic curve, such as a cosine function. Alternatively, separate transition probabilities can be used for nearly stationary portions of the cycle, for example four 3-month seasons or 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).

8.2.2 Two-State, First-Order Markov Chains

The simplest kind of discrete random variable pertains to the situation of 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 [Eq. (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 simple 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 shown in Table 8.1. 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 tend to cluster in time. As was illustrated in Example 2.2, this clustering is an expression of the serial dependence present in the time series. 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 common and often quite good stochastic model for data of this kind is a first-order, two-state Markov chain. A two-state Markov chain is natural for

Table 8.1
Time Series of a Dichotomous Random Variable Derived from the January 1987 Ithaca Precipitation Data in Table A.1[a]

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

[a] Days on which nonzero precipitation was reported yield $X_t = 1$; days with zero precipitation yield $X_t = 0$.

dichotomous data since each of the two states will pertain to one of the possible data values. A first-order Markov chain has the property that the transition probabilities governing each observation in the time series depends only on the value of the previous member of the time series.

Figure 8.1 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 8.1. 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 Fig. 8.1 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

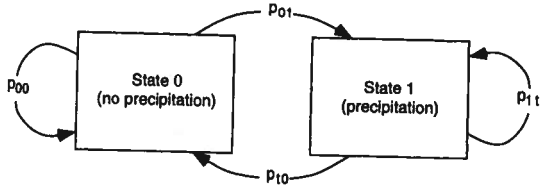

Fig. 8.1 Schematic representation of a two-state, first-order Markov chain, illustrated in terms of daily precipitation occurrence or nonoccurrence. The two states are labeled "0" for no precipitation, and "1" for precipitation occurrence. For a first-order Markov chain, there are four transition probabilities 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 such as day-to-day precipitation occurrence that exhibit positive serial correlation, $p_{10} < p_{00}$, and $p_{01} < p_{11}$.

$t + 1$ (e.g., whether precipitation will occur tomorrow) given the state at time t (e.g., whether precipitation occurred today). That is,

$$p_{00} = \Pr\{X_{t+1} = 0 \mid X_t = 0\} \tag{8.2a}$$

$$p_{01} = \Pr\{X_{t+1} = 1 \mid X_t = 0\} \tag{8.2b}$$

$$p_{10} = \Pr\{X_{t+1} = 0 \mid X_t = 1\} \tag{8.2c}$$

$$p_{11} = \Pr\{X_{t+1} = 1 \mid X_t = 1\}. \tag{8.2d}$$

Together, Eqs. (8.2a) and (8.2b) constitute the conditional probability distribution for the value of the time series at time $t + 1$, given that $X_t = 0$ at time t. Similarly, Eqs. (8.2c) and (8.2d) 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 Eqs. (8.2) 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 necessary to focus on only 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 the two pairs of conditional probabilities must sum to 1. The parameter estimation procedure consists simply of computing the conditional relative frequencies, which yield the maximum-likelihood estimators (MLEs)

$$\hat{p}_{01} = \frac{\#\text{ of 1's following 0's}}{\text{total \# of 0's}} = \frac{n_{01}}{n_{0\cdot}} \tag{8.3a}$$

and

$$\hat{p}_{11} = \frac{\#\text{ of 1's following 1's}}{\text{total \# of 1's}} = \frac{n_{11}}{n_{1\cdot}}. \tag{8.3b}$$

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, $n_{0\cdot}$ is the number of 0's in the series followed by another data point, and $n_{1\cdot}$ is the number of 1's in the series followed by another data point. That is, the subscript "$\cdot$" indicates the total over all values of the index replaced by this symbol, so that $n_{1\cdot} = n_{10} + n_{11}$ and $n_{0\cdot} = n_{00} + n_{01}$. Equations (8.3) state that the parameter p_{01} is estimated by looking at the conditional 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 that fraction of points for which $x_t = 1$ is followed by points with $x_{t+1} = 1$. The somewhat labored definitions of $n_{0\cdot}$ and $n_{1\cdot}$ are necessary to account for the "edge" effect of a finite sample size. The final point in the time series is not counted in the denominator of Eq. (8.3a) or (8.3b), 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 (8.3) 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 [Eq. (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 Chapter 4 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 binomial 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},

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}}, \tag{8.4}$$

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

The transition probabilities also imply a degree of serial correlation, or persistence, for the time series. In terms of the transition probabilities, the lag-1 autocorrelation [Eq. (3.23)] of the time series is simply

$$r_1 = p_{11} - p_{01}. \tag{8.5}$$

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. This implies 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 $r_1 = p_{11} - p_{01} = 0$, or $p_{11} = p_{01} = \pi_1$. In this case the two conditional probability distributions specified by Eq. (8.2) are the same, and the time series is simply a string of independent binomial realizations, each with $N = 1$. The binomial distribution with $N = 1$ can be viewed as a two-state, *zero*-order Markov chain.

The full autocorrelation function, Eq. (3.24), follows easily from the lag-1 correlation 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,

$$r_k = (r_1)^k. \tag{8.6}$$

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 (8.6) 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 Eq. (8.1). 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.

8.2.3 Test for Independence versus First-Order Serial Dependence

Even if a series of binary data is generated by a mechanism producing serially independent values, the sample lag-1 autocorrelation [Eq. (8.5)] computed from a finite sample is unlikely to be exactly zero. A formal test, similar to the χ^2 goodness-of-fit test [Eq. (5.18)], can be conducted 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 binomial variables with $N = 1$); the alternative is 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 situation is illustrated in Fig. 8.2, 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

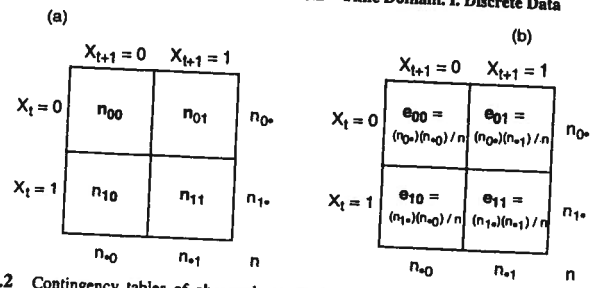

Fig. 8.2 Contingency tables of observed transition counts n_{ij} (a) for a binary time series and (*b*) transition counts e_{ij} expected if the time series actually consists of serially independent values with the same marginal totals. The transition counts are shown in boldface, and the marginal totals are in plain type.

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

$$e_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}. \tag{8.7}$$

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

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}}, \tag{8.8}$$

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 Eq. (8.8) for each of the four pairs of contingency table cells in Fig. 8.2. Note that Eq. (8.8) is analogous to Eq. (5.18), 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 the parameter $\nu = 1$ 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 Eq. (8.8) 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 tables of the χ^2 distribution, or by using Table B.2 (in Appendix B) since the χ^2 is a special case of the gamma distribution. Note that the test statistic must be transformed to the standard gamma distribution (with $\beta = 1$), as described in Chapter 5 and illustrated in the following example.

Example 8.1. Fitting a Two-State, First-Order Markov Chain

Consider summarizing the time series in Table 8.1, derived from the January 1987 Ithaca precipitation series in Table A.1, using a first-order Markov chain. The parameter estimates in Eq. (8.3) are easily obtained from the transition counts. For example, the number of 1's following 0's in the time series of Table 8.1 is $n_{01} = 5$. Similarly, $n_{00} = 11$, $n_{10} = 4$, and $n_{11} = 10$. The resulting sample estimates for the transition probabilities [Eqs. (8.3)] 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 the extra effort of fitting the first-order Markov chain to the data in Table 8.1 is justified can be investigated using the χ^2 test in Eq. (8.8). Here the null hypothesis is that these data resulted from an independent (i.e., binomial) 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 $n_{0\cdot} = 11 + 5 = 16$, $n_{1\cdot} = 4 + 10 = 14$, $n_{\cdot 0} = 11 + 4 = 15$, and $n_{\cdot 1} = 5 + 10 = 15$. 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 Eq. (8.8), one finds $\chi^2 = (11 - 8)^2/8 + (5 - 8)^2/8 + (4 - 7)^2/7 + (10 - 7)^2/7 = 4.82$. The degree of unusualness of this result with reference to the null hypothesis can be assessed with the aid of the gamma distribution probabilities in Table B.2. Recall that the χ^2 distribution is a special case of the gamma distribution, with $\alpha = \nu/2 = 0.5$, and $\beta = 2$. Transforming to the standard (i.e., $\beta = 1$) gamma distribution using $\xi = \chi^2/\beta = 4.82/2 = 2.41$, one finds that the result lies between the 95th and 99th percentiles of the standard gamma distribution with $\alpha = 0.5$. 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 persistence exhibited by this data sample can be summarized using the persistence parameter, which is also the lag-1 autocorrelation, $r_1 = p_{11} - p_{01} = 0.714 - 0.312 = 0.402$. This is a fairly large value, indicative of substantial serial correlation in the time series. It also implies the full autocorrelation function, through Eq. (8.6). Figure 8.3 shows that the implied theoretical correlation function for this Markov process, shown as 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.

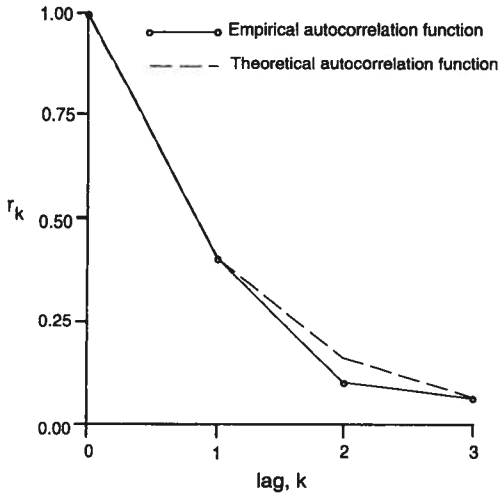

Fig. 8.3 Sample autocorrelation function for the January 1987 Ithaca binary precipitation occurrence series, Table 8.1 (solid, with circles), and theoretical autocorrelation function (dashed) specified by the fitted first-order Markov chain model [Eq. (8.6)]. The correlations are 1.00 for $k = 0$, since the unlagged data are perfectly correlated with themselves.

Finally, the stationary (i.e., climatological) probability for precipitation implied for this data by the Markov chain model is, using Eq. (8.4), $\pi_1 = 0.312/(1 + 0.312 - 0.714) = 0.522$. This value agrees closely with the relative frequency $16/30 = 0.533$, obtained by counting the number of 1's in the last thirty values of the series in Table 8.1. ¤

8.2.4 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 generated using the Markov chain as an algorithm. To generate sequences of numbers statistically resembling those in Table 8.1, for example, the parameters $p_{01} = 0.312$ and $p_{11} = 0.714$, estimated in Example 8.1, would be used together with a uniform [0,1] random-number generator. The uniform random-number generator is a computer algorithm producing numbers in the interval [0,1] with equal probability. 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, each new uni-

form random number is 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 precipitation data, simulated wet days are more likely to follow wet days than dry days, as occurs 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). Commonly a gamma distribution (Chapter 4) is fit to the precipitation amounts on wet days in the data record (e.g., Katz, 1977; Richardson, 1981; Stern and Coe, 1984), although other distributions can also be used (Woolhiser and Roldan, 1982). Computer algorithms are available to generate random variables drawn from gamma distributions (e.g., Bratley *et al.*, 1983; Johnson, 1987), and these can be used to simulate 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 those instances where it has been investigated (Katz, 1977; Stern and Coe, 1984). 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).

Inferences concerning properties of longer-term precipitation quantities (e.g., the monthly frequency distributions of numbers of wet days in a month, or of total monthly precipitation) can be drawn 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 occurrence and amount. 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).

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 (Chapter 7). Usually one of two reference forecasts is used: (1) the climatological probability of the forecast event, in this case π_1; or (2) 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 on, 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}, whereas the conditional forecast probability for precipitation following a day with precipitation would be p_{11}. Note that for meteorological 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 persistence (either 0 or 1) and climatology (π_1). 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 a strong serial correlation will produce transition probabilities much closer to 0 and 1.

8.2.5 Multistate 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 Fig. 8.4. Here the three states are arbitrarily labeled 1, 2, and 3. At each time t, the random variable in the series

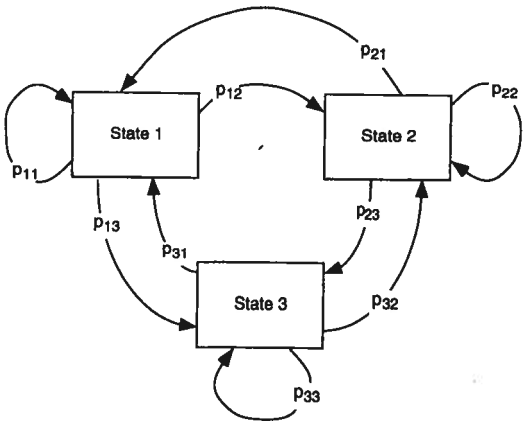

Fig. 8.4 Schematic illustration of a three-state, first-order Markov chain. There are nine possible transitions among the three states, including the possibility that two consecutive points in the time series will be in the same state. First-order time dependence implies that the transition probabilities depend only on the current state of the system, or present value of the time series.

can take on one of the three values $X_t = 1$, $X_t = 2$, or $X_t = 3$, and each of these values corresponds to a 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 $3^2 = 9$ transition probabilities, p_{ij}. In general, for a first-order, k-state Markov chain, there are k^2 transition probabilities.

As is the case for the two-state Markov chain discussed above, the transition probabilities for multiple-state Markov chains are conditional probabilities. For example, the transition probability p_{12} in Fig. 8.4 is the conditional probability that state 2 will occur at time $t + 1$, given that state 1 occurred at time t. Therefore, in a k-state Markov chain it must be the case that the probabilities for the k 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 Eqs. (8.3) for two-state chains. Each of these estimates is simply obtained from the conditional relative frequencies of the transition counts:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i\cdot}}; \qquad i, j = 1, \ldots, s. \tag{8.9}$$

As before, the dot indicates summation over all values of the replaced subscript so that, for example, $n_{1\cdot} = \Sigma_j n_{1j}$. For the $s = 3$-state Markov chain represented in Fig. 8.4, for example, $\hat{p}_{12} = n_{12}/(n_{11} + n_{12} + n_{13})$. In general, a contingency table of transition counts, corresponding to Fig. 8.2a 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 Eq. (8.8). 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 Eq. (8.7). Under the null hypothesis of no serial correlation, the distribution of the test statistic in Eq. (8.8) 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 for temperature and precipitation, as defined by the U.S. Climate Analysis Center (Examples 4.6 and 4.8), 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.

8.2.6 Higher-Order Markov Chains

First-order Markov chains often provide good representations of daily precipitation occurrence; however, it is not obvious just from inspection of the series in Table 8.1, for example, that this simple model will be adequate to capture the exhibited correlation structure. More generally, an m-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 Eq. (8.1) to the m-order Markov chain is

$$\Pr\{X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \ldots, X_1\} = \Pr\{X_{t+1} \mid X_t, X_{t-1}, \ldots, X_{t-m}\}. \tag{8.10}$$

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 denoting the state at time $t - 1$, the second denoting the state in time t, and the third specifying the state at (the future) time $t + 1$. The notation for the transition probabilities of a second-order Markov chain can be defined as

$$p_{hij} = \{X_{t+1} = j \mid X_t = i, X_{t-1} = h\}. \tag{8.11}$$

In general, the notation for a m-order Markov chain requires $m + 1$ subscripts on the transition counts and transition probabilities. If Eq. (8.11) is being applied to a binary time series such as that in Table 8.1, 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, Eq. (8.11) 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, m-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 Fig. 8.2a, is shown in Table 8.2 for a $s = 2$ state, $m =$ 2-order Markov chain. The transition counts are determined from the observed

Table 8.2
Arrangement of the $2^{2+1} = 8$ Transition Counts for a Two-State, Second-Order Markov Chain in a Table of the Form of Fig. 8.2a[a]

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}	$n_{11\cdot} = n_{110} + n_{111}$

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

data series by examining consecutive groups of $m + 1$ data points. For example, the first three data points in Table 8.1 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 8.1 exhibits three transitions of this kind, so the final transition count $n_{011} = 3$ for this data set. The second triplet in the data set in Table 8.1 would contribute one count to n_{110}. There is only one other triplet in this data for which $X_{t-1} = 1$, $X_t = 1$, $X_{t+1} = 0$, so 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

$$\hat{p}_{hij} = \frac{n_{hij}}{n_{hi\cdot}}. \tag{8.12}$$

That is, given that 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 future value of the time series $X_{t+1} = j$ is p_{hij}, and the sample estimate of this probability is given in Eq. (8.12). Just as the two-state first-order Markov chain consists essentially of two conditional binomial distributions, a two-state second-order Markov chain amounts to four conditional binomial distributions, with parameters $N = 1$ and $p = p_{hi1}$, for each of the four distinct combinations of the indices h and i.

Note that the small data set in Table 8.1 is really too short to fit a second-order Markov chain. Since there are no triplets in this series for which $X_{t-1} = 1$, $X_t = 0$, $X_{t+1} = 1$ (i.e., a single dry day following and followed by a wet day) the transition count $n_{101} = 0$. This would lead to the sample estimate for the transition probability $\hat{p}_{101} = 0$, even though there is no physical reason why this particular sequence of wet and dry days should not occur.

8.2.7 Deciding among Alternative Orders of Markov Chains

How is one to know what order m is appropriate for a Markov chain to represent a particular data series? One approach is to use a hypothesis test. For example, the χ^2 test in Eq. (8.8) can be used to assess the significance of a first-order Markov chain model versus a zero-order, or binomial model (Example 8.1). While 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, the overall significance of a collection of such tests would be difficult to evaluate. This difficulty arises in part because of the issue multiplicity. As discussed in the context of the topic of field significance in Chapter 5, the overall significance level of a collection of simultaneous, correlated tests is difficult to evaluate.

Two criteria are in common use for choosing among alternative orders of Markov chain models: 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 for the transition

probabilities of the fitted Markov chains. These log-likelihoods depend on tl transition counts and the estimated transition probabilities. The log-likelihoods f Markov chains of order 0, 1, 2, and 3 are

$$L_0 = \sum_{j=0}^{s-1} n_j \ln(\hat{p}_j) \tag{8.13}$$

$$L_1 = \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} n_{ij} \ln(\hat{p}_{ij}) \tag{8.13}$$

$$L_2 = \sum_{h=0}^{s-1} \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} n_{hij} \ln(\hat{p}_{hij}) \tag{8.13}$$

and

$$L_3 = \sum_{g=0}^{s-1} \sum_{h=0}^{s-1} \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} n_{ghij} \ln(\hat{p}_{ghij}), \tag{8.13}$$

with obvious extension for fourth-order and higher Markov chains. Here the sun mations are over all s states of the Markov chain, and so will include only tw terms each for two-state (binary) time series. Equation (8.13a) is simply the lo likelihood for the independent binomial model.

Example 8.2. Likelihood Ratio Test for the Order of a Markov Chain

To illustrate the application of Eqs. (8.13), consider a likelihood ratio test first-order dependence of the time series in Table 8.1, versus the null hypothes of zero serial correlation. The test involves computation of the log-likelihoods Eqs. (8.13a) and (8.13b). The resulting two log-likelihoods are compared usin the test statistic given by Eq. (5.19).

In the last 30 data points in Table 8.1, there are $n_0 = 14$ zeros and $n_1 = 1$ ones, yielding the unconditional relative frequencies of rain and no rain $\hat{p}_0 = 14/30 = 0.467$ and $\hat{p}_1 = 16/30 = 0.533$, respectively. The last 30 points are use because the first-order Markov chain amounts to two conditional binomial distr butions, given the previous day's value, and the value for December 31, 1986 not available in Table A.1. The log-likelihood in Eq. (8.13a) for these data is $L_0 = 14 \ln(0.467) + 16 \ln(0.533) = -20.73$. Values of n_{ij} and $\hat{p}_{ij}$ were compute previously, and can be substituted into Eq. (8.13b) to yield $L_1 = 11 \ln(0.688) + 5 \ln(0.312) + 4 \ln(0.286) + 10 \ln(0.714) = -18.31$. Necessarily, $L_1 \geq L_0$ be cause the greater number of parameters in the more elaborate first-order Marko chain model provide more flexibility for a closer fit to the data at hand. The stati tical significance of the difference in log-likelihoods can be assessed since it known that the null distribution of $\Lambda = 2(L_1 - L_0) = 4.83$ is χ^2, with $\nu = [s^{m(H_A)} - s^{m(H_0)}](s - 1)$ degrees of freedom. Since the time series being teste in binary, $s = 2$. The null hypothesis is that the time dependence is zero-order, s

$m(H_0) = 0$, 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. This likelihood test result is consistent with the χ^2 goodness-of-fit test conducted in Example 8.1. ¤

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 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 then computed for each trial order m, using

$$\mathrm{AIC}(m) = -2L_m + 2s^m(s - 1), \tag{8.14}$$

or

$$\mathrm{BIC}(m) = -2L_m + s^m(\ln n), \tag{8.15}$$

respectively. The order m is chosen as appropriate that minimizes either Eq. (8.14) or (8.15). 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 may be 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).

8.3 Time Domain. II. Continuous Data

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, windspeed, and geopotential height are continuous variables. The correlation structure of such time series can often be represented successfully using a class of time-series models known as *Box–Jenkins models*, after the classic text by Box and Jenkins (1976).

8.3.1 First-Order Autoregression

The simplest Box–Jenkins model is the *first-order autoregression* [AR(1)] model. It is the continuous analog of the first-order Markov chain. As the term suggests, one way of viewing the AR(1) model is as a simple linear regression (Chapter 6), 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

$$x_{t+1} - \mu = \phi(x_t - \mu) + \epsilon_{t+1}, \tag{8.16}$$

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 Eq. (8.16) 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 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 Eq. (3.5). The ϵ values are mutually independent random quantities having mean $\mu_\epsilon = 0$ and variance σ_ϵ^2. Very often it is further assumed that the ϵ values follow a Gaussian distribution.

As illustrated in Fig. 8.5, the autoregressive model in Eq. (8.16) 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. Plotted on the horizontal are the first 30 data values, for January 1–30. The corresponding temperatures for the following days, January 2–31, are plotted on the vertical. The expected serial correlation, or persistence, is evident from the appearance of

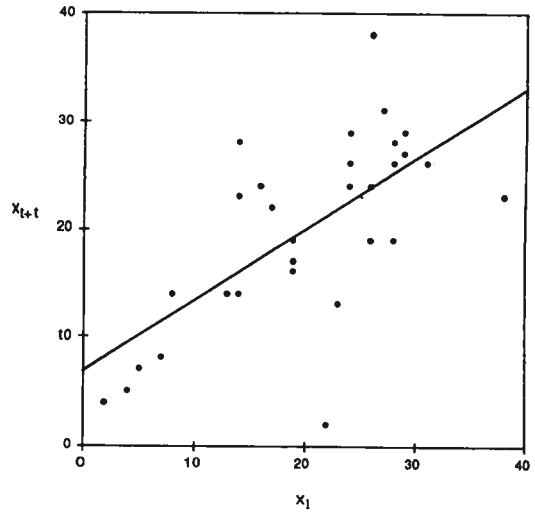

Fig. 8.5 Scatterplot of January 1–30, 1987 minimum temperatures (°F) at Canandaigua, New York (x_t, horizontal) paired with minimum temperatures for the following days, January 2–31 (x_{t+1}, vertical). The data are from Table A.1. The simple linear regression equation corresponding to the first term of the AR(1) time-series model [Eq. (8.16)] is also shown.