Chapter 8

STOCHASTIC MODELS

The deterministic dynamic models of Part II of this book do not allow for the explicit representation of uncertainty. When random effects are taken into account, the resulting model is called a *stochastic model*. Several kinds of general stochastic models are in wide use today. In this chapter we will introduce the most important and commonly used stochastic models.

8.1 Markov Chains

A Markov chain is a discrete–time stochastic model. It is a generalization of the discrete–time dynamical system model introduced in Section 4.3. Although the model is simple, the number and diversity of applications are surprisingly large. In this section we will introduce the general Markov chain model. We will also introduce a concept of steady state appropriate for a stochastic model.

Example 8.1. A pet store sells a limited number of 20–gallon aquariums. At the end of each week, the store manager takes inventory and places orders. Store policy is to order three new 20–gallon aquariums at the end of the week if all of the current inventory has been sold. If even one of the 20–gallon aquarium remains in stock, no new units are ordered. This policy is based on the observation that the store only sells an average of one of the 20–gallon aquariums per week. Is this policy adequate to guard against potential lost sales of 20–gallon aquariums due to a customer requesting one when they are out of stock?

We will use the five-step method. Step 1 is to ask a question. The store begins each sales week with an inventory of between one and three of the 20–gallon aquariums. The number of sales in one week depends on both the supply and the demand. The demand averages one per week but is subject to random fluctuations. It is possible that on some weeks demand will exceed supply, even if we start the week with the maximum inventory of three units. We would like to calculate the probability that demand exceeds supply. In order to get a specific answer, we need to make an assumption about the probabilistic nature

Fourth edition

Variables:	S_n = supply of aquariums at the beginning of week n D_n = demand for aquariums during week n
Assumptions:	If $D_{n-1} < S_{n-1},$ then $S_n = S_{n-1} - D_{n-1}$ If $D_{n-1} \geq S_{n-1},$ then $S_n = 3$ $\Pr\{D_n = k\} = e^{-1}/k!$
Objective:	Calculate $P_r\{D_n > S_n\}$

Figure 8.1: Results of step 1 of the inventory problem.

of the demand. It seems reasonable to assume that potential buyers arrive at random at a rate of one per week. Hence, the number of potential buyers in one week will have a Poisson distribution with mean one. (The Poisson distribution was introduced in Exercise 6 of Chapter 7.) Figure 8.1 summarizes the results of step 1. Step 2 is to select the modeling approach. We will use a Markov chain model.

A *Markov chain* can best be described as a sequence of random jumps. For the purposes of this book, we will assume that these jumps can only involve a finite discrete set of locations or states. Suppose that the random variables X_n take values in a finite discrete set. There is no harm in assuming that

$$X_n \in \{1, 2, 3, \ldots, m\}.$$

We say that the sequence $\{X_n\}$ is a Markov chain provided that the probability that $X_{n+1} = j$ depends only on X_n. If we define

$$p_{ij} = \Pr\{X_{n+1} = j | X_n = i\}, \tag{8.1}$$

then the entire future history of the process $\{X_n\}$ is determined by the p_{ij} and the probability distribution of the initial X_0. Of course, when we say "determined", we mean that the probabilities $\Pr\{X_n = i\}$ are determined. The actual value of X_n depends on random factors.

Example 8.2. Describe the behavior of the following Markov chain. The state variable

$$X_n \in \{1, 2, 3\}.$$

If $X_n = 1$, then $X_{n+1} = 1,\ 2,$ or 3 with equal probability. If $X_n = 2$, then $X_{n+1} = 1$ with probability 0.7, and $X_{n+1} = 2$ with probability 0.3. If $X_n = 3$, then $X_{n+1} = 1$ with probability 1.

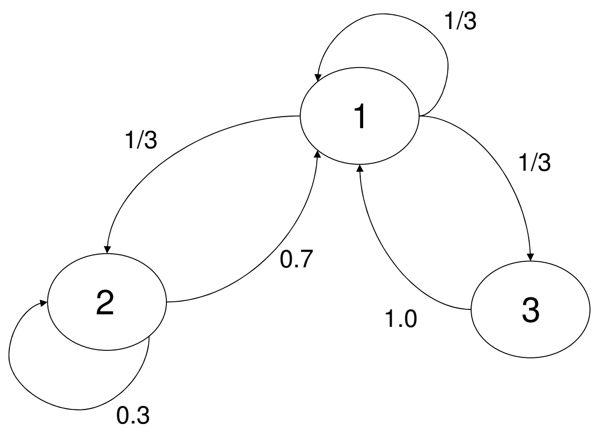

Figure 8.2: State transition diagram for Example 8.2.

The state transition probabilities p_{ij} are given by

$$\begin{aligned} p_{11} &= \frac{1}{3} \\ p_{12} &= \frac{1}{3} \\ p_{13} &= \frac{1}{3} \\ p_{21} &= 0.7 \\ p_{22} &= 0.3 \\ p_{31} &= 1, \end{aligned}$$

and the rest are zero. It is customary to write the p_{ij} in matrix form:

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}. \tag{8.2}$$

Here

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0.7 & 0.3 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Another convenient method is called the *state transition diagram* (see Figure 8.2). This makes it easy to visualize the Markov chain as a sequence of random jumps. Suppose $X_0 = 1$. Then $X_1 = 1$, 2, or 3 with probability 1/3 each. The probability that $X_2 = 1$ is obtained by calculating the probability associated with each individual sequence of jumps that transitions from state 1 to state 1 in two steps. Thus,

$$\Pr\{X_2 = 1\} = \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)(0.7) + \left(\frac{1}{3}\right)(1) = 0.67\overline{7}.$$

Similarly,

$$\Pr\{X_2 = 2\} = \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)(0.3) = 0.21\overline{1},$$

and

$$\Pr\{X_2 = 3\} = \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{9}.$$

To calculate $\Pr\{X_n = j\}$ for larger n, it is useful to observe that

$$\Pr\{X_{n+1} = j\} = \sum_i p_{ij} \Pr\{X_n = i\}. \tag{8.3}$$

The only way to get to state j at time $n+1$ is to be in some state i at time n, and then jump from i to j. Hence, we could have calculated

$$\Pr\{X_2 = 1\} = p_{11} \Pr\{X_1 = 1\} + p_{21} \Pr\{X_1 = 2\} + p_{31} \Pr\{X_1 = 3\},$$

and so forth. This is where the matrix notation comes in handy.

If we let

$$\pi_n(i) = \Pr\{X_n = i\},$$

then Eq. (8.3) can be written in the form

$$\pi_{n+1}(j) = \sum_i p_{ij}\, \pi_n(i). \tag{8.4}$$

If we let π_n denote the vector with entries $\pi_n(1), \pi_n(2), \ldots$ and let P denote the matrix in Eq. (8.2), then the set of equations relating π_{n+1} to π_n can be written most compactly in the form

$$\pi_{n+1} = \pi_n P. \tag{8.5}$$

For example, we have that $\pi_2 = \pi_1 P$, or

$$(0.67\overline{7},\, 0.21\overline{1},\, \frac{1}{9}) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0.7 & 0.3 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Now we can calculate $\pi_3 = \pi_2 P$ to get

$$\pi_3 = (0.485,\, 0.289,\, 0.226)$$

to three decimal places. Continuing on, we obtain

$$\begin{aligned}
\pi_4 &= (0.590,\, 0.248,\, 0.162)\\
\pi_5 &= (0.532,\, 0.271,\, 0.197)\\
\pi_6 &= (0.564,\, 0.259,\, 0.177)\\
\pi_7 &= (0.546,\, 0.266,\, 0.188)\\
\pi_8 &= (0.556,\, 0.262,\, 0.182)\\
\pi_9 &= (0.551,\, 0.264,\, 0.185)\\
\pi_{10} &= (0.553,\, 0.263,\, 0.184)\\
\pi_{11} &= (0.553,\, 0.263,\, 0.184)\\
\pi_{12} &= (0.553,\, 0.263,\, 0.184).
\end{aligned}$$

Notice that the probabilities $\pi_n(i) = \Pr\{X_n = i\}$ tend to a specific limiting value as n increases. We say that the stochastic process approaches steady state. This concept of steady state or equilibrium differs from that for a deterministic dynamic model. Because of random fluctuations, we cannot expect that the state variable will stay at one value when the system is in equilibrium. The best we can hope for is that the probability distribution of the state variable will tend to a limiting distribution. We call this the *steady–state distribution*. In Example 8.2 we have

$$\pi_n \to \pi,$$

where the steady–state probability vector

$$\pi = (0.553,\, 0.263,\, 0.184) \tag{8.6}$$

to three decimal places.

A faster way to calculate the steady–state vector π is as follows. Suppose that

$$\pi_n \to \pi.$$

Certainly

$$\pi_{n+1} \to \pi$$

too, so if we let $n \to \infty$ on both sides of Eq. (8.5), we obtain the equation

$$\pi = \pi P. \tag{8.7}$$

We can calculate π simply by solving this linear system of equations. For Example 8.2 we have

$$(\pi_1,\, \pi_2,\, \pi_3) = (\pi_1,\, \pi_2,\, \pi_3) \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0.7 & 0.3 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

and it is not hard to calculate that Eq. (8.6) is the only solution to this system of equations for which

$$\sum \pi_i = 1.$$

Not every Markov chain tends to steady state. For example, consider the two–state Markov chain for which

$$\Pr\{X_{n+1} = 2 | X_n = 1\} = 1$$

and

$$\Pr\{X_{n+1} = 1 | X_n = 2\} = 1.$$

The state variable alternates between states 1 and 2. Certainly π_n does not tend to a single limiting vector. We say that this Markov chain is *periodic* with period two. Generally we say state i is periodic with period δ if, starting at $X_n = i$, the chain can return to state i only at times $n + k\delta$. If $\{X_n\}$ is *aperiodic* (every state i has period $\delta = 1$), and if, for each i and j, it is possible to transition from i to j in a finite number of steps, we say that X_n is *ergodic.* There is a theorem that guarantees that an ergodic Markov chain tends to steady state. Furthermore, the distribution of X_n tends to the same steady–state distribution regardless of the initial state of the system (see, e.g., Çinlar (1975) p. 152). Thus, in Example 8.2, if we had started with $X_0 = 2$ or $X_0 = 3$, we would still see π_n converge to the same steady–state distribution π given by Eq. (8.6). The problem of calculating the steady–state probability vector π is mathematically equivalent to the problem of locating the equilibrium of a discrete–time dynamical system with state space $\pi \in \mathbb{R}^m$, $0 \le \pi_j \le 1$,

$$\sum \pi_i = 1$$

and iteration function

$$\pi_{n+1} = \pi_n P.$$

The previously cited theorem states that there is a unique asymptotically stable equilibrium π for this system whenever P represents an ergodic Markov chain.

We return now to the inventory problem of Example 8.1. We will model this problem using a Markov chain. Step 3 is to formulate the model. We begin with a consideration of the state space. The concept of state here is much the same as for deterministic dynamical systems. The state contains all of the information necessary in order to predict the (probabilistic) future of the process. We will take $X_n = S_n$, the number of 20–gallon aquariums in stock at the beginning of our sales week, as the state variable. The demand D_n relates to the dynamics of the model and will be used to construct the state transition matrix P. The state space is

$$X_n \in \{1, 2, 3\}.$$

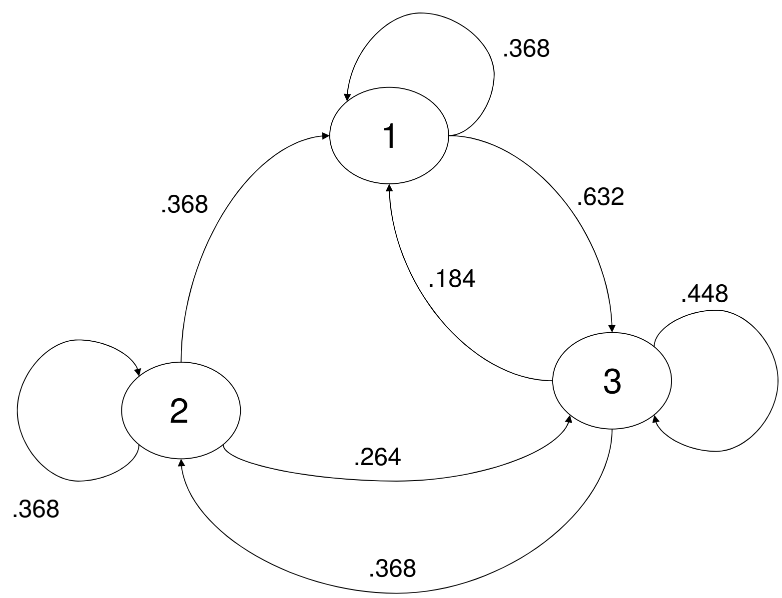

Figure 8.3: State transition diagram for the inventory problem.

We do not know the initial state, but it seems reasonable to assume that $X_0 = 3$. In order to determine P, we will begin by drawing the state transition diagram. See Figure 8.3. The distribution of the demand D_n yields

$$\begin{aligned}
\Pr\{D_n = 0\} &= 0.368 \\
\Pr\{D_n = 1\} &= 0.368 \\
\Pr\{D_n = 2\} &= 0.184 \\
\Pr\{D_n = 3\} &= 0.061 \\
\Pr\{D_n > 3\} &= 0.019,
\end{aligned} \tag{8.8}$$

so that if $X_n = 3$, then

$$\begin{aligned}
\Pr\{X_{n+1} = 1\} &= \Pr\{D_n = 2\} = 0.184 \\
\Pr\{X_{n+1} = 2\} &= \Pr\{D_n = 1\} = 0.368 \\
\Pr\{X_{n+1} = 3\} &= 1 - (0.184 + 0.368) = 0.448.
\end{aligned}$$

The remaining state transition probabilities are computed similarly. The state transition matrix is

$$P = \begin{pmatrix} .368 & 0 & .632 \\ .368 & .368 & .264 \\ .184 & .368 & .448 \end{pmatrix}. \tag{8.9}$$

Now for step 4. The analysis objective was to calculate the probability

$$\Pr\{D_n > S_n\}$$

that demand exceeds supply. In general this probability depends on n. More specifically, it depends on X_n. If $X_n = 3$, then

$$\Pr\{D_n > S_n\} = \Pr\{D_n > 3\} = 0.019,$$

and so forth. To get a better idea of how often demand will exceed supply, we need to know more about X_n.

Since $\{X_n\}$ is an ergodic Markov chain, we know that there is a unique steady–state probability vector π that can be computed by solving the steady–state equations. Substituting Eq. (8.9) back into Eq. (8.7), we obtain

$$\begin{aligned}\pi_1 &= .368\pi_1 + .368\pi_2 + .184\pi_3 \\ \pi_2 &= .368\pi_2 + .368\pi_3 \\ \pi_3 &= .632\pi_1 + .264\pi_2 + .448\pi_3,\end{aligned} \tag{8.10}$$

which we need to solve along with the condition

$$\pi_1 + \pi_2 + \pi_3 = 1,$$

to obtain the steady–state distribution of X_n. Since we now have four equations in three variables, we can delete one of the equations in Eq. (8.10) and then solve, which yields

$$\pi = (\pi_1, \pi_2, \pi_3) = (.285, .263, .452).$$

For all large n it is approximately true that

$$\begin{aligned}\Pr\{X_n = 1\} &= .285 \\ \Pr\{X_n = 2\} &= .263 \\ \Pr\{X_n = 3\} &= .452.\end{aligned}$$

Putting this together with our information about D_n, we obtain

$$\begin{aligned}\Pr\{D_n > S_n\} &= \sum_{i=1}^{3} \Pr\{D_n > S_n | X_n = i\} \Pr\{X_n = i\} \\ &= (.264)(.285) + (.080)(.263) + (.019)(.452) = .105\end{aligned}$$

for large n. In the long run, demand will exceed supply about 10% of the time.

It is easy to compute the steady–state probabilities using a computer algebra system. Figure 8.4 illustrates the use of the computer algebra system Maple to solve the system of equations in Eq. (8.10) to find the steady–state probabilities.

Computer algebra systems are quite useful in such problems, especially when performing sensitivity analysis. If you have access to a computer algebra system, it will be useful in the exercises at the end of this chapter. Even if you prefer to solve systems of equations by hand, you will have the ability to verify your results.

Finally, step 5. The current inventory policy results in lost sales about 10% of the time, or at least five lost sales per year. Most of this is due to the fact

```
> s:={pi1=.368*pi1+.368*pi2+.184*pi3,
>      pi2=.368*pi2+.368*pi3,
>      pi1+pi2+pi3=1};
```

$s := \{\pi 1 = 0.368\ \pi 1 + 0.368\ \pi 2 + 0.184\ \pi 3, \pi 2 = 0.368\ \pi 2 + 0.368\ \pi 3, \pi 1 + \pi 2 + \pi 3 = 1\}$

```
> solve(s,{pi1,pi2,pi3});
```

$\{\pi 2 = 0.2631807648, \pi 1 = 0.2848348783, \pi 3 = 0.4519843569\}$

```
>
```

Figure 8.4: Calculation of the steady–state distribution of the number of 20–gallon aquariums in stock at the beginning of the week for the inventory problem, using the computer algebra system Maple.

that we do not order more aquariums when only one is left. Although we only sell an average of one unit per week, the actual number of potential sales per week (demand) fluctuates from one week to the next. Hence, when we start the week with only one unit in stock, we run a significant risk (about a one in four chance) of losing potential sales due to insufficient inventory. In the absence of other factors, such as a discount for orders of three or more, it seems reasonable to try out a new inventory policy in which we never start out a week with only one aquarium.

We now come to the subject of sensitivity analysis and robustness. The main sensitivity issue is the effect of the arrival rate λ of potential buyers on the probability that demand exceeds supply. Currently $\lambda = 1$ customer per week. For arbitrary λ, the state transition matrix for X_n is given by

$$P = \begin{pmatrix} e^{-\lambda} & 0 & 1 - e^{-\lambda} \\ \lambda e^{-\lambda} & e^{-\lambda} & 1 - (1+\lambda)e^{-\lambda} \\ \lambda^2 e^{-\lambda}/2 & \lambda e^{-\lambda} & 1 - (\lambda + \lambda^2/2)e^{-\lambda} \end{pmatrix}, \tag{8.11}$$

using the fact that D_n has a Poisson distribution. While it would be possible to carry through the calculation of $p = \Pr\{D_n > S_n\}$ from this point, it would be very messy. It makes more sense simply to repeat the calculations of step 4 for a few selected values of λ near 1. The results of this exercise are shown in Figure 8.5. They confirm that our basic conclusions are not particularly sensitive to the exact value of λ. The sensitivity $S(p, \lambda)$ is around 1.5. (Another sensible option for sensitivity analysis is to use a computer algebra system to perform the messy calculations. See Exercise 2 at the end of this chapter.)

Finally, we should consider the robustness of our model. We have assumed a Markov chain model based on a Poisson process model of the arrival process. The robustness of the Poisson process model as a representative of a more general arrival process was discussed briefly at the end of Section 7.2. It is reasonable to conclude that our results would not be altered significantly if the arrival process were not exactly Poisson. The basic assumption here is that the arrival process represents the merging of a large number of independent arrival processes. Many kinds of customers arrive at the shop from time to

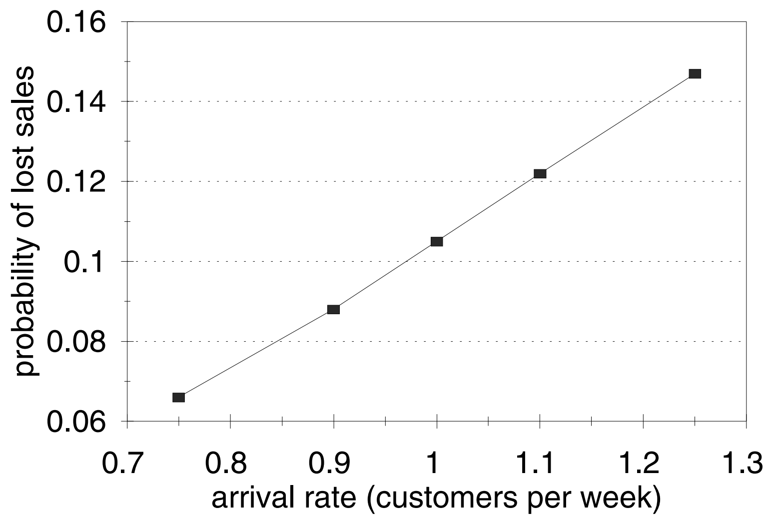

Figure 8.5: Graph showing the sensitivity of the probability of lost sales to arrival rate in the inventory problem.

time to buy a 20–gallon aquarium, and it is reasonable to assume that they do not coordinate this activity with each other. Of course, certain store activities, such as an advertised sale price on 20–gallon aquariums, would invalidate this assumption, resulting in the need to reexamine the conclusions of our modeling exercise. It may also be that there are significant seasonal variations in the demand for this item.

The other basic modeling assumption is that the inventory level S_n represents the state of the system. A more sophisticated model might take into account the store manager's response to long–term fluctuations in sales, such as seasonal variations. The mathematical analysis of such a model is more complex, but not essentially different from what we did here. We just expand the state space to include information on past sales; say, S_n, S_{n-1}, S_{n-2}, S_{n-3}. Of course, our transition matrix P is now 81×81 instead of 3×3.

Many different inventory policies are possible. Several of these are explored in the exercises at the end of the chapter. Which inventory policy is the best? One way to approach this question is to formulate an optimization model based on a generalized version of our Markov chain model. A range of inventory policies is described in terms of one or more decision variables, and the objective is defined in terms of the resulting steady–state probabilities. The study of such models is called *Markov decision theory*. Details can be found in any introductory text on operations research (e.g., Hillier et al. (1990)).

Variables:	X_t = number of forklifts in repair at time t months.
Assumptions:	Vehicles arrive for repair at the rate of 4.5 per month. Maximum repair rate is 7.3 vehicles per month.
Objective:	Calculate EX_t, $\Pr\{X_t > 0\}$

Figure 8.6: Results of step 1 of the forklift problem.

8.2 Markov Processes

A Markov process model is the continuous–time analogue of the Markov chain model introduced in the preceding section. It may also be considered as a stochastic analogue to a continuous–time dynamical system model.

Example 8.3. A mechanic working for a heavy equipment repair facility is responsible for the repair and maintenance of forklift trucks. When forklifts break down, they are taken to the repair facility and serviced in the order of their arrival. There is space for 27 forklift trucks at the facility, and last year the facility repaired 54 trucks. Average repair time for a single vehicle is about three days. In the past few months certain questions have been raised about the effectiveness and the efficiency of this operation. The two central issues are the time it takes to have a machine repaired, and the percentage of time the mechanic devotes to this part of his duties.

We will analyze the situation using a mathematical model of the repair facility. Forklift trucks arrive at the facility for repair at a rate of $54/12 = 4.5$ per month. The maximum rate at which they can be repaired is $22/3 \approx 7.3$ vehicles per month, based on an average of 22 working days per month. Let X_t denote the number of vehicles in the repair shop at time t. We are interested in the average number in service, EX_t, and the proportion of time that the mechanic is busy repairing machines, represented by $\Pr\{X_t > 0\}$. Figure 8.6 summarizes the results of step 1.

We will model the repair facility using a Markov process.

A *Markov process* is the continuous–time analogue of the Markov chain introduced in the previous section. As before, we will assume that the state space is finite; i.e., we will suppose

$$X_t \in \{1, 2, 3, \ldots, m\}.$$

The stochastic process $\{X_t\}$ is a Markov process if the current state X_t really represents the state of the system; i.e., it totally determines the probabilistic future of the process. This condition is formally written as

$$\Pr\{X_{t+s} = j | X_u : u \leq t\} = \Pr\{X_{t+s} = j | X_t\}. \tag{8.12}$$

The Markov property, Eq. (8.12), has two important implications. First of all, the time until the next transition does not depend on how long the process has been in the current state. In other words, the distribution of time spent in a particular state has the memoryless property. Let T_i denote the time spent in state i. Then the Markov property says that

$$\Pr\{T_i > t + s | T_i > s\} = \Pr\{T_i > t\}. \tag{8.13}$$

In Section 7.2 we showed that the exponential distribution has this property, so T_i could have density function

$$F_i(t) = \lambda_i e^{-\lambda_i t}. \tag{8.14}$$

In fact, the exponential distribution is the only continuous distribution having the memoryless property. (This is a deep theorem in real analysis. See Billingsley (1979) p. 160.) Hence, for a Markov process the distribution of time in a particular state is exponential with parameter λ_i, which depends in general on the state i.

The second important implication of the Markov property has to do with state transition. The probability distribution that describes the identity of the next state can depend only on the current state. Thus, the sequence of states visited by the process forms a Markov chain. If we let p_{ij} denote the probability that the process jumps from state i to state j, then the embedded Markov chain has state transition probability matrix $P = (p_{ij})$.

Example 8.4. Consider a Markov chain with state transition probability

$$P = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 1/2 & 0 & 1/2 \\ 3/4 & 1/4 & 0 \end{pmatrix}, \tag{8.15}$$

and form a Markov process by assuming that the jumps of $\{X_t\}$ follow this Markov chain with the average times in states 1, 2, 3 equal to 1, 2, and 3, respectively.

Solving the steady–state equation $\pi = \pi P$ shows that the proportion of jumps that land in states 1, 2, and 3 are 0.396, 0.227, and 0.377, respectively. However, the proportion of time spent in each state also depends on how long we wait in one state before the next jump. Correcting for this produces the relative proportions 1(.396), 2(.227), and 3(.377). If we normalize to 1 (divide each term by the sum), we get 0.200, 0.229, and 0.571. Hence the Markov process spends about 57.1% of the time in state 3, and so on. We call this the steady–state distribution for the Markov process. Generally, if $\pi = (\pi_1, \ldots, \pi_m)$ is the steady–state distribution for the embedded Markov chain and $\lambda = (\lambda_1, \ldots, \lambda_m)$ is the vector of rates, then the proportion of time spent in state i is given by

$$P_i = \frac{(\pi_i/\lambda_i)}{(\pi_1/\lambda_1) + \cdots + (\pi_m/\lambda_m)}. \tag{8.16}$$

The reciprocal of the rate λ_i represents mean time in state i. In summary, a Markov process can be thought of as a Markov chain where the time between jumps has an exponential distribution, depending on the current state.

An equivalent model can be formulated as follows. Given that $X_t = i$, let T_{ij} be exponential with parameter $a_{ij} = \lambda_i p_{ij}$. Furthermore, suppose that $T_{i1}, \ldots, T_{im}$ are independent. Then the time T_i until the next jump is the minimum of $T_{i1}, \ldots, T_{im}$, and the next state is the state j such that T_{ij} is the minimum of $T_{i1}, \ldots, T_{im}$. The mathematical equivalence between the two forms of the Markov process model follows from the fact that

$$T_i = \min(T_{i1}, \ldots, T_{im})$$

is exponential with parameter

$$\lambda_i = \sum_j a_{ij}$$

and that

$$\Pr\{T_i = T_{ij}\} = p_{ij}.$$

(The proof is left to the reader. See Exercise 7 at the end of the chapter.) The parameter

$$a_{ij} = \lambda_i p_{ij}$$

denotes the rate at which the process tends to go from state i to state j. It is customary to depict the rates a_{ij} in a *rate diagram.* The rate diagram for Example 8.4 is given in Figure 8.7. Often the structure of a Markov process is originally specified by such a diagram.

This alternative formulation for the Markov process leads to a convenient method of computing the steady–state distribution, based on the rate diagram. As before we define

$$a_{ij} = \lambda_i p_{ij}$$

to be the rate at which the process tends to jump from state i to state j. Also, let

$$a_{ii} = -\lambda_i$$

denote the rate at which the process tends to leave state i. It can be shown that the probability functions

$$P_i(t) = \Pr\{X_t = i\} \tag{8.17}$$

must satisfy the differential equations

$$\begin{aligned} P_1'(t) &= a_{11}P_1(t) + \cdots + a_{m1}P_m(t) \\ &\vdots \\ P_m'(t) &= a_{1m}P_1(t) + \cdots + a_{mm}P_m(t) \end{aligned} \tag{8.18}$$

(see Çinlar (1975) p. 255). This basic condition can be most easily understood by using the *fluid–flow analogy.* Visualize the probabilities $P_i(t)$ as the amount of fluid (probability mass) at each state i. The rates a_{ij} represent the rate of fluid flow, and the fact that

$$P_1(t) + \cdots + P_m(t) = 1$$

means that the total amount of fluid stays equal to 1. In the case of Example 8.4, we have

$$\begin{aligned} P_1'(t) &= -P_1(t) + \frac{1}{4}\ P_2(t) + \frac{1}{4}\ P_3(t) \\ P_2'(t) &= \frac{1}{3}\ P_1(t) - \frac{1}{2}\ P_2(t) + \frac{1}{12}\ P_3(t) \\ P_3'(t) &= \frac{2}{3}\ P_1(t) + \frac{1}{4}\ P_2(t) - \frac{1}{3}\ P_3(t). \end{aligned} \tag{8.19}$$

The steady–state distribution for the Markov process corresponds to the steady–state solution to this system of differential equations. Setting $P_i' = 0$ for all i, we obtain

$$\begin{aligned} 0 &= -P_1 + \frac{1}{4}\ P_2 + \frac{1}{4}\ P_3 \\ 0 &= \frac{1}{3}\ P_1 - \frac{1}{2}\ P_2 + \frac{1}{12}\ P_3 \\ 0 &= \frac{2}{3}\ P_1 + \frac{1}{4}\ P_2 - \frac{1}{3}\ P_3. \end{aligned} \tag{8.20}$$

Solving the system of linear equations in Eq. (8.20) together with the condition

$$P_1 + P_2 + P_3 = 1$$

yields

$$P = \left(\frac{7}{35}, \frac{8}{35}, \frac{20}{35}\right),$$

which is the same as before.

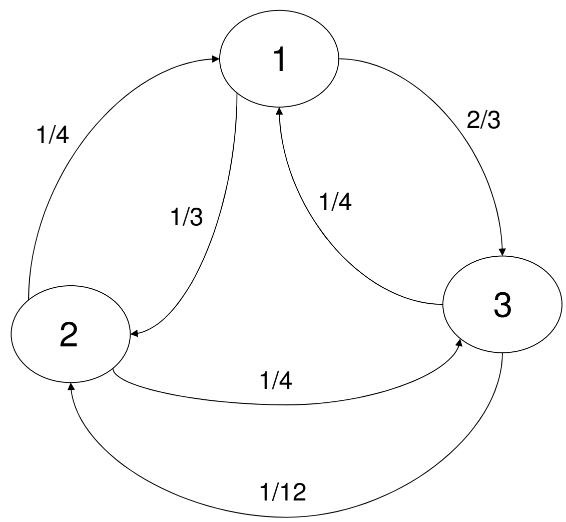

Figure 8.7: Rate diagram for Example 8.4, illustrating the rate at which the process tends to jump from one state to another.

One simple way to determine the system of equations that must be solved to get the steady–state distribution for a Markov process is to use the fluid–flow analogy. Fluid flows into and out of each state. In order for the system to remain in equilibrium, the rate at which fluid flows into each state must equal the rate at which it flows back out. For example, in Fig. 8.7 fluid is flowing out of state 1 at the rate of $1/3 + 2/3 = 1 \times P_1$. Fluid flows from state 2 back into state 1 at the rate of $1/4 \times P_2$ and from state 3 to state 1 at the rate of $1/4 \times P_3$, so we have the condition

$$P_1 = 1/4\ P_2 + 1/4\ P_3.$$

By applying this principle of

$$[\text{Rate out}] = [\text{Rate in}]$$

to the other two states as well, we obtain the system of equations

$$\begin{aligned} P_1 &= \frac{1}{4}\ P_2 + \frac{1}{4}\ P_3 \\ \frac{1}{2}\ P_2 &= \frac{1}{3}\ P_1 + \frac{1}{12}\ P_3 \\ \frac{1}{3}\ P_3 &= \frac{2}{3}\ P_1 + \frac{1}{4}\ P_2. \end{aligned} \tag{8.21}$$

which is equivalent to the system of equations in Eq. (8.20). We call the system of equations in Eq. (8.21) the *balance equations* for our Markov process model. They express the condition that the rates into and out of each state are in balance.

In Section 8.1 we remarked that an ergodic Markov chain will always tend to steady state. Now we will state the corresponding result for Markov processes. A Markov process is called *ergodic* if, for every pair of states i and j, it is possible to jump from i to j in a finite number of transitions. There is a theorem that guarantees that an ergodic Markov process always tends to steady state. Furthermore, the distribution of X_t tends to the same steady–state distribution regardless of the initial state of the system (see, for example, Çinlar (1975) p. 264). Let

$$P(t) = (P_1(t), \dots, P_m(t))$$

denote the current probability distribution of our Markov process. Then, this theorem says that for any initial probability distribution $P(0)$ on the state space, we will always see the probability distribution $P(t)$ of the Markov process state vector X_t converge to the same steady–state distribution

$$P = (P_1, \dots, P_m)$$

as $t \to \infty$. The system of differential equations in Eq. (8.18) that describes the dynamics of the probability distribution $P(t)$ can be written in matrix form as

$$P(t)' = P(t)A, \tag{8.22}$$

where $A = (a_{ij})$ is the matrix of rates. This is a linear system of differential equations on the space

$$S = \{x \in \mathbb{R}^m : 0 \le x_i \le 1; \sum x_i = 1\}. \tag{8.23}$$

Our theorem says that if this dynamical system in Eq. (8.22) represents an ergodic Markov process, then there exists a unique stable equilibrium solution P. Furthermore, for any initial condition $P(0)$ we will have $P(t) \to P$ as $t \to \infty$. More detailed information about the transient (time–dependent) behavior of the Markov process can be obtained by explicitly solving the linear system in Eq. (8.22) by the usual methods.

Now we return to the forklift problem of Example 8.3. We want to formulate a Markov process model for X_t = the number of forklifts in repair at time t months. Since the facility can handle only 27 forklifts, we have

$$X_t \in \{0, 1, 2, \dots, 27\}.$$

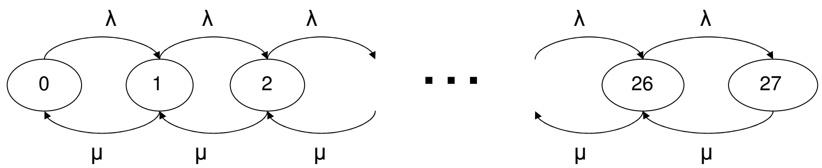

Figure 8.8: Rate diagram for the forklift problem, illustrating the rate at which the number of forklifts in service tends to increase or decrease.

The only allowed transitions are from $X_t = i$ to $X_t = i + 1$ or $i - 1$. The rates up and down are $\lambda = 4.5$ and $\mu = 7.3$, respectively, except that we cannot transition up from state 27 or down from state 0. The rate diagram for this problem is shown in Figure 8.8.

The steady–state equations $PA = 0$ can be obtained from the rate diagram by using the

$$[\text{Rate out}] = [\text{Rate in}]$$

principle. From Fig. 8.8 we obtain

$$\begin{aligned}
\lambda P_0 &= \mu P_1 \\
(\mu + \lambda)P_1 &= \lambda P_0 + \mu P_2 \\
(\mu + \lambda)P_2 &= \lambda P_1 + \mu P_3 \\
&\vdots \\
(\mu + \lambda)P_{26} &= \lambda P_{25} + \mu P_{27} \\
\mu P_{27} &= \lambda P_{26}.
\end{aligned} \tag{8.24}$$

Solving along with

$$\sum P_i = 1$$

will yield the steady–state $\Pr\{X_t = i\}$. We are interested in

$$\Pr\{X_t > 0\} = 1 - \Pr\{X_t = 0\} = 1 - P_0$$

and in

$$EX_t = \sum iP_i.$$

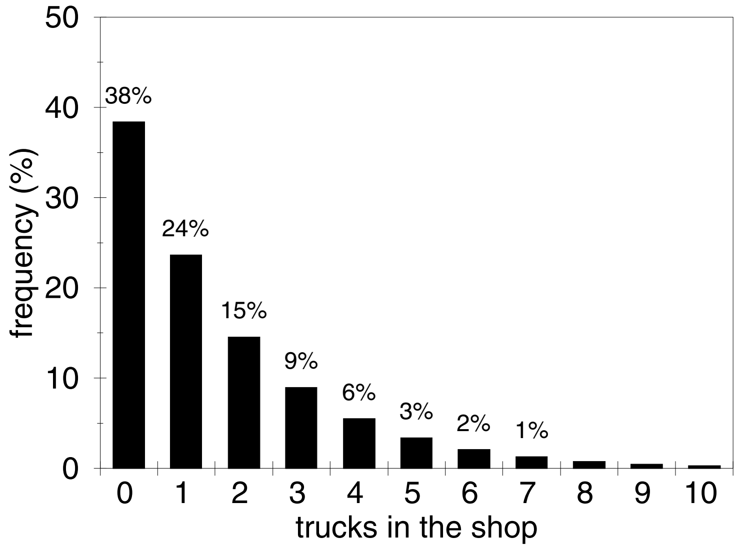

Figure 8.9: Histogram showing the distribution of the number of forklifts in service at any given time.

Moving on to step 4, we first solve for P_1 in terms of P_0, then for P_2 in terms of P_1, and so forth, to obtain

$$P_n = (\lambda/\mu)\, P_{n-1}$$

for all $n = 1, 2, 3, \ldots, 27$. Then

$$P_n = (\lambda/\mu)^n\, P_0$$

for all such n. Since

$$\sum_{n=0}^{27} P_n = P_0 \sum_{n=0}^{27} \left(\frac{\lambda}{\mu}\right)^n = 1,$$

we must have

$$P_0 = \frac{1-\rho}{1-\rho^{28}},$$

where $\rho = \lambda/\mu$. Here we have used the standard formula for the sum of a finite geometric series. For $n = 1, 2, 3, \ldots, 27$ we have

$$P_n = \rho^n P_0 = \frac{\rho^n(1-\rho)}{1-\rho^{28}}. \tag{8.25}$$

Now $\rho = \lambda/\mu = 4.5/7.3 \approx 0.616$, so $1-\rho^{28} \approx 0.9999987$. Hence, we may assume for all practical purposes that $P_0 = 1-\rho$ and $P_n = \rho^n(1-\rho)$ for $n \geq 1$.

Now we calculate our two measures of performance. First, we have

$$\Pr\{X_t > 0\} = 1 - P_0 = \rho \approx 0.616.$$

Next, we have

$$\begin{aligned} EX_t &= \sum_{n=0}^{27} nP_n \\ &= \sum_{n=0}^{27} n\rho^n(1-\rho), \end{aligned} \tag{8.26}$$

which yields $EX_t = 1.607$.

To summarize (step 5), we consider a system where forklift trucks break down at a rate of 4.5 per month and are taken to a repair facility with the capacity to service up to 7.3 per month. Since the rate at which vehicles arrive for repair is only about 60% of the potential service rate, the mechanic is busy with this activity only about 60% of the time. However, since breakdowns occur essentially at random, there will be times when there is more than one vehicle in the shop at one time, through no fault of the mechanic. In fact, on an average day we would expect to see 1.6 vehicles in the repair facility. By this, we mean that if we kept track on a daily basis of the number of vehicles in the repair facility, then at the end of the year this number would average about 1.6. More specifically, we would expect the distribution indicated by Fig. 8.9.

Murphy's Law is certainly in evidence here. Even if the mechanic does his job flawlessly, on 8% of the working days in a year, there will be 5 or more forklift trucks in the shop. Since it takes three days to fix one truck, this represents a backlog of about three weeks. Assuming 250 working days a year, this unfortunate situation will occur about 20 days out of the year. Meanwhile, a time study will show the mechanic busy on this part of his job only about 60% of the time. This seeming discrepancy can be explained by the simple fact that breakdowns are not always nicely spaced out. Sometimes, due to sheer bad luck, several machines will break down in rapid succession, and the mechanic will be swamped. At other times there will be long periods between breakdowns, and the mechanic will have no repairs to perform.

There are really two management problems here. One is the problem of idle time. There is no way to avoid the sporadic nature of this job activity, so idle time will continue to be a significant concern. Management might consider the extent to which the mechanic can perform other duties during this idle time. Of course, these duties should be easily interruptible in case there is a breakdown.

The second problem is backlogs. Sometimes a number of forklift trucks will be waiting for repair. Additional study may be required to determine whether this problem should be addressed by devoting additional manpower to the repair activity during particularly busy periods. (This problem is the subject of Exercise 8 at the end of this chapter.)

Finally, we come to the very important questions of sensitivity and robustness. Both measures of performance depend on the ratio

$$\rho = \frac{\lambda}{\mu},$$

currently equal to 0.616. The relation

$$\Pr\{X_t > 0\} \approx \rho \tag{8.27}$$

requires no further comment. If we let $A = EX_t$ denote the average number in the system, then (using $1 - \rho^{28} \approx 1$) we have

$$A = \sum_{n=0}^{27} n\rho^n(1-\rho), \tag{8.28}$$

which reduces to

$$A = \rho + \rho^2 + \rho^3 + \cdots + \rho^{27} - 27\rho^{28}.$$

This is approximately

$$\rho(1 + \rho + \rho^2 + \rho^3 + \cdots) = \frac{\rho}{1-\rho},$$

and so

$$\frac{dA}{d\rho} \approx \frac{1}{(1-\rho)^2},$$

so that

$$S(A, \rho) \approx 2.6.$$

A small error in ρ would not significantly alter our conclusions based on $A = EX_t$.

We should also consider the size of the facility storage area, currently at $K = 27$ forklift trucks. We have seen that for moderate ρ (not too close to 1) this parameter makes little difference, in the sense that we have continually used the approximation

$$1 - \rho^{K+1} \approx 1.$$

Indeed, this approximation is equivalent to the assumption that there is unlimited storage capacity for forklift trucks to repair. Essentially, we are assuming that $K = \infty$. If more storage capacity is made available, this will not change the rate λ at which vehicles arrive to be repaired. Thus, $\rho = \lambda/\mu$ remains the same, so both our measures of performance are insensitive to K. The only effect of increasing K (i.e., obtaining aditional storage space) will be to increase the number of trucks that can be waiting for service.

The model we have used to represent the repair facility is a special case of a *queuing model*. A queuing model represents a system consists of one or more service facilities at which arrivals are processed, and those arrivals that cannot be processed have to wait for service in a queue. There is a large body of literature on queuing models, including much current research. A textbook on operations research (e.g., Hillier et al. (1990)) is a good place to begin. The most important assumption we might try to relax is that service times have an exponential distribution. There is ample reason to suspect that the arrivals are more or less random. There is a result (based on the simplifying assumption $K = \infty$, which we have made before) for a general service time distribution with variance σ^2 that states that $\rho = \lambda/\mu$ is the probability that the server is busy and the steady–state

$$EX_t = \rho + \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)}. \tag{8.29}$$

Of course, this reduces to $\rho/(1-\rho)$ in the case of an exponential service time. In that case, $\sigma = 1/\mu$. The general conclusion to be drawn from this formula is that the average number of vehicles in repair grows with the variance of service time. Thus, more uncertainty about the length of a repair will result in longer waiting times.

8.3 Linear Regression

The single most commonly used stochastic model assumes that the expected value of the state variable is a linear function of time. The model is attractive not only because of its wide range of applications, but also because of the availability of good software implementations.

Example 8.5. Adjustable–rate mortgages on private homes are commonly based on one of several market indices tabulated by the federal home loan bank. The author's mortgage is adjusted yearly on the basis of the U.S. Treasury one–year Constant Maturity (CM1) index for May of each year. Historical data for the three–year period beginning June 1986 are shown in Table 8.1 (source: Board of Governors of the Federal Reserve). Use this information to project the estimated value of this index in May 1990, the date of the next adjustment.

We will use the five-step method. Step 1 is to ask a question. We are attempting to estimate the future trend in a variable that exhibits a tendency to grow with time, along with some random fluctuation. Let X_t denote the U.S. Treasury one–year Constant Maturity (CM1) index at time t months after May 1986. A graph of X_t for $t = 1, \ldots, 37$ is shown in Figure 8.10. We want to estimate X_{48}. If we assume that X_t depends in part on a random element, then we cannot expect to predict X_{48} exactly. The best we can hope for is an average value EX_{48}, along with some kind of measure of the magnitude of uncertainty. For the moment we will concentrate on obtaining an estimate of EX_{48}. We will leave the other matter for the section on sensitivity analysis.

Step 2 is to select the modeling approach. We will model this problem using linear regression.

month	TB3	TB6	CM1	CM2	CM3	CM5
6/86	6.21	6.28	6.73	7.18	7.41	7.64
7/86	5.84	5.85	6.27	6.67	6.86	7.06
8/86	5.57	5.58	5.93	6.33	6.49	6.80
9/86	5.19	5.31	5.77	6.35	6.62	6.92
10/86	5.18	5.26	5.72	6.28	6.56	6.83
11/86	5.35	5.42	5.80	6.28	6.46	6.76
12/86	5.49	5.53	5.87	6.27	6.43	6.67
1/87	5.45	5.47	5.78	6.23	6.41	6.64
2/87	5.59	5.60	5.96	6.40	6.56	6.79
3/87	5.56	5.56	6.03	6.42	6.58	6.79
4/87	5.76	5.93	6.50	7.02	7.32	7.57
5/87	5.75	6.11	7.00	7.76	8.02	8.26
6/87	5.69	5.99	6.80	7.57	7.82	8.02
7/87	5.78	5.86	6.68	7.44	7.74	8.01
8/87	6.00	6.14	7.03	7.75	8.03	8.32
9/87	6.32	6.57	7.67	8.34	8.67	8.94
10/87	6.40	6.86	7.59	8.40	8.75	9.08
11/87	5.81	6.23	6.96	7.69	7.99	8.35
12/87	5.80	6.36	7.17	7.86	8.13	8.45
1/88	5.90	6.31	6.99	7.63	7.87	8.18
2/88	5.69	5.96	6.64	7.18	7.38	7.71
3/88	5.69	5.91	6.71	7.27	7.50	7.83
4/88	5.92	6.21	7.01	7.59	7.83	8.19
5/88	6.27	6.53	7.40	8.00	8.24	8.58
6/88	6.50	6.76	7.49	8.03	8.22	8.49
7/88	6.73	6.97	7.75	8.28	8.44	8.66
8/88	7.02	7.36	8.17	8.63	8.77	8.94
9/88	7.23	7.43	8.09	8.46	8.57	8.69
10/88	7.34	7.50	8.11	8.35	8.43	8.51
11/88	7.68	7.76	8.48	8.67	8.72	8.79
12/88	8.09	8.24	8.99	9.09	9.11	9.09
1/89	8.29	8.38	9.05	9.18	9.20	9.15
2/89	8.48	8.49	9.25	9.37	9.32	9.27
3/89	8.83	8.87	9.57	9.68	9.61	9.51
4/89	8.70	8.73	9.36	9.45	9.40	9.30
5/89	8.40	8.39	8.98	9.02	8.98	8.91
6/89	8.22	8.00	8.44	8.41	8.37	8.29

Table 8.1: Possible adjustable–rate mortgage loan indices.

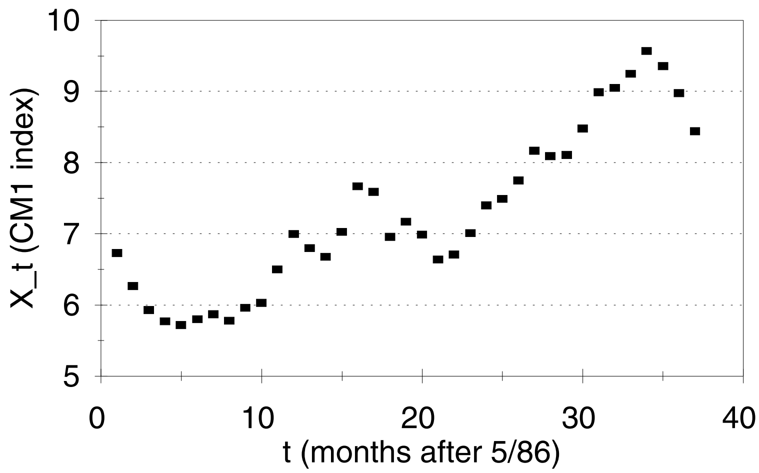

Figure 8.10: Graph of CM1 index versus time for the ARM problem.

The *linear regression* model assumes that

$$X_t = a + bt + \varepsilon_t, \tag{8.30}$$

where a and b are real constants and ε_t is a random variable that represents the effect of random fluctuations. It is assumed that

$$\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$$

are independent and identically distributed with mean zero. It is also common to assume that ε_t is normal; i.e., that for some $\sigma > 0$ the random variable

$$\varepsilon_t/\sigma$$

has the standard normal density. In the case where the random fluctuations represented by ε_t involve the additive effects of a fairly large number of independent random factors, this normal assumption is justified by the central limit theorem. (The normal density and the central limit theorem were introduced in Section 7.3.)

Since the error term ε_t has mean zero,

$$EX_t = a + bt, \tag{8.31}$$

so the problem of estimating EX_t reduces to the estimation of the parameters a and b. If we were to graph the line

$$y = a + bt \tag{8.32}$$

on the graph of Fig. 8.10, we would expect the data points to lie near this line, with some above and some below. The best–fitting line, representing our best estimate of the parameters a and b, should minimize the extent to which the data points deviate from the line.

Given a set of data points

$$(t_1,\, y_1), \ldots, (t_n,\, y_n),$$

we measure the goodness of fit of the regression line in terms of the vertical distance

$$|y_i - (a + bt_i)|$$

between the data point $(t_i,\, y_i)$ and the point on the regression line, Eq. (8.32), at $t = t_i$. To avoid absolute value signs, which are troublesome in an optimization problem, we measure the overall goodness of fit by

$$F(a,\, b) = \sum_{i=1}^{n} (y_i - (a + bt_i))^2. \tag{8.33}$$

The best–fitting line is characterized by a global minimum of the objective function in Eq. (8.33). Setting the partial derivatives $\partial F/\partial a$ and $\partial F/\partial b$ equal to zero yields

$$\begin{aligned} \sum_{i=1}^{n} y_i &= na + b \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} t_i y_i &= a \sum_{i=1}^{n} t_i + b \sum_{i=1}^{n} t_i^2. \end{aligned} \tag{8.34}$$

Solving these two linear equations in two unknowns determines a and b.

An estimate of the predictive power of the regression equation (8.32) can be obtained as follows. Let

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{8.35}$$

denote the average or mean value of the y data points, and for each i let

$$\hat{y}_i = a + bt_i. \tag{8.36}$$

The total variation $y_i - \overline{y}$ between any one data value and the mean value can be expressed as the sum

$$(y_i - \overline{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})\,. \tag{8.37}$$

The first term on the right-hand side of Eq. (8.37) represents the error (vertical distance of the data point from the regression line), and the second term represents the amount of deviation in y accounted for by the regression line. A little algebra shows that

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2. \tag{8.38}$$

The statistic

$$R^2 = \frac{\displaystyle\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\displaystyle\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{8.39}$$

measures the portion of the total variation in the data accounted for by the regression line. The remaining portion of the total variation is attributed to random errors; i.e., the effect of ε_t. If R^2 is close to 1, then the data are very nearly linear. If R^2 is close to 0, the data are very nearly random.

Most educational computer laboratories have statistical packages that will automatically compute a, b, and R^2 from a data set. Inexpensive software of the same kind is available for most personal computers, and some hand–held calculators have built–in linear regression functions. For the linear regression problems in this book, any of these methods will suffice. It is not recommended that these problems be solved by hand.

Step 3 is to formulate the model. We will let X_t represent the value of the CM1 index t months after May 1986, and we will assume the linear regression model in Eq. (8.30). The data are

$$\begin{aligned} (t_1,\, y_1) &= (1,\, 6.73) \\ (t_2,\, y_2) &= (2,\, 6.27) \\ &\;\vdots \\ (t_{37},\, y_{37}) &= (37,\, 8.44). \end{aligned} \tag{8.40}$$

The best–fitting regression line can be obtained by solving the linear system of equations in Eq. (8.34) to obtain a and b. Then the goodness–of–fit statistic R^2 can be obtained from Eq. (8.39). Using a computer implementation of this linear regression model will allow us to avoid a lot of tedious calculation.

Step 4 is to solve the model. We used the Minitab statistical package to obtain the regression line

$$y = 5.45 + 0.0970t \tag{8.41}$$

and

$$R^2 = 83.0\%$$

(see Figure 8.11 for selected outputs). To do this, first we entered the CM1 data into one column of a Minitab worksheet and entered the time index numbers

$t = 1, 2, 3, \ldots, 37$ into another column. Then we used the pull-down menus to issue the command `Stat > Regression > Regression` and specified the CM1 data as the response and the time index data as the predictor. To get the prediction interval for $t = 48$, we selected the `Options` button in the regression window and entered "48" in the box labelled `Prediction intervals for new observations`. The details of this procedure and the resulting outputs are quite similar if you use a different statistical package, a spreadsheet with a regression tool, or even a hand calculator (although some features, like prediction intervals, may not be available).

```
The regression equation is
cm1 = 5.45 + 0.0970 t

Predictor       Coef   SE Coef      T      P
Constant      5.4475    0.1615  33.73  0.000
t           0.096989  0.007409  13.09  0.000

S = 0.481203   R-Sq = 83.0%   R-Sq(adj) = 82.6%

Unusual Observations

Obs    t     cm1     Fit  SE Fit  Residual  St Resid
  1  1.0  6.7300  5.5445  0.1551    1.1855     2.60R

R denotes an observation with a large standardized residual.

Predicted Values for New Observations

New
Obs      Fit  SE Fit        95% CI              95% PI
  1  10.1030  0.2290  (9.6381, 10.5678)  (9.0211, 11.1848)X

X denotes a point that is an outlier in the predictors.

Values of Predictors for New Observations

New
Obs     t
  1  48.0
```

Figure 8.11: Solution to the ARM problem using the statistical package Minitab.

Equation (8.41) represents the best–fitting straight line through the data points in Eq. (8.40). See Figure 8.12 for a graphical illustration. The average trend in the CM1 index over the period June 1986 to June 1989 has been to

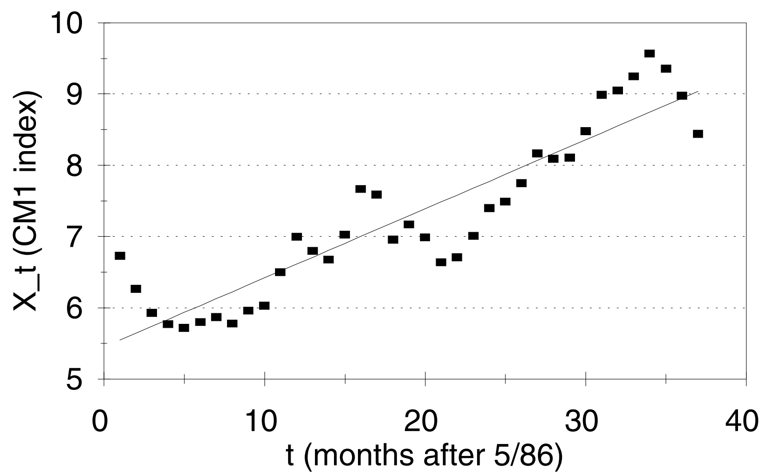

Figure 8.12: Graph of CM1 index versus time showing the regression line for the ARM problem.

increase by 0.0970 per month. Substituting $t = 48$ into Eq. (8.41), we obtain the estimate

$$EX_{48} = 5.45 + 0.0970(48) = 10.1$$

for the May 1990 CM1 index value. Since $R^2 = 83.0\%$, the regression equation accounts for 83% of the total variation in our CM1 index data. This gives us a fairly high level of confidence in our estimate of EX_{48}. Of course, the actual value of X_{48} will differ due to random fluctuations. More details about the likely magnitude of these fluctuations will be provided below in the sensitivity analysis.

Finally, step 5. We have concluded that the CM1 index shows a general trend, increasing by about 0.097 points per month. This figure is based on historical observations over the last three years. Projecting on this basis, we obtain the estimate of 10.1 for the May 1990 index figure. This is about 1.1 points higher than the May 1989 index, so in 1990 the author should expect his ARM payments to increase again.

The most important sensitivity analysis question here is the amount of random fluctuation in X_t. We are assuming the linear regression model in Eq. (8.30), where ε_t is mean zero normal. Our regression package estimated the standard deviation $\sigma \approx 0.4812$ on the basis of the data. In other words, $\varepsilon_t/0.4812$ is approximately standard normal. About 95% of the data points were no more than $\pm 2\sigma$ away from the line in Eq. (8.41). If this is representative of the magnitude of future fluctuations, then we would expect X_{48} to lie between $10.1 \pm 2\sigma$ with

95% confidence; i.e., we should see

$$9.1 \leq X_{48} \leq 11.1.$$

There is also a more sophisticated method built into the statistics package that takes into account the additional uncertainty involved in estimating EX_{48}. This method yields $9.02 \leq X_{48} \leq 11.19$ at the 95% confidence level. See Fig. 8.11.

Next we consider the sensitivity of our model to unusual data values. We are assuming the linear regression model from Eq. (8.30). Most of the time the random error ε_t will be small, but there is a small probability that ε_t will be rather large, so that one or more of our data points may lie far off the regression line. We need to consider the sensitivity of our procedure to such anomalous points, which are called *outliers*.

It is not hard to show that the regression line for a set of data points (t_1, y_1), …, (t_n, y_n) will always pass through the point $(\bar{t}, \bar{y})$ defined by

$$\begin{aligned} \bar{t} &= \frac{t_1 + \cdots + t_n}{n} \\ \bar{y} &= \frac{y_1 + \cdots + y_n}{n}. \end{aligned} \tag{8.42}$$

In our model we have $\bar{t} = 19$ and $\bar{y} = 7.29$. The regression procedure selects the best–fitting line through the point (19, 7.29). Since the essence of the procedure is to minimize the vertical distance between the regression line and the data points, an outlier will tend to pull the regression line toward itself, regardless of the location of the other data points. The situation gets worse as n gets smaller, because then each individual data point has more influence. It also gets worse as the distance from the base point $(\bar{t}, \bar{y})$ gets larger, since points farther out on the regression line have more leverage.

In Fig. 8.11 the statistical package Minitab has flagged the data point (1, 6.73) as being unusual. If we plug $t = 1$ into the regression equation, we get

$$\hat{y}_1 = 5.45 + 0.0970(1) = 5.547.$$

The vertical distance or *residual* $y_1 - \hat{y}_1$ is 1.18, which means that this data point is about 2.6 standard deviations above the regression line. In order to ascertain the sensitivity of our model to this outlier, we repeat the regression calculation, leaving out the data point (1, 6.73). Figure 8.13 shows the results of this sensitivity run.

The new regression equation is

$$EX_t = 5.30 + 0.103t$$

with $R^2 = 86.2\%$. The predicted value $EX_{48} = 5.30 + 0.103(48) = 10.24$ is our new estimate of the CM1 index on May 1990. The new residual standard deviation is 0.438450, so we expect that the CM1 index on May 1990 will be between $10.24 \pm 2(0.44)$. In other words, we are 95% sure that we will see

$$9.36 < X_{48} < 11.12.$$

```
The regression equation is
cm1 = 5.30 + 0.103 t

Predictor       Coef   SE Coef      T      P
Constant      5.3045    0.1554  34.13  0.000
t           0.102634  0.007034  14.59  0.000

S = 0.438450   R-Sq = 86.2%   R-Sq(adj) = 85.8%
```

Figure 8.13: Sensitivity analysis for the ARM problem using the statistical package Minitab.

The more sophisticated prediction interval computed by Minitab (not shown in Figure 8.13) is a bit wider: $9.24 \leq X_{48} \leq 11.22$. Either way, these conclusions are about the same as before, so we conclude that our model is not too sensitive to this outlier.

The main robustness issue here concerns our choice of a linear model in Eq. (8.30). More generally, we might assume that

$$X_t = f(t) + \varepsilon_t, \tag{8.43}$$

where $f(t)$ represents the true value of one–year U.S. Treasury bonds at time t, and ε_t represents fluctuations in the market. In this more general setting the linear regression model represents a linear approximation

$$f(t) \approx a + bt, \tag{8.44}$$

which is good near the base point $(\bar{t}, \bar{y})$. In Figure 8.11, Minitab has flagged the point $t = 48$ as being far away from the center point $t = 19$. We have data for $1 \leq t \leq 37$ and strong evidence of a linear relationship over this interval ($R^2 = 83\%$). In other words, the linear approximation in Eq. (8.44) involves at most a small percent error over this interval. As we move away from this interval, however, it must be expected that the error involved in this linear approximation gets worse. Another robustness issue arises from our assumption that the random errors ε_t are independent and identically distributed. A more complex model might take into account the dependence between these random variables. We will explore this issue in Section 8.5.

Our linear regression model is a simple example of a *time series model.* A time series model is a stochastic model of one or more variables that evolve over time. Most economic forecasting is done using time series models. More complex time series models represent the interaction of several variables and dependence in the random fluctuations of these variables. Time series analysis is a branch of statistics. In the next section, we provide an introduction to the essentials of time series analysis. For more details about time series models, a good reference is Box et al. (1976).

8.4 Time Series

A time series is a stochastic process that varies over time, usually observed at fixed intervals. Daily temperature and rainfall, monthly unemployment levels, and annual income are some typical examples of time series. The basic tool for modeling time series is linear regression, introduced in Section 8.3. For that reason, this section may be considered as a follow-up to Section 8.3 that introduces some additional applications and methods. Indeed, the example we consider next is an extension of the ARM problem from that section. This section requires a numerical implementation of multiple regression; i.e., linear regression with more than one predictor. This is available in a statistical package (e.g., Minitab, SAS, SPSS) or a spreadsheet (e.g., Excel). Since it is not reasonable to perform these problems by hand, a full set of computational formulas will not be given.

Example 8.6. Reconsider the ARM problem of Example 8.5 but now consider the relationship between the mortgage index at different times. Answer the same question as before: Estimate the value of the CM1 index in May 1990 using the data provided on the CM1 index values from June 1986 to June 1989.

We will use the five-step method. Step 1 is the same as before, except that now we also want to take into account the dependence between the CM1 index at different times. We are attempting to estimate the future trend in a variable that exhibits a tendency to grow with time, along with some random fluctuation. Let X_t denote the U.S. Treasury one–year Constant Maturity (CM1) index at time t months after May 1986. A graph of X_t for $t = 1, \ldots, 37$ was shown in Figure 8.10. We want to estimate X_{48}. We will assume that X_t depends on the time t, the previous values $X_{t-1}, X_{t-2}, \ldots$ and a random element. Then we want to predict the average EX_{48} along with a suitable estimate of uncertainty.

Step 2 is to select the modeling approach. We will model this problem as a time series and fit an autoregressive model.

> A time series is a sequence of random variables $\{X_t\}$ that varies over time $t = 0, 1, 2, \ldots$ according to some random pattern. The key to time series modeling is to recognize the pattern. A typical assumption is that the pattern involves a *trend* added to a *stationary time series*. The trend is a non-random function that varies over time, and it represents the mean value of the series. Once the trend is removed, we are left with a mean zero time series, and we want to model its dependence structure. The simplest case is where the remaining time series consists of independent random variables. However, it is typical to find dependence between these variables. Dependence is measured in terms of the *covariance*. For two random variables X_1 and X_2, the covariance
>
> $$\mathrm{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)],$$
>
> where $\mu_i = E(X_i)$, the expected value or mean. The covariance measures the linear relation between the two variables. If X_1 and

X_2 are independent, then the covariance $\mathrm{Cov}(X_1, X_2) = 0$. A positive covariance indicates that a higher–than–average value of X_1 is likely to be found in the company of a higher–than–average value of X_2. Likewise, a low value of X_1 will typically be found along with a low value of X_2. For example, if X_1 is the income of an individual and X_2 is their income tax, then $\mathrm{Cov}(X_1, X_2)$ would be positive. You cannot infer the income tax knowing only the income, but it is a good bet that someone with a high income pays more tax, and someone with a low income pays less. Mathematically, $(X_1 - \mu_1)$ is the deviation of income from the average, and likewise for $(X_2 - \mu_2)$. The covariance averages the product of these. If one tends to be positive when the other is positive, and negative when the other is negative, then the covariance is positive, indicating a *positive relation*. For another example, the median price X_1 of a home in a town is negatively correlated to the percentage of families X_2 owning their own home. Here μ_1 is the median home price over all towns, and μ_2 is the average percentage of families owning their own home, taken over all the towns. When $X_1 - \mu_1$ is positive, then $X_2 - \mu_2$ is likely to be negative, and vice versa, so their average $\mathrm{Cov}(X_1, X_2)$ will be negative, indicating a *negative relation* between these two variables. We also note that correlation only captures a linear relation. Consider the case where X_1 is the air pressure in a car tire and X_2 is the tread life. If X_1 is near the mean μ_1, the recommended air pressure, then X_2 will be the highest. If X_1 is either smaller or larger than its mean, then X_2 will decrease. The covariance does not capture this kind of dependence. Finally, we note that the covariance is also a generalization of the variance, in the sense that $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

A close cousin of the covariance is the *correlation*

$$\rho = \mathrm{corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} = E\left[\frac{(X_1 - \mu_1)}{\sigma_1}\frac{(X_2 - \mu_2)}{\sigma_2}\right]$$

which is a dimensionless version of the covariance. Here $\sigma_i^2 = \mathrm{Var}(X_i)$ is the variance, so that σ_i is the standard deviation of the random variable X_i. Since both μ_i and σ_i have the same units as X_i, the units cancel, leaving a dimensionless measure of dependence. Again, if X_1 and X_2 are independent, then $\mathrm{corr}(X_1, X_2) = 0$, and we say that X_1 and X_2 are uncorrelated. It can also be shown that the correlation satisfies $-1 \le \rho \le 1$ in all cases, the extreme cases $\rho = \pm 1$ corresponding to the case of perfect dependence where X_2 is a linear function of X_1. If $\rho > 0$, we say that X_1 and X_2 are *positively correlated*, and if $\rho < 0$, we say that they are *negatively correlated*.

Correlation is a useful measure of dependence in a time series. In that context we call $\rho(t, h) = \mathrm{corr}(X_t, X_{t+h})$ the *autocorrelation* *function* of the time series. It measures the serial dependence between the times series at different times. The time series is called *stationary* (or sometimes *weakly stationary*) if the mean $E(X_t)$ and the autocorrelation function $\rho(h, t)$ are constant over time. In time series analysis, it is often necessary to *detrend* the series to get something that is stationary. In Example 8.5 we detrended the CM1 times series X_t by using regression to identify a linear trend $a + bt$, leaving a zero mean (centered) error term ε_t that we modeled as independent and identically distributed. In the context of time series analysis, this is called *random noise*. It is the simplest centered times series. More generally, one might hope that the centered series is at least stationary, with a correlation structure that remains the same over time. There are various tests for stationarity, but the simplest is just to plot the errors ε_t over time to see whether they appear to follow a consistent (random) pattern. One typical indication of non-stationarity would be a widening or narrowing of the distribution of ε_t over time. This is called *heteroscedasticity*, which just means that the variance changes. Once we are satisfied that a centered time series is stationary, we attempt to model its covariance structure. The simplest useful model for this is called an *autoregressive process*

$$X_t = a + bt + c_1 X_{t-1} + \cdots + c_p X_{t-p} + \varepsilon_t \tag{8.45}$$

where for convenience, we have also included the trend. The parameter p is called the *order* of the autoregressive process, sometimes abbreviated as AR(p). In the context of linear regression, we can fit the parameters of an autoregressive process by regressing the observations X_t against multiple predictors. The first is the time t, as in Section 8.3. The remaining predictors are the previous observations $X_{t-1}, \ldots, X_{t-p}$. Now the trick is to choose a reasonable value for the parameter p, and there are two main ways to do this. One way is to look at the R^2 value, which measures how much of the variability in X_t is captured by the predictors. Since any additional data will give at least a slightly better prediction, adding a predictor will always make R^2 increase. However, a very small increase is not worth the bother of carrying another predictor, so one could just add predictors X_{t-1}, X_{t-2}, and so forth one by one, until the additional improvement in R^2 is minimal. Some packages also output an adjusted R^2 value that includes a penalty for an increased number of predictors. If this is available, one can simply add predictors until the adjusted R^2 begins to decrease (or more generally, consider several models and pick the one with the largest adjusted R^2). Such packages also typically list the *sequential sums of squares* that can be interpreted as an extension of the R^2 statistic. Recall that for a simple (one predictor) linear regression model, the formula (8.39) for R^2 is the sum of the squares of the regressions $(\hat{y}_i - \bar{y})$ divided by the sum of the squares of the total variations $(y_i - \bar{y})$. The sequential sums of squares are just the individual components in the sum of the squares of the regressions, one predictor at a time, to measure the additional variation captured by each additional predictor. Looking at the sequential sums of squares is another way to measure the additional value of another predictor, essentially equivalent to looking at the change in the R^2 value. Additional information is contained in the *p–values* for each of the estimated regression coefficients a, b, and c_i. The p–value indicates the likelihood that this parameter value could have occurred by chance, even though the predictor does not belong in the model (or equivalently, it belongs in the model with a coefficient of zero). Hence, a small p–value (say, $p < 0.05$) indicates strong evidence that the predictor belongs in the model. However, this is less important then the R^2 indicator, since a predictor can be statistically significantly related to the dependent variable X_t we are trying to predict, and still add just a small amount of information. Hence, we might not consider it worthwhile to include. This is the principle of *parsimony*: Make the model as simple as possible, without sacrificing the ability to predict.

The second method for autoregressive modeling is to consider the *residuals* from the time series model, which are simply our estimates of the error term ε_t. Once we have determined our estimates of the parameters a, b, and c_i from the regression, we can use the formula

$$\varepsilon_t = X_t - (a + bt + c_1 X_{t-1} + \cdots + c_p X_{t-p})$$

to estimate the errors. Since our goal is to include enough predictors in the model to capture the dependence structure, we would like the resulting ε_t sequence to be an uncorrelated noise sequence. We can check this by computing the autocorrelation function of the residuals. Most statistical packages will automatically compute the residuals and their autocorrelation function, as well as p–values or error bars for the autocorrelation function $\rho(h)$. The error bars indicate the likely values of $\rho(h)$ in the case of a noise sequence, and values outside these bars (also indicated by a low p–value, typically less than 0.05) indicate a statistically significant correlation. Since the autocorrelation function assumes a stationary process, it is also advisable to check for this using graphical displays of the residuals.

Step 3 is to formulate the model. We will model the CM1 index X_t at time t months after May 1986 as an autoregressive time series with a linear trend. Hence, we are assuming that Eq. (8.45) holds for some constants a, b, $c_1, \ldots, c_p$ and some noise sequence ε_t. In order to choose a suitable value for the parameter p, we will consider a sequence of increasingly complex models $p = 0, 1, 2, \ldots$ until we achieve a satisfactory result, indicated by a model with (hopefully) a small number of predictors that appears to have an uncorrelated noise sequence of residuals. Then we can proceed to estimate (forecast) the

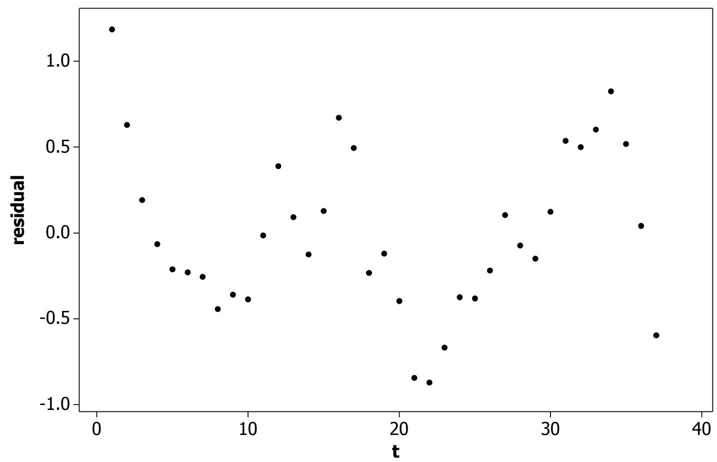

Figure 8.14: Graph of CM1 residuals ε_t versus time for the simple regression model of equation (8.30).

value X_t of the CM1 index at month $t = 48$, which is May 1990, along with appropriate error bounds.

Step 4 is to solve the problem. We will use the statistical package Minitab which has convenient facilities for multiple regression, time series analysis, and graphical display. We begin by examining in more detail the results of the modeling in Example 8.5. There, we fit a simple linear regression model of the form $X_t = a + bt + \varepsilon_t$. The results are summarized in Figure 8.11. The best–fit regression line is given by $a = 5.45$ and $b = 0.097$, indicating an upward trend in the CM1 index. The statistic $s = 0.48$ estimates the standard deviation of the errors ε_t, and the R^2 statistic indicates that the trend predicts 83% of variations in the CM1 index X_t. The p–values for both a and b are given as 0.000, indicating strong statistical evidence that these parameters are different from zero. Our subsequent analysis and predictions were based on the simple regression model, which assumes that the errors ε_t form a noise sequence, independent and identically distributed. We will now test this assumption using graphical displays and the correlation function.

The residuals or estimated errors are computed as described in Section 8.3, using the estimated values of a and b and the equation $\varepsilon_t = X_t - (a + bt)$. The predicted values $\hat{y}_t = a + bt$ are plotted in Figure 8.12 along with the original data. The residuals are simply the vertical deviations $(y_t - \hat{y}_t)$ of the data values from the regression line, where $y_t = X_t$ is the t-th observation of the CM1 index. For example, the second data value is $y_2 = 6.27$ and the fitted value is $\hat{y}_2 = 5.45 + 0.097(2) = 5.64$, so the residual is $y_2 - \hat{y}_2 = 0.63$, indicating the the second data point in Figure 8.12 is 0.63 units above the regression line.

Figure 8.14 shows a plot of the residuals. They were computed in Minitab by clicking the `Storage` button in the regression window and checking the box labelled `Residuals`. The graph was also prepared in Minitab using the command `Graph > Scatterplot`. The data appear stationary in the sense that the spread of values does not seem to increase or decrease with t. However, it does appear that there may be some serial dependence, particularly for $t \geq 20$ where there seems to be an upward trend. This could indicate many things, including nonstationarity, a change in the dependence structure of the series, or some correlation.

To investigate further, we compute the autocorrelation function of the residuals. Figure 8.15 shows the results of this calculation using the Minitab command `Stat > Time Series > Autocorrelation` applied to the residuals stored as part of the regression calculation. The vertical bars indicate the correlation function $\rho(h) = \text{corr}(\varepsilon_t, \varepsilon_{t+h})$ as a function of the time lag $h = 1, 2, 3, \ldots$ and the dotted curve represents the 95% error bars. The autocorrelation function plotted here is, of course, a statistical estimate of the model autocorrelation, and the error bars show the range of normal variation for this statistical estimate for an uncorrelated noise sequence. Hence, a value outside the bars indicates strong statistical evidence for a nonzero autocorrelation. In Figure 8.15, the first value $\rho(1)$ is well outside the error bars. This is an indication of serial correlation in the residuals ε_t from our simple regression, evidence that a more complex model is needed to obtain a simple uncorrelated (white) noise sequence. This kind of strong positive correlation can cause a pattern that looks like a trend, as seen in Figure 8.14, since a large positive value of ε_t makes it more likely that the next value is large, and so forth.

Now we proceed to consider a more complicated autoregressive times series model (8.45) for the CM1 index time series data. Our goal is to find a number of predictors p in terms of the past history of the process that is in some sense optimal. We begin by repeating the regression procedure with $p = 1$ additional predictor X_{t-1}. First, we prepare another column of data X_{t-1}; that is, the CM1 index shifted downward by one place. Use a simple copy and paste or the Minitab command `Stat > Time Series > Lag`. It is necessary to leave off the last CM1 index value in the lag one column X_{t-1}, since the statistical package requires predictors to be the same data length as the predicted data (and, of course, the first entry is blank or missing, in Minitab this is denoted by a $*$ in that data cell). The details vary with different packages, but the steps are similar. Now we repeat the regression command and store the residuals.

Figure 8.16 displays portions of the Minitab computer output. The regression equation is $X_t = 1.60 + 0.033t + 0.698X_{t-1}$, where X_t is the CM1 index t months after May 1986. This indicates an upward trend along with some serial dependence between the CM1 index on subsequent months. The statistic $R^2 = 94.1\%$ indicates that the combination of trend and the CM1 index from last month predicts 94.1% of the variations in the CM1 index this month. This is a significant improvement over the value $R^2 = 83.0\%$ for the simple linear regression model of Example 8.5.

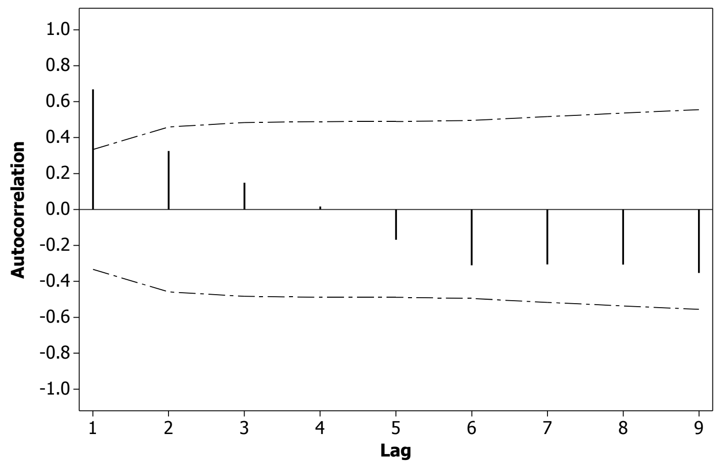

Figure 8.15: Autocorrelation function for CM1 residuals ε_t versus lag for the simple regression model of equation (8.30).

The adjusted $R^2 = 93.8\%$ here is also higher than the 82.6% figure from Example 8.5, additional evidence that the autoregressive model is superior. The analysis of variance statistics give more detail on the R^2 calculation. Recalling the formula (8.39), we see that R^2 is the ratio of two sums of squares. The sum of squared regressions $\sum_i (\hat{y}_i - \bar{y})^2 = 44.676$ and the sum of squares of total variation is $\sum_i (y_i - \bar{y})^2 = 47.460$. The ratio of these two is $R^2 = 44.676/47.460 = 0.941$. The sequential sums of squares table indicates that 40.924 of the 44.676 residual sum of squares comes from the first predictor t, and an additional 3.752 comes from the addition of a second predictor X_{t-1}. If we had listed the two predictors in the reverse order, then these two figures would change, since the two predictors t and X_{t-1} are not completely independent, but they would still add to 44.767. We also note that the p-value for the constant $a = 1.5987$ is 0.008, the p–value for the coefficient $b = 0.03299$ of the predictor t is 0.007, and the p–value for the coefficient $c_1 = 0.6977$ of X_{t-1} is 0.000. This is additional evidence that all these coefficients are statistically significantly different than zero and should be included in the model.

Figure 8.17 shows the fitted model $1.60 + 0.033t + 0.698X_{t-1}$ plotted along with the data X_t to check the quality of the fit. It appears that the fit is better than the simple regression line; compare Figure 8.12. Now a prediction of X_{48} can be accomplished by iterating the equation $X_t = 1.60 + 0.033t + 0.698X_{t-1}$

```
The regression equation is
CM1 = 1.60 + 0.0330 t + 0.698 X(t-1)

Predictor      Coef  SE Coef     T      P
Constant     1.5987   0.5652  2.83  0.008
t           0.03299  0.01144  2.88  0.007
X(t-1)       0.6977   0.1046  6.67  0.000

S = 0.290471   R-Sq = 94.1%   R-Sq(adj) = 93.8%

Analysis of Variance

Source          DF      SS      MS       F      P
Regression       2  44.676  22.338  264.75  0.000
Residual Error  33   2.784   0.084
Total           35  47.460

Source  DF  Seq SS
t        1  40.924
X(t-1)   1   3.752
```

Figure 8.16: Autoregressive model for the ARM problem using the statistical package Minitab.

starting at $t = 37$ (e.g., use a hand calculator or a spreadsheet). This yields

$$\begin{aligned}
X_{38} &= 1.60 + 0.0330(38) + 0.698(8.44) = 8.75 \\
X_{39} &= 1.60 + 0.0330(39) + 0.698(8.75) = 8.99 \\
&\vdots \\
X_{48} &= 1.60 + 0.0330(48) + 0.698(10.16) = 10.28
\end{aligned}$$

so we predict that the value of the CM1 index for May 1990 will be 10.28. Using the model $X_t = 1.60 + 0.033t + 0.698X_{t-1} + \varepsilon_t$ along with the fact that the estimated standard deviation of ε_t is 0.290471, we predict (with 95% certainty) that the value of the CM1 index on May 1990 will be between $10.28 \pm 2(0.29)$, or in other words, between 9.7 and 10.9. This is a more precise estimate than Example 8.5 because the regression model fits more closely, giving a smaller error standard deviation. Graphically, the standard deviation is simply the typical vertical variation of the data points from the prediction line, so a closer fit gives a smaller standard deviation. However, the bottom line is the same as before: The author's ARM interest rate is likely to increase again in 1990. In fact, this improved model gives a slightly higher estimate for the May 1990 CM1 index.

Next we examine the model residuals to determine whether they resemble an

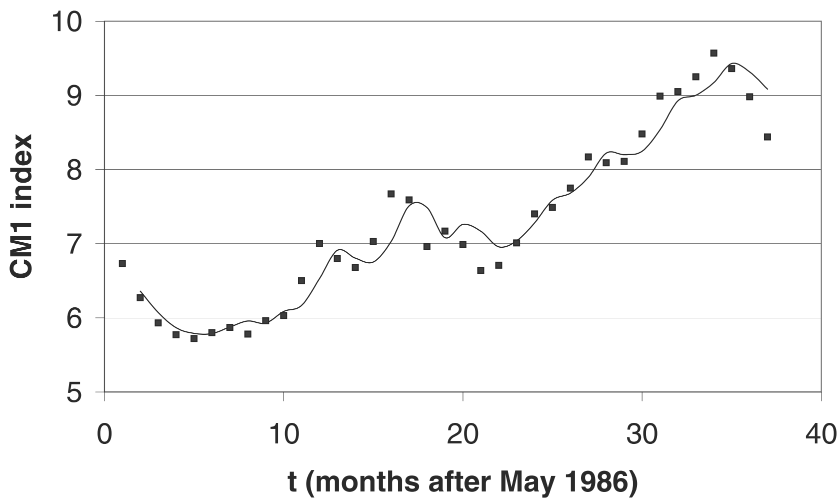

Figure 8.17: Graph of CM1 index versus time t showing the fitted autoregressive model of equation (8.45) with $p = 1$.

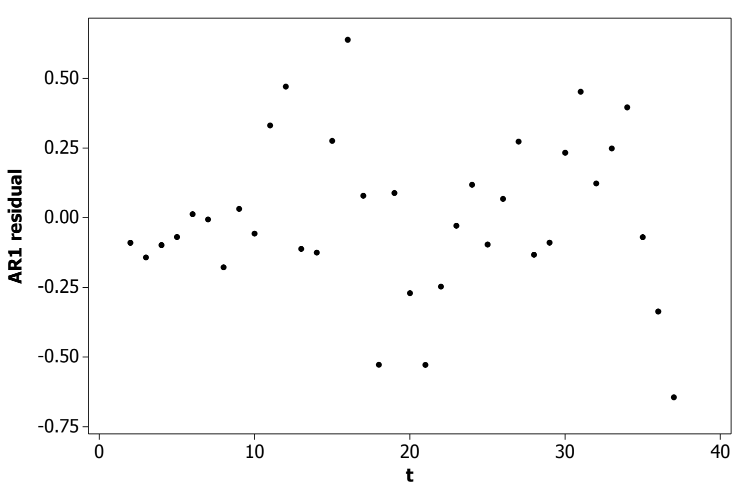

Figure 8.18: Graph of CM1 residuals ε_t versus time for the autoregressive model of equation (8.45) with $p = 1$.

uncorrelated (white) noise sequence. Figure 8.18 shows a plot of the residuals for this AR(1) model versus time. This plot is more satisfactory than Figure 8.14, with no significant evidence of serial correlation. We can also examine the autocorrelation function of these residuals as we did before, and this plot (not shown) indicates no serial correlation, since all the autocorrelations are within the error bars. Hence, it is reasonable to conclude that the AR(1) model (8.45) with $p = 1$ captures nearly all of the dependence in the CM1 time series. To validate this conclusion, we may also consider a further regression model with three predictors: t, X_{t-1}, and X_{t-2}. We leave the details for an exercise (see Exercise 8.19).

Finally, we come to Step 5. Example 8.6 may be considered as part of the robustness analysis of Example 8.5. There we predicted that the value of the CM1 index on May 1990 would be 10.1, about 1.1 points higher than the May 1989 index value of 8.98. The more sophisticated model considered here gives a refined estimate of 10.2, a little higher. We also recall that in Example 8.5 we produced a prediction interval of (9.1,11.1) with 95% confidence. The refined model yields a significantly tighter interval of (9.7,10.9). Putting this all together, we expect that the CM1 index for May 1990 is likely to be a bit more than one point higher than the May 1989 value, and we are reasonably sure it will go up by at least 3/4 of a percent.

Sensitivity analysis could include many factors. For example, two of the observations at $t = 16, 37$ were flagged by Minitab for large residuals (both were more than 2.2 standard deviations from the mean). Hence, we could repeat the analysis with these values deleted, to see if this makes any significant difference. We could consider a different trend function like $a + bt + ct^2$ (see Exercise 16) or at^b (see Exercise 18 where we apply this trend to the response time data for the facility location problem). We could also add more predictors $X_{t-2}, X_{t-3}, \ldots$ and check to see if this makes a big difference (see Exercise 8.19). The possibilities are literally endless, and hence, some judgment is called for. This is where the principle of parsimony comes in. Our goal is to get a reasonable estimate of the CM1 index for May 1990, based on the available data. Refining the simple regression model of Example 8.5 to the AR(1) model of Example 8.6 was probably worthwhile, not so much for the improved point estimate of 9.2 instead of 9.1, but rather for the significantly tighter prediction interval. If we only cared about the point estimate, then simple linear regression is probably good enough. Whether it is worth while to continue in this vein, to consider more alternative models with more predictors and/or more complicated trend functions, is less clear. In the real world, sensitivity analysis can continue as long as time and funds permit, but at some point the intelligent modeler will declare victory and move on to a new challenge.

Many interesting robustness questions are important in the analysis of real–world time series. One question is the trend. While we use a linear trend, other options such as a higher order polynomial (see Exercise 16) or a nonlinear trend (see Exercise 18) can also be considered. Adding more parameters will always improve the fit, so care is required. Examining the adjusted R^2 value is one way to avoid over-parameterization. Often the choice of trend function will depend on the applications. For example, for income or population data, one might expect exponential rather than linear growth. Another important issue is change point analysis: Did the underlying correlation structure or trend function of the time series change at some point during the data collection period? For example, this is an important part of the debate over global warming. Time series is a growing field both in applications and theory. A good place to start learning more about the underlying theory is Brockwell and Davis (1991).

8.5 Exercises

1. Reconsider the inventory problem of Example 8.1, but now suppose that the store policy is to order more aquariums if there are less than two left in stock at the end of the week. In either case (zero or one remaining) the store orders enough to bring the total number of aquariums in stock back up to three.

   (a) Calculate the probability that the demand for aquariums in a given week exceeds the supply. Use the five-step method, and model as a Markov chain in steady state.

   (b) Perform a sensitivity analysis on the demand rate λ. Calculate the steady–state probability that demand exceeds supply, assuming $\lambda = 0.75,\ 0.9,\ 1.0,\ 1.1,$ and 1.25, and display in graphical form, as in Fig. 8.5.

   (c) Let p denote the steady–state probability that demand exceeds supply. Use the results of part (b) to estimate $S(p,\,\lambda)$.

2. (Requires a computer algebra system) Reconsider the inventory problem of Example 8.1. In this problem we will explore the sensitivity of the probability p that demand exceeds supply to the demand rate λ.

   (a) Draw the state transition diagram for an arbitrary λ. Show that Eq. (8.11) is the appropriate state transition probability matrix for this problem.

   (b) Write down the system of equations analogous to Eq. (8.10) that must be solved to get the steady–state distribution for a general λ. Use a computer algebra system to solve these equations.

   (c) Let p denote the steady–state probability that demand exceeds supply. Use the results of part (b) to obtain a formula for p in terms of λ. Graph p versus λ for the range $0 \le \lambda \le 2$.

   (d) Use a computer algebra system to differentiate the formula for p obtained in part (c). Calculate the exact sensitivity $S(p,\,\lambda)$ at $\lambda = 1$.

3. Reconsider the inventory problem of Example 8.1, but now suppose that the inventory policy depends on recent sales history. Whenever inventory

drops to zero, the number of units ordered is equal to two plus the number sold over the past week, up to a maximum of four.

(a) Determine the steady–state probability distribution of the number of aquariums in stock. Use the five-step method, and model as a Markov chain.

(b) Determine the steady–state probability that demand exceeds supply.

(c) Determine the average size of a resupply order.

(d) Repeat parts (a) and (b), but now suppose that weekly demand is Poisson with a mean of 2 customers per week.

4. Reconsider the inventory problem of Example 8.1, but now suppose that three additional aquariums are ordered any time that there are less than two in stock at the end of the week.

(a) Determine the probability that demand exceeds supply on any given week. Use the five-step method, and model as a Markov chain in steady state.

(b) Use the steady–state probabilities from part (a) to calculate the expected number of aquariums sold per week under this inventory policy.

(c) Repeat part (b) for the inventory policy in Example 8.1.

(d) Suppose that the store makes a profit of \$5 per 20–gallon aquarium sold. How much would the store gain by implementing the new inventory policy?

5. For the purposes of this problem, we will consider the stock market to be in one of three states:

1	Bear market
2	Strong bull market
3	Weak bull market

Historically, a certain mutual fund gained -3%, 28%, and 10% annually when the market was in states 1, 2, and 3 respectively. Assume that the state transition probability matrix

$$P = \begin{pmatrix} 0.90 & 0.02 & 0.08 \\ 0.05 & 0.85 & 0.10 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$$

applies to the weekly change of state in the stock market.

(a) Determine the steady–state distribution of market state.

(b) Suppose that \$10,000 is invested in this fund for ten years. Determine the expected total yield. Does the order of state transitions make any difference?

(c) In the worst–case scenario, the long–run expected proportion of time in each market state is 40%, 20%, and 40%, respectively. What is the effect on the answer to part (b)?

(d) In the best–case scenario, the long–run expected proportion of time in each market state is 10%, 70%, and 20%, respectively. What is the effect on the answer to part (b)?

(e) Does this mutual fund offer a better investment opportunity than a money market fund currently yielding about 8%? Consider that the money market fund offers a lower risk.

6. A Markov chain model of floods uses the state variable $X_n = 0, 1, 2, 3, 4$ where state 0 means the average daily flow is below 1,000 cubic feet per second, state 1 is 1,000–2,000 cfs, state 2 is 2,000–5,000 cfs, 3 is 5,000–10,000 cfs, and 4 is over 10,000 cfs. The state transition probability matrix for this model is

$$P = \begin{pmatrix} 0.9 & 0.05 & 0.025 & 0.015 & 0.01 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{pmatrix}.$$

(a) Draw the state transition probability diagram for this model.

(b) Find the steady–state probability distribution for this model.

(c) How often are severe floods (over 10,000 cfs) expected to occur?

(d) A reservoir used for drought storage depends on the flow of this river. When the flow is over 5,000 cfs, the reservoir is allowed to store 1,000 acre–feet per day. When the flow is under 1,000 cfs, the reservoir is required to release 100 acre–feet per day back into the river. Find the expected annual number of acre–feet of water stored in the reservoir. Is this number positive or negative? What does this mean?

7. This exercise shows the equivalence of the two formulations of a Markov process model.

(a) Suppose that $T_{i1}, \ldots, T_{im}$ are independent random variables and that T_{ij} is exponential with rate parameter $a_{ij} = p_{ij}\lambda_i$. Assume that $\sum p_{ij} = 1$ and $\lambda_i > 0$. Show that

$$T_i = \min(T_{i1}, \ldots, T_{im})$$

has an exponential distribution with rate parameter λ_i. [Hint: Use the fact that

$$\Pr\{T_i > x\} = \Pr\{T_{i1} > x, \ldots, T_{im} > x\}$$

for all $x > 0$.]

(b) Suppose that $m = 2$ so that $T_i = \min(T_{i1}, T_{i2})$ and show that $\Pr\{T_i = T_{i1}\} = p_{i1}$. [Hint: Use the fact that

$$\begin{aligned}\Pr\{T_i = T_{i1}\} &= \Pr\{T_{i2} > T_{i1}\} \\ &= \int_0^\infty \Pr\{T_{i2} > x\}\, f_{i1}(x)\, dx,\end{aligned}$$

where $f_{i1}(x)$ is the probability density function of the random variable T_{i1}.]

(c) Use the results of (a) and (b) to show that in general $\Pr\{T_i = T_{ij}\} = p_{ij}$.

8. Reconsider the forklift problem of Example 8.3, but now suppose that a second mechanic is called in whenever two or more trucks need repair.

   (a) Determine the steady–state distribution of the number of trucks needing repair. Use the five-step method and model as a Markov process.

   (b) Use the results of part (a) to calculate the steady–state expected number of forklift trucks in the repair facility, the probability that the first mechanic is busy, and the probability that the second mechanic is called in. Compare to the results obtained for one mechanic in Example 8.3.

   (c) The cost of the second mechanic is \$250 per day, and the mechanic is only paid for the days worked. Another option which can be used in case of backlog (two or more trucks in repair) is to lease replacement vehicles for customers whose trucks are in repair, at a cost of \$125 per week per vehicle. Which of the two plans is most cost-effective?

   (d) There is some uncertainty as to the actual cost of bringing in a second mechanic during periods of backlog. What is the minimum cost per day for a second mechanic that makes leasing a better alternative?

9. Five locations are connected by radio. The radio link is active 20% of the time, and there is no radio activity the remaining 80% of the time. The main location sends radio messages with an average duration of 30 seconds, and the remaining four locations send messages that average 10 seconds in length. Half of all radio messages originate at the main location, with the remaining proportion equally divided among the other four locations.

   (a) For each location, determine the steady–state probability that this location is sending a message at any given time. Use the five-step method, and model as a Markov process.

   (b) A monitoring station samples from the radio emissions on this network once every five minutes. How long on average does it take until the monitor finds a message in progress from one particular location?

(c) The monitor can identify the source of a radio message if the message lasts at least three seconds after the monitor begins listening. How long does it take until the monitor locates one particular location?

(d) Perform a sensitivity analysis to see how the results of part (c) are affected by the percent utilization of the radio frequency (currently 20%).

10. A gasoline service station has two pumps, each of which can service two cars at a time. If all pumps are busy, cars will queue up in a single line to await a free pump. Since the station operates in a competitive environment, it can be expected that customers encountering long lines at this station will take their business elsewhere.

    (a) Construct a model that can be used to predict both the steady–state probability of a waiting line and the expected length of the line. Use the five-step method, and model as a Markov process. You will have to make some additional assumptions about customer demand, service time, and balking (refusing to join a queue).

    (b) Use the model of part (a) to estimate the fraction of potential business lost because of customer balking. Consider a range of possible levels of customer demand.

    (c) What is the easiest way to infer the level of customer demand (i.e., potential sales) from data to which the station manager has access?

    (d) Under what circumstances would you recommend that the station purchase additional pumps?

11. A certain form of one–celled creature reproduces by cell division, producing two offspring. The mean lifetime before cell division is one hour, and each individual cell has a 10% chance of dying before it can reproduce.

    (a) Construct a model that represents the evolution of population size over time. Use a Markov process, and draw a rate diagram.

    (b) Describe in general terms what you would expect to happen to the population level over time.

    (c) What is the problem with applying steady–state results to this model?

12. Table 8.2 gives per capita income in 1982 dollars and population density in population per square miles for the ten poorest counties in Asia (source: Webster's New World Atlas, (1988)).

    (a) Does the data support the proposition that prosperity is linked to population density? Use linear regression to obtain a formula that predicts per capita income as a linear function of population density.

    (b) What percentage of the total variation in per capita income can be attributed to variations in population density?

Country	Per Capita Income (in 1982 dollars)	Population Density (pop. per sq. mi.)
Nepal	168	290
Kampuchea	117	101
Bangladesh	122	1740
Burma	171	139
Afghanistan	172	71
Bhutan	142	76
Vietnam	188	458
China	267	284
India	252	578
Laos	325	45

Table 8.2: Per capita income versus population density for the 10 poorest Asian nations.

(c) What is the effect on your answers to parts (a) and (b) if we leave out of our analysis the country (Bangladesh) with population density of more than 1,000 per square mile?

(d) On the basis of your regression model, estimate the potential benefit for the citizens of a poor Asian country that manages to reduce its population by 25%.

13. Reconsider Exercise 12(a), but now obtain the regression line by solving the underlying optimization problem by hand. Letting t_i denote the population density and y_i the per capita income of country i, the goodness–of–fit for a candidate regression line

$$y = a + bt$$

is given by Eq. (8.33) in the text. Plug in the data points $(t_i,\, y_i)$ and compute the function $F(a,\, b)$. Then obtain the best–fitting line by minimizing $F(a,\, b)$ over the set of all $(a,\, b) \in \mathbb{R}^2$.

14. Reconsider the ARM problem of Example 8.5. Suppose that only the data for the period June 1986 to June 1988 were known, and we were attempting to predict the May 1989 value of the CM1 index.

(a) Use a computer or calculator implementation of linear regression to obtain the regression line for this data.

(b) What is the predicted value of the May 1989 CM1 index according to the regression model of part (a)?

(c) What is the value of R^2 for the model in part (a)? How would you interpret this value?

(d) Compare to the actual value of the May 1989 CM1 index. How close was the predicted value? Was it within two standard deviations?

15. Repeat Exercise 14, but use the numbers for the TB3 index.

16. Reconsider the ARM problem of Example 8.5. Use a computer program for multiple regression to predict future trends in the CM1 index by fitting a second–degree polynomial to the data. After you input the time index $t = 1, 2, 3, \ldots$ in one column, and the CM1 data in a second column, prepare a third column of data containing the second power of the time index numbers $t^2 = 1, 4, 9, \ldots$ Multiple regression will give the best–fitting second–degree polynomial

$$\text{CM1} = a + bt + ct^2,$$

and R^2 can be interpreted as before. This technique is known as *polynomial least squares.*

(a) Use a computer implementation of multiple linear regression to obtain a formula that predicts the CM1 index as a quadratic function of t.

(b) Use the formula obtained in part (a) to predict the expected May 1990 value for the CM1 index.

(c) What percent of the total variation in the CM1 index can be accounted for by using this model?

(d) Compare the R^2 value for this multiple regression model with what was done in Section 8.3. Which model gives the best fit?

17. (Response–time formula from Example 3.2) A suburban community intends to replace its old fire station with a new facility. As part of the planning process, response–time data were collected for the past quarter. It took an average of 3.2 minutes to dispatch the fire crew. The dispatch time was found to vary only slightly. The time for the crew to reach the scene of the fire (drive time) was found to vary significantly depending on the distance to the scene. The data on drive time are displayed in Table 8.3.

(a) Use linear regression to obtain a formula that predicts drive time as a linear function of distance traveled. Then determine a formula for total response time, including dispatch time.

(b) What percentage of the total variation in drive time is accounted for by the formula you came up with in part (a)?

(c) Draw a graph of drive time versus distance for the data in Table 8.3. Do the data seem to indicate a linear trend?

(d) Plot the regression line from part (a) on the graph made in part (c). Does the line seem to be a good predictor for this data?

Distance (miles)	Drive Time (minutes)
1.22	2.62
3.48	8.35
5.10	6.44
3.39	3.51
4.13	6.52
1.75	2.46
2.95	5.02
1.30	1.73
0.76	1.14
2.52	4.56
1.66	2.90
1.84	3.19
3.19	4.26
4.11	7.00
3.09	5.49
4.96	7.64
1.64	3.09
3.23	3.88
3.07	5.49
4.26	6.82
4.40	5.53
2.42	4.30
2.96	3.55

Table 8.3: Response time data for the facility location problem.

18. (Continuation of Exercise 17) Another way to get a formula that relates drive time d to distance r is to use a power law model. Suppose that the underlying relationship between d and r is of the form $d = ar^b$. Taking logarithms of both sides yields the relationship

$$\ln d = \ln a + b \ln r.$$

Then linear regression can be used to estimate the parameters $\ln a$ and b in this linear equation.

(a) Transform the data in Table 8.3 by taking logarithms of both the drive time d and the distance r. Plot $\ln d$ versus $\ln r$. Does your graph suggest a linear relationship between $\ln d$ and $\ln r$?

(b) Use linear regression to obtain a formula that predicts $\ln d$ as a linear function of $\ln r$. Then determine a formula for total response time, including dispatch time, as a function of the distance r to the fire. Compare to the formula in Example 3.2.

(c) What is the value of R^2 for your regression model in part (b)? How do you interpret this number?

(d) Plot drive time d versus distance r, and then sketch a graph of the formula for d as a function of r, which you determined in part (b). Does the power law model seem to give a good fit to this data?

(e) Compare the results of parts (c) and (d) with the results of Exercise 17. Which model seems to give a better fit to this data? Justify your answer.

19. (Model selection for CM1 data) Reconsider the ARM problem of Example 8.6, but now consider a time series model with two previous months of data as predictors.

(a) Use a computer package for multiple linear regression to fit the model

$$X_t = a + bt + c_1 X_{t-1} + c_2 X_{t-2} + \varepsilon_t$$

to the CM1 data.

(b) Interpret the R^2 value for this model, and compare with the results of Example 8.6.

(c) Make a plot for the residuals for this model, similar to Figure 8.18 in the text. Do the residuals appear to form a stationary uncorrelated noise sequence?

(d) Iterate the model equation from part (a) and estimate the CM1 index on May 1990. Use the reported standard deviation to give a 95% prediction interval. Compare with the results of Example 8.6. Is this new model significantly better?

20. (Model selection for TB3 data) Reconsider the ARM problem of Examples 8.5 and 8.6, but now consider the TB3 data from Table 8.1.

(a) Use simple linear regression to fit the model $X_t = a + bt + \varepsilon_t$ as in Example 8.5. Predict the TB3 index for September 1989 and give a 95% prediction interval.

(b) Compute the residuals for the model in part (a) and check for stationarity and serial dependence. Does the model seem adequate?

(c) Repeat parts (a) and (b) using the autoregressive model $X_t = a + bt + c_1 X_{t-1} + \varepsilon_t$ as in Example 8.6 and compare with the results of (a) and (b).

(d) Which of the two models would you recommend? Justify your answer.

Further Reading

1. Arrow, K. et al. (1958) *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, California.
2. Billingsley, P. (1979) *Probability and Measure*. Wiley, New York.
3. Box, G. and Jenkins, G. (1976) *Time Series Analysis, Forecasting, and Control*. Holden-Day, San Francisco.
4. Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods*. 2nd Ed., Springer–Verlag, New York.
5. Çinlar, E. (1975) *Introduction to Stochastic Processes*. Prentice–Hall, Englewood Cliffs, New Jersey.
6. Cornell, R., Flora, J. and Roi, L. *The Statistical Evaluation of Burn Care*. UMAP module 553.
7. Freedman, D. et al. (1978) *Statistics*. W. W. Norton, New York.
8. Giordano, F., Wells, M. and Wilde, C. *Dimensional Analysis*. UMAP module 526.
9. Giordano, M., Jaye, M. and Weir, M. *The Use of Dimensional Analysis in Mathematical Modeling*. UMAP module 632.
10. Hillier, F. and Lieberman, G. (1990) *Introduction to Operations Research*. 5th Ed., Holden-Day, Oakland CA.
11. Hogg, R. and Tanis, E. (1988) *Probability and Statistical Inference*. Macmillan, New York.
12. Huff, D. (1954) *How to Lie with Statistics*. W. W. Norton, New York.
13. Kayne, H. *Testing a Hypothesis: t–Test for Independent Samples*. UMAP module 268.
14. Keller, M. *Markov Chains and Applications of Matrix Methods: Fixed–Point and Absorbing Markov Chains*. UMAP modules 107 and 111.
15. Knapp, T. *Regression Toward the Mean*. UMAP module 406.
16. Meerschaert, M. and Cherry, W. P. (1988) Modeling the behavior of a scanning radio communications sensor. *Naval Research Logistics Quarterly*, Vol. 35, 307–315.
17. Travers, K., and Heeler, P. *An Iterative Approach to Linear Regression*. UMAP module 429.
18. Yates, F. *Evaluating and Analyzing Probabilistic Forecasts*. UMAP module 572.