Chapter 12
Markov Decision Processes

Markov chains provide a useful modeling tool for determining expected profits or costs associated with certain types of systems. The key characteristic that allows for a Markov model is a probability law in which the future behavior of the system is independent of the past behavior given the present condition of the system. When this Markov property is present, the dynamics of the process can be described by a matrix containing the one-step transition probabilities and a vector of initial conditions. In some circumstances, the transition probabilities may depend on decisions made just before the transition time. Furthermore, not only the transition probabilities, but also associated costs or profits per transition may depend on decisions made at the transition times. For example, consider a slight variation of the first homework problem from Chap. 5. This problem had Joe and Pete playing a game of matching pennies (the pennies were biased), and the Markov model for the game used a state space representing the number of pennies in Joe’s pocket. We shall generalize the previous homework problem by having two rules for the game instead of one: Rule 1 states that Joe wins when the coins match and Pete wins when coins do not match; Rule 2 is the opposite, namely Pete wins when the coins match and Joe wins when they do not match. Now, before each play of the game, the previous winner gets to decide which rule is to be used. The dynamics of this game can no longer be modeled as a simple Markov chain because we need to know how Joe and Pete will make their decisions before transition probabilities can be determined.

Processes that involve decisions which affect the transition probabilities often yield models in which optimization questions naturally arise. When the basic structure of Markov chains is combined with decision processes and optimization questions, a new model called Markov decision processes is formed. In Markov decision processes, in contrast to Markov chains, the future depends not only on the current state but also on the decisions made. The purpose of this chapter is to present some of the basic concepts and techniques of Markov decision theory and indicate the types of problems that are amenable to modeling as such processes.

1 This chapter would normally be skipped in a one-semester undergraduate course.
12.1 Basic Definitions

The basic structure and dynamics of a Markov decision process will be introduced by way of an example. In this example, we introduce the basic elements of a Markov decision process; namely, a stochastic process with a state space denoted by $E$ and a decision process with an action space denoted by $A$.

**Example 12.1.** Let $X = \{X_0, X_1, \cdots \}$ be a stochastic process with a four-state state space $E = \{a, b, c, d\}$. This process will represent a machine that can be in one of four operating conditions denoted by the states $a$ through $d$ indicating increasing levels of deterioration. As the machine deteriorates, not only is it more expensive to operate, but also production is lost. Standard maintenance activities are always carried out in states $b$ through $d$ so that the machine may improve due to maintenance; however, improvement is not guaranteed. In addition to the state space, there is an action space which gives the decisions possible at each step. (We sometimes use the words “decisions” and “actions” interchangeably referring to the elements of the action space.) In this example, we shall assume the action space is $A = \{1, 2\}$; that is, at each step there are two possible actions: use an inexperienced operator (Action 1) or use an experienced operator (Action 2). To complete the description of a Markov decision problem, we need a cost vector and a transition matrix for each possible action in the action space. For our example, define the two cost vectors $f_1$ and $f_2$ and two Markov matrices as

$$f_1 = (100, 125, 150, 500)^T,$$

$$f_2 = (300, 325, 350, 600)^T,$$

$$P_1 = \begin{bmatrix} 0.1 & 0.3 & 0.6 & 0.0 \\ 0.0 & 0.2 & 0.5 & 0.3 \\ 0.0 & 0.1 & 0.2 & 0.7 \\ 0.8 & 0.1 & 0.0 & 0.1 \end{bmatrix},$$

$$P_2 = \begin{bmatrix} 0.6 & 0.3 & 0.1 & 0.0 \\ 0.75 & 0.1 & 0.1 & 0.05 \\ 0.8 & 0.2 & 0.0 & 0.0 \\ 0.9 & 0.1 & 0.0 & 0.0 \end{bmatrix}.$$  

The dynamics of the process are illustrated in Fig. 12.1 and are as follows: if, at time $n$, the process is in state $i$ and the decision $k$ is made, then a cost of $f_k(i)$ is incurred and the probability that the next state will be $j$ is given by $P_k(i,j)$. To illustrate, if $X_n = a$ and decision 1 is made, then a cost of $\$100$ is incurred (representing the operator cost, lost production cost, and machine operation cost) and $\Pr\{X_{n+1} = a\} = 0.1$; or, if $X_n = d$ and decision 2 is made, then a cost of $\$600$ is incurred (representing the operator cost, machine operation cost, major maintenance cost, and lost-production cost) and $\Pr\{X_{n+1} = a\} = 0.9$.  

---

2 A superscript T denotes transpose.