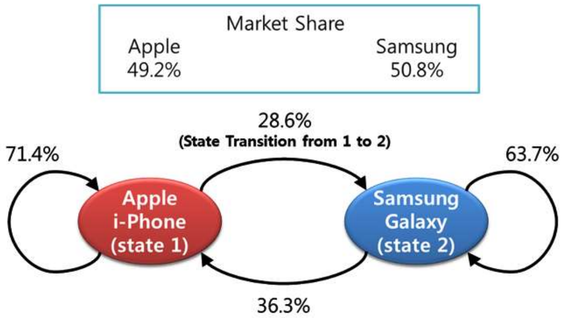
Chapter 3. Markov Chains



Introduction to Markov Chains

Consider a system that can be in any one of a finite or countably infinite number of states. Let $\mathscr{S}$ denote this set of states. We can assume that $\mathscr{S}$ is a subset of the integers. The set $\mathscr{S}$ is called the   of the system. Let the system be observed at the discrete moments of time $n = 0, 1, 2, \ldots$, and let X_n denote the state of the system at time n.

Since we are interested in non-deterministic systems, we think of X_n, $n \geq 0$, as random variables defined on a common probability space. Little can be said about such random variables unless some additional structure is imposed upon them.

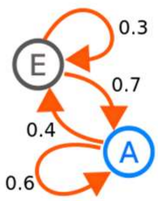
The simplest possible structure is that of independent random variables. This would be a good model for such systems as repeated experiments in which future states of the system are independent of past and present states. In most systems that arise in practice, however, past and present states of the system influence the future states even if they do not uniquely determine them.

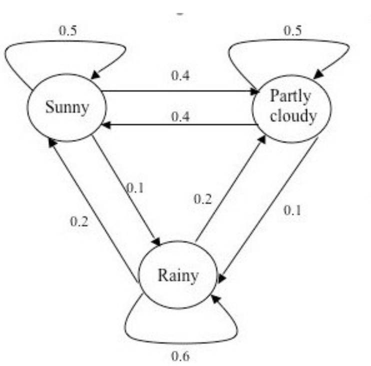
Many systems have the property that given the present state, the past states have no influence on the future. This property is called the [   ] and systems having this property are called [   ] The Markov property is defined precisely by the requirement that

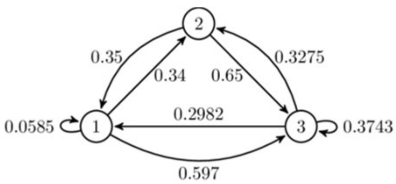
(1) [   ]

for every choice of the nonnegative integer n and the numbers $x_0, \ldots, x_{n+1}$, each in $\mathscr{S}$. The conditional probabilities [   ] are called the *transition probabilities* of the chain. In this book we will study Markov chains having [   ] transition probabilities, i.e., those such that $P(X_{n+1} = y \mid X_n = x)$ is independent of n. From now on, when we say that X_n, $n \geq 0$, forms a Markov chain, we mean that these random variables satisfy the Markov property and have stationary transition probabilities.

The study of such Markov chains is worthwhile from two viewpoints. First, they have a rich theory, much of which can be presented at an elementary level. Secondly, there are a large nurnber of systems arising in practice that can be modeled by Markov chains, so the subject has many useful applications.







Markov chains having two states

For an example of a Markov chain having two states, consider a machine that at the start of any particular day is either broken down or in operating condition. Assume that if the machine is broken down at the start of the nth day, the probability is p that it will be successfully repaired and in operating condition at the start of the $(n+1)$th day. Assume also that if the machine is in operating condition at the start of the nth day, the probability is q that it will have a failure causing it to be broken down at the start of the $(n+1)$th day. Finally, let $\pi_0(0)$ denote the probability that the machine is broken down initially, i.e., at the start of the 0th day.

Let the state 0 correspond to the machine being broken down and let the state 1 correspond to the machine being in operating condition. Let $\boxed{X_n}$ be the random variable denoting the state of the machine at time n.

According to the above description

$$P(X_{n+1} = 1 \mid X_n = 0) = p,$$

$$P(X_{n+1} = 0 \mid X_n = 1) = q,$$

and

$$P(X_0 = 0) = \pi_0(0).$$

Since there are only two states, 0 and 1, it follows immediately that

and that the probability $\pi_0(1)$ of being initially in state 1 is given by

$$\pi_0(1) = P(X_0 = 1) = 1 - \pi_0(0).$$

From this information, we can easily compute $P(X_n = 0)$ and $P(X_n = 1)$. We observe that

$$P(X_{n+1} = 0) = P(X_n = 0 \text{ and } X_{n+1} = 0) + P(X_n = 1 \text{ and } X_{n+1} = 0)$$

$$= (1 - p)P(X_n = 0) + qP(X_n = 1)$$

$$= (1 - p)P(X_n = 0) + q(1 - P(X_n = 0))$$

$$= (1 - p - q)P(X_n = 0) + q.$$

Now $P(X_0 = 0) = \pi_0(0)$, so

$$P(X_1 = 0) = (1 - p - q)\pi_0(0) + q$$

and

It is easily seen by repeating this procedure n times that

$$P(X_n = 0) = (1 - p - q)^n \pi_0(0) + q \sum_{j=0}^{n-1} (1 - p - q)^j. \tag{2}$$

In the trivial case $p = q = 0$, it is clear that for all n

$$P(X_n = 0) = \pi_0(0) \quad \text{and} \quad P(X_n = 1) = \pi_0(1).$$

Suppose now that $p + q > 0$. Then by the formula for the sum of a finite geometric progression,

$$\sum_{j=0}^{n-1} (1 - p - q)^j = \frac{1 - (1 - p - q)^n}{p + q}.$$

We conclude from (2) that

(3)

and consequently that

(4)

Suppose that p and q are neither both equal to zero nor both equal to 1. Then $0 < p + q < 2$, which implies that $|1 - p - q| < 1$. In this case we can let $n \to \infty$ in (3) and (4) and conclude that, for any *initial dist.* π_0,

(5)

We can also obtain the probabilities $q/(p + q)$ and $p/(p + q)$ by a different approach. Suppose we want to choose $\pi_0(0)$ and $\pi_0(1)$ so that $P(X_n = 0)$ and $P(X_n = 1)$ are independent of n. It is clear from (3) and (4) that to do this we should choose

$$\pi_0(0) = \frac{q}{p + q} \quad \text{and} \quad \pi_0(1) = \frac{p}{p + q}.$$

Thus we see that if X_n, $n \geq 0$, starts out with the initial distribution

$$P(X_0 = 0) = \frac{q}{p+q} \quad \text{and} \quad P(X_0 = 1) = \frac{p}{p+q},$$

then for all n

$$P(X_n = 0) = \frac{q}{p+q} \quad \text{and} \quad P(X_n = 1) = \frac{p}{p+q}.$$

$$\begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

The description of the machine is vague because it does not really say whether X_n, $n \geq 0$, can be assumed to satisfy the Markov property. Let us suppose, however, that the Markov property does hold. We can use this added information to compute the joint distribution of $X_0, X_1, \ldots, X_n$.

For example, let $n = 2$ and let x_0, x_1, and x_2 each equal 0 or 1. Then

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2) \\ &\quad = P(X_0 = x_0 \text{ and } X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) \\ &\quad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1). \end{aligned}$$

$$\begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Now $P(X_0 = x_0)$ and $P(X_1 = x_1 \mid X_0 = x_0)$ are determined by p, q, and $\pi_0(0)$; but without the Markov property, we cannot evaluate $P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1)$ in terms of p, q, and $\pi_0(0)$. If the Markov property is satisfied, however, then

$$P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) = P(X_2 = x_2 \mid X_1 = x_1),$$

which is determined by p and q. In this case

For example,

$$\begin{aligned} P(X_0 = 0, X_1 = 1, \text{ and } X_2 = 0) & \\ &= P(X_0 = 0)P(X_1 = 1 \mid X_0 = 0)P(X_2 = 0 \mid X_1 = 1) \\ &= \pi_0(0)pq. \end{aligned}$$

The reader should check the remaining entries in the following table, which gives the joint distribution of X_0, X_1, and X_2.

x_0	x_1	x_2	$P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2)$
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Transition function and initial distribution

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$. (The restriction to two states is now dropped.) The function $P(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, defined by

(6) $$P(x, y) = P(X_1 = y \mid X_0 = x), \qquad x, y \in \mathscr{S},$$

is called the *transition function* of the chain. It is such that

(7) $$P(x, y) \geq 0, \qquad x, y \in \mathscr{S},$$

and

(8) $$\sum_y P(x, y) = 1, \qquad x \in \mathscr{S}.$$

Since the Markov chain has stationary probabilities, we see that

(9) $$P(X_{n+1} = y \mid X_n = x) = P(x, y), \qquad n \geq 1.$$

It now follows from the Markov property that

(10) $$P(X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x) = P(x, y).$$

In other words, if the Markov chain is in state x at time n, then no matter how it got to x, it has probability $P(x, y)$ of being in state y at the next step. For this reason the numbers $P(x, y)$ are called the *one-step transition probabilities* of the Markov chain.

The function $\pi_0(x)$, $x \in \mathscr{S}$, defined by

$$\pi_0(x) = P(X_0 = x), \qquad x \in \mathscr{S}, \tag{11}$$

is called the *initial distribution* of the chain. It is such that

$$\pi_0(x) \geq 0, \qquad x \in \mathscr{S}, \tag{12}$$

and

$$\sum_x \pi_0(x) = 1. \tag{13}$$

The joint distribution of $X_0, \ldots, X_n$ can easily be expressed in terms of the transition function and the initial distribution. For example,

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1) &= P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0) \\ &= \pi_0(x_0)P(x_0, x_1). \end{aligned}$$

Also,

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, X_2 = x_2) \\ &\qquad = P(X_0 = x_0, X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) \\ &\qquad = \pi_0(x_0)P(x_0, x_1)P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1). \end{aligned}$$

$$\begin{aligned} P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) &= P(X_2 = x_2 \mid X_1 = x_1) \\ &= P(X_1 = x_2 \mid X_0 = x_1) \\ &= P(x_1, x_2). \end{aligned}$$

Thus

By induction it is easily seen that

$$(14) \qquad P(X_0 = x_0, \ldots, X_n = x_n) = \pi_0(x_0)P(x_0, x_1)\cdots P(x_{n-1}, x_n).$$

It is usually more convenient, however, to reverse the order of our definitions. We say that $P(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, is a *transition function* if it satisfies (7) and (8), and we say that $\pi_0(x)$, $x \in \mathscr{S}$, is an *initial distribution* if it satisfies (12) and (13). It can be shown that given any transition function P and any initial distribution π_0, there is a probability space and random variables X_n, $n \geq 0$, defined on that space satisfying (14). It is not difficult to show that these random variables form a Markov chain having transition function P and initial distribution π_0.

It will soon be apparent that the transition function of a Markov chain plays a much greater role in describing its properties than does the initial distribution. For this reason it is customary to study simultaneously all Markov chains having a given transition function. In fact we adhere to the usual convention that by "a Markov chain having transition function P," we really mean the family of all Markov chains having that transition function.