

## Markov Chains

Consider a process that can go through several possible states through time. We may be interested in describing the law by which the process evolves from one state to another. There are two main approaches to this question: (1) the *deterministic* approach in which the evolution from each state to another is known with certainty, and (2) the *stochastic* approach in which that evolution obeys a probabilistic law.

The choice between a deterministic and a stochastic approach is usually dictated by the nature of the process under study. The movement of the planets around the Sun, for instance, seems to call for a deterministic approach since there is little uncertainty about which way the planet is moving. But the movement of elementary particles, because of the Heisenberg uncertainty principle is better described in stochastic terms.

The set of states involved in the process can range from the finite, to the discrete, to the continuum. Likewise, time can be counted in discrete or continuous fashion. Finally the law of evolution can involve more or less dependence on past events. A stochastic process that has finitely or countably many states and involves no memory of the past is called a Markov Chain. There are, however, two cases: (1) the discrete time Markov chains, and (2) the continuous time Markov chains. We first recall some basic definitions.

To define a probability space one needs a set  $\Omega$  (the sample space) together with the algebra  $\mathcal{S}$  of all its subsets and a map  $P: \mathcal{S} \rightarrow [0,1]$  satisfying

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \quad \text{whenever } A \cap B = \emptyset \\ P(\Omega) &= 1 \end{aligned}$$

Now, let  $I$  be a countable (perhaps finite) set with generic element denoted  $i \in I$ . A random variable  $X$  with values in  $I$  is simply a function  $X: \Omega \rightarrow I$ . We can naturally define the probability

$$\lambda_i = P(X = i) = P(\{\omega \in \Omega: X(\omega) = i\})$$

The set  $\lambda = (\lambda_i: i \in I)$  is the *distribution* of the random variable  $X$ . It clearly satisfies  $\lambda_i \geq 0$  and  $\sum_{i \in I} \lambda_i = 1$ . We will now say that a matrix  $T = (p_{ij})_{i,j \in I}$  is stochastic

if every row of  $T$  is a distribution. For instance  $T = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$  is stochastic.

## Math 500: Markov Processes, Decisions, and Evolution: Class #9

With those definitions, we can now define a *discrete* time Markov chain as a sequence  $X_n$  ( $n = 0, 1, \dots$ ) of random variables such that

- (1)  $X_0 = \lambda$ , with  $\lambda$  a given "initial" distribution, and
- (2)  $X_{n+1} = X_n T$ , for all  $n \geq 0$ , with  $T$  a stochastic "transition" matrix.

Of course, if  $I$  is finite, so is the matrix  $T$ , and it has as many rows and columns as any  $X_n$ , and that is the cardinal of  $I$ . We will mostly work with such finite sets.

With the above transition matrix  $T$ , there is a nice visualization of how such a Markov chain works. If the current time is  $n$  and if the current state is 1 with certainty, then  $X_n = (1,0,0)$ . Therefore,  $X_{n+1} = X_n T = (\frac{1}{2}, 0, \frac{1}{2})$  and there is 50-50 chance that the state at time  $(n + 1)$  will be 1 or 3. Similarly, state 3 will lead to state 1 with probability  $\frac{1}{2}$ , to state 2 with probability  $\frac{1}{3}$ , and will stay at state 3 with probability  $\frac{1}{6}$ . The way to visualize this probabilistic law is with a diagram (Figure 1 below) showing all three possible states of the Markov chain together with the possible moves from one state to another with corresponding probabilities (one usually does not show the moves from a state to itself).

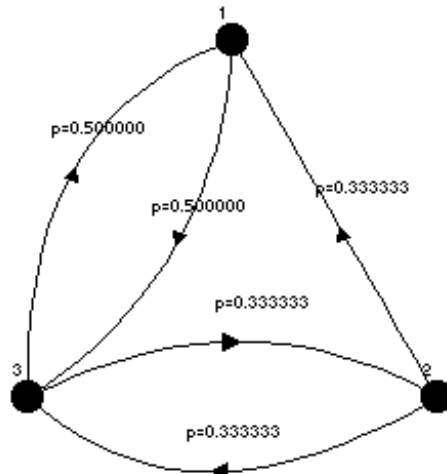


Figure 1

In the study of Markov chains, one asks questions such as "what is the probability of being at such state and at such time?" To answer this kind of questions, we begin by some characterizations. First, when can we say that a discrete time random process is a Markov chain with initial distribution  $\lambda$  and transition matrix  $T$ ? The answer is

*Theorem:*  $(X_n)_{n \geq 0}$  is a Markov chain  $(\lambda, T)$  if and only if for any  $N$  and  $i_0, \dots, i_N$ :

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

## Math 500: Markov Processes, Decisions, and Evolution: Class #9

*Proof:* One applies the basic probability law that  $P(A \cap B) = P(A)P(B|A)$ . See Norris, pp. 2-3 for details.<sup>1</sup>

Next, instead of starting the analysis of the effect of a Markov chain at time 0, one may want to start it at time  $m$ . Suppose that at that time, conditional on  $X_m = i$ , we want to know how the  $(X_{m+n})_{n \geq 0}$  behave. The answer is

*Theorem:*  $(X_{m+n})_{n \geq 0}$  is a Markov chain  $(\delta, T)$  with  $\delta$  the distribution such that  $\delta_i = 1$ , and  $\delta_j = 0$  if  $j \neq i$ . It is independent of the random variables  $(X_n)_{0 \leq n < m}$ .

*Proof:* One shows that

$$P(X_m = i, X_{m+1} = i_{m+1}, \dots, X_{m+n} = i_{m+n}) = \delta_i p_{i i_{m+1}} \dots p_{i_{m+n-1} i_{m+n}}.$$

See Norris, pp. 3-4 for details. We will later generalize this property to so-called "stopping times."

In other words, if we have reached the specified distribution  $\delta$  at time  $m$ , then the subsequent developments are a Markov chain with initial distribution  $\delta$  and same transition matrix  $T$ . All prior events  $(X_n)_{0 \leq n < m}$  become irrelevant. The Markov chain has no memory of the past.

Finally, in order to turn to examples, we need to develop a calculation device. Clearly, if we start with  $X_0 = \lambda$  and if  $X_{n+1} = X_n T$ , for all  $n \geq 0$ , we can write inductively

$$X_1 = X_0 T, X_2 = X_1 T = X_0 T^2, \dots, X_n = X_0 T^n = \lambda T^n$$

So, the successive powers of the transition matrix  $T$  give the distribution of  $X_n$  at any time  $n$  from the initial distribution  $\lambda$ . Let us use this idea with our current example. Evidently, we need to find some systematic calculation for  $T^n$ . But we recall from earlier studies that most matrices can be diagonalized. Why not try here?

Our usual technique is to write the characteristic polynomial and find eigenvalues and eigenvectors. Here we get ( $\mu$  stands for an eigenvalue, *not* an initial distribution)

$$\begin{aligned} p(\mu) &= \begin{vmatrix} \frac{1}{2} - \mu & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} - \mu & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} - \mu \end{vmatrix} = \\ &= \frac{1}{36} \left\{ (1 - 2\mu) \left( (1 - 3\mu)(1 - 6\mu) - 2 \right) + 2 - 3(1 - 3\mu) \right\} \\ &= \left( \mu^2 - \frac{1}{18} \right) (\mu - 1) = 0 \end{aligned}$$

---

<sup>1</sup>Norris refers to *Markov Chains* by J.R. Norris, 1997, Cambridge University Press

## Math 500: Markov Processes, Decisions, and Evolution: Class #9

So, we find  $\mu = 1$ , and  $\pm \frac{\sqrt{2}}{6}$  as the eigenvalues of  $T$ . With some appropriate nonsingular matrix  $B$ , and the diagonal matrix  $M$  (with the eigenvalues on the diagonal) we may write

$$T = BMB^{-1} \quad \text{and} \quad T^n = (BMB^{-1})^n = BM^nB^{-1}$$

But the matrix  $M^n$  has the eigenvalues of  $T$  raised to the exponent  $n$  on its diagonal. So

$$X_n = \lambda T^n = \lambda BM^nB^{-1}$$

is a (row) vector that is a *fixed* linear combination of the eigenvalues raised to the  $n$ . If we ask, for instance, "what is the probability  $x_n$  of being at state 1 after  $n$  turns, starting from state 1?" we will write, for some constants  $a$ ,  $b$ , and  $c$ , to be determined

$$x_n = a + b\left(\frac{\sqrt{2}}{6}\right)^n + c\left(-\frac{\sqrt{2}}{6}\right)^n$$

Now, clearly  $x_0 = 1 = a + b + c$ . At time  $n = 1$ ,  $x_1 = \frac{1}{2} = a + \frac{\sqrt{2}}{6}(b - c)$ . And at time  $n = 2$ ,  $x_2 = \frac{1}{2} = a + \frac{1}{18}(b + c)$ . Solving yields

$$a = \frac{8}{17}, \quad b = \frac{18+3\sqrt{2}}{68}, \quad c = \frac{18-3\sqrt{2}}{68}$$

So, we have an explicit formula for  $x_n$  and could get similar formulas for any other such probability of being at state 2 or 3, starting anywhere. In particular, we see that the long run probability of being at state 1 is  $a = \frac{8}{17}$ .

### Homework

1. With the same example, find the long run probabilities to be at 2 or 3.
2. Do #1.1.4 p. 9 of Norris.
3. Do #1.1.7 p. 10 of Norris.