Appendix II

Review of Markov chains

A ***stochastic*** or ***random process*** is defined as a family of random variables, $\{X(t)\}$, describing an empirical process, the development of which in time is governed by probabilistic laws. The ***state space***, S, of the stochastic process is the set of all possible values that the random variable $X(t)$ can take. The parameter t is often interpreted as time and may be either discrete or continuous. When the set of possible values of t forms a countable set, the process $\{X(t), t = 0, 1, 2, \ldots\}$, is ***discrete.*** If t forms an interval of real line, the process $\{X(t),\ t \geq 0\}$ is said to be ***continuous.*** In the discrete case, the state space can be finite or infinite.

Among many different discrete stochastic processes, we are interested in a special class called Markov chains. The basic concepts of Markov chains were introduced in 1907 by the Russian mathematician A.A. Markov.

Let $i_1, i_2, \ldots$ represent the states of the chain. The sequence of random variables $X_1, X_2, \ldots$ is called a ***Markov chain*** if:

$$P(X_n = i_{k_n} | X_1 = i_{k_1}, \ldots, X_{n-1} = i_{k_{n-1}}) = P(X_n = i_{k_n} | X_n = i_{k_{n-1}})$$

An intuitive interpretation is that a stochastic process $\{X(t)\}$ has the Markov property if the conditional probability of any future state, given the present and past states, is independent of the past states and depends only on the present state. Thus, a Markov chain can be used to model the position of an object in a discrete set of possible states over time, in which the subsequent position is chosen at random from a distribution that depends only on the current location of the chain and not on any previous locations of the chain.

The conditional probabilities that the chain moves to state j at time n, given that it is in state i at time $n-1$, are called ***transition probabilities*** and are denoted by p_{ij},

$$p_{ij} = P(X_n = j | X_{n-1} = i),$$

with the subscript ij of p indicating the direction of transition $i \to j$. Sometimes, p_{ij} may also be represented by $p(i,j)$, and if we need to represent the time points, then we use the notation, $p_{n-1,n}(i,j) = P(X_n = j | X_{n-1} = i)$.

Two basic assumptions we make are that (1) $p_{ij} \geq 0$ for all i and j; the transition probabilities are nonnegative. Also, (2) for every i,

$$\sum_{j=1}^{\infty} p_{ij} = 1 \left(\sum_{j=1}^{n} p_{ij} = 1 \text{ if the state space is finite} \right),$$

that is, the chain makes a transition to some state in the state space.

If the transition probabilities p_{ij} depend only on the sates i and j and not on the time n, then the conditional probabilities are called ***stationary***. Markov chains with stationary probabilities are called (time) ***homogeneous Markov chains***. We shall consider only homogeneous Markov chains.

The behavior of homogeneous Markov chains is described by the transition or stochastic matrices of the processes in which the transition probabilities are arranged as elements of a matrix. The ***transition*** or ***stochastic matrix*** of a chain having transition probabilities $i, j = 1, 2, \ldots n$ is:

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

In the infinite state space case, we represent the transition matrix in the following manner:

$$\begin{pmatrix} p_{11} & \cdots & p_{ln} \cdots \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mn} \cdots \\ \vdots & & \vdots \end{pmatrix}$$

Each element of the matrix is nonnegative, and each row sums to 1. If we look at any particular row, say the mth row, then we can see the probabilities of going from state m to the various other states including the state m.

EXAMPLE AII.1

Four quarterbacks are warming up by throwing a football to one another. Let 1, 2, 3, and 4 denote the four quarterbacks. It has been observed that 1 is as likely to throw the ball to 2 as to 3 and 4. Player 2 never throws to 3 but splits his throws between 1 and 4. Quarterback 3 throws twice as many passes to 1 as to 4 and never to 2, but 4 throws only to 1. This process forms a Markov chain because the player who is about to throw the ball is not influenced by the player who had the ball before him. The one-step transition matrix is

$$\begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

The following is a standard example of a chain with infinite state space.

EXAMPLE AII.2

Consider a chain with state space $S = (0, 1, 2, 3, \ldots)$ and transition matrix:

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & 0 & \cdots \\ 0 & 0 & q_3 & r_3 & p_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix},$$

where $p_i, q_i, r_i \geq 0$ for all $i \geq 0$, $p_0 + r_0 = 1$, and $p_i + q_i + r_i = 1$ for all $i \geq 1$. Thus, for this Markov chain, the transition probabilities are $p_{00} = r_0$, $p_{01} = p_0$, and for $i, j \neq 0$,

$$p_{ij} = \begin{cases} p_i, & j = i+1 \\ r_i, & j = i \\ q_i, & j = i-1 \\ 0, & \text{otherwise.} \end{cases}$$

This chain is known as the ***random walk chain*** (***with barrier at 0***).

The following example gives a transition matrix for the random walk chain in a special case. We can think of this as a chain resulting from tossing a fair coin. If we are not at state 0, then if heads comes up, we take a step to the right, and if tails comes up, we take a step to the left. If at state 0, we remain at 0 for a tails outcome and move a step to the right for heads.

EXAMPLE AII.3

Consider a Markov chain with state space $S = (0, 1, 2, 3, \ldots)$ and the transition probabilities given by:

$$p_{00} = \frac{1}{2}, p_{ij} = \begin{cases} \frac{1}{2} & j = i - 1 \\ \frac{1}{2} & j = i + 1 \\ 0, & \text{otherwise.} \end{cases}$$

This results in the symmetric transition matrix with elements:

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \ldots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \ldots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \ldots \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

The ***n*-step transition probability**, $p_{ij}^{(n)}$, is defined as the probability that the chain is in state i and will go to state j in n steps. If p_{ij} is the one-step transition probability, $p_{ij}^{(n)}$ can be obtained as follows. Let i be the state of the process at time m, that is $X_m = i$. Then, the n-step transition probability is:

$$\begin{aligned} p_{ij}^{(n+m)} &= P(X_{n+m} = j | X_0 = i) \\ &= \sum_{k=0}^{\infty} P(X_{n+m} = j, X_n = k | X_0 = i) \\ &= \sum_{k=0}^{\infty} P(X_{n+m} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i) \\ &= \sum_{k=0}^{\infty} p_{kj}^{m} p_{ik}^{n}. \end{aligned}$$

This can be rewritten in the matrix notation as:

$$P^{(n+m)} = P^{(m)} P^{(n)} = P^{(n)} P^{(m)}.$$

This is known as the *Chapman–Kolmogorov equation*.

The following example shows how to compute an n-step transition matrix.

EXAMPLE AII.4

Consider the one-step transition matrix given in Example AII.1,

$$\begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

The two-step transition matrix, P^2, is:

$$P^2 = P.P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{13}{18} & 0 & 0 & \frac{5}{18} \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

The three-step transition matrix, P^3, is:

$$P^3 = P^2P = \begin{pmatrix} \frac{5}{18} & \frac{13}{54} & \frac{13}{54} & \frac{13}{54} \\ \frac{13}{36} & \frac{1}{6} & \frac{1}{6} & \frac{11}{36} \\ \frac{13}{27} & \frac{1}{9} & \frac{1}{9} & \frac{8}{27} \\ \frac{13}{18} & 0 & 0 & \frac{5}{18} \end{pmatrix}.$$

For instance, the third row of P^3,

$$\begin{pmatrix} \frac{13}{27} & \frac{1}{9} & \frac{1}{9} & \frac{8}{27} \end{pmatrix},$$

denotes that, after three throws, the ball is in the hands of player 1, 2, 3, or 4 with respective probabilities 13/27, 1/9, 1/9, and 8/27.

A transition matrix, P, all entries of which are positive, is called a ***positive transition matrix***. A state j of a Markov chain is ***accessible*** from a state i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$. If state j is accessible from state i, and state i is accessible from state j, the states are said to ***communicate***. If all the states communicate, then the Markov chain is called ***irreducible***. A state i is ***periodic*** (of period d) if the only way to revisit it is through steps of length $k.d$ for some value of k and a fixed value of $d > 1$. Thus, the period, d, is the greatest common divisor of the number of steps n needed for the chain, starting at state i, to revisit the state i:

$$d = GCD\{n \geq 1 | p_{ii}^n > 0\}.$$

If a state is not periodic, then it is called ***aperiodic***. A state i is ***recurrent*** if it will be revisited by the chain with probability 1. That is,

$$P(X_n = i \text{ for infinitely many } n | X_0 = i) = 1.$$

If a state is not recurrent, it is called ***transient***. Recurrent, aperiodic states are called ***ergodic***. It is necessary to impose an extra condition for ergodicity, that the expected recurrence time be finite. This is satisfied for recurrent states in a finite-state Markov chain. A Markov chain is called ***ergodic*** if every state is ergodic. It is clear that a finite-state Markov chain with a positive transition matrix is ergodic.

The following result is of fundamental importance.

Theorem AII.1. For an ergodic Markov chain, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ exists, and this limit is independent of the initial state i. Let the vector $\boldsymbol{\pi}$ with elements (π_j) be the limiting or the stationary distribution of the chain. Then, this stationary probability vector is the unique solution of the equation:

$$\boldsymbol{\pi} = \boldsymbol{\pi} P,$$

and satisfies the normalization condition:

$$\sum_{j\in S} \pi_j = 1.$$

If, at any transition step n, the distribution of the chain is the same as π obtained in Theorem AII.1, we say that the chain has reached the ***steady state***. Thus, the vector $\boldsymbol{\pi}$ would be the unique steady-state probability vector of the Markov chain.

Analogous to the law of large numbers for a sequence of independent random variables, for Markov chains we can obtain the following so-called ***ergodic theorem***:

Theorem AII.2. For any ergodic Markov chain $\{X_n\}$ with stationary distribution $\boldsymbol{\pi}$:

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \sum_{i\in S} f(i)\pi_i = Ef(X) \;\; w.p.1.$$

The validity of the Markov chain Monte Carlo method lies in this ergodic theorem.