MARKOV CHAIN

1. Markov Chain

Outlines

Markov Chain

Let $\{X_n\}$, $n \geq 0$, be a sequence of random variables which take values in some finite set $S = \{s_1, s_2, \cdots, s_k\}$, called the **state space**. Each X_n is a discrete random variable that takes one of k possible states and its probability distribution can be represented by the (probability) vector $\pi_n = (\pi_n(1), \pi_n(2), \cdots, \pi_n(k)) \in \mathbb{R}^k$, in which $\pi_n(j) = P(X_n = s_j)$, for $j = 1, 2, \cdots, k$.
It is clear that $\pi_n(j) \geq 0$, for all j, and $\sum_{j=1}^{k} \pi_n(j) = 1$.
Definition 1. The sequence $\{X_n\}$ is called a **Markov chain** if it satisfies the **Markov condition**,

$$P(X_n = s | X_0 = x_0, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}) = P(X_n = s | X_{n-1} = x_n \quad (1)$$

for all $n \geq 1$ and all $s, x_0, x_1, \cdots, x_{n-1} \in S$.

Remarks. (a) The Markov property (1) is equivalent to each of the stipulations (2) and (3) below: for each $s \in S$ and for every sequence $\{x_i : i \geq 0\}$ in S, for all $n_1 < n_2 < \cdots < n_k \leq n$,

$$P\left(X_{n+1} = s \middle| X_{n_1} = x_{n_1}, X_{n_2} = x_{n_2}, \cdots, X_{n_k} = x_{n_k}\right) = P\left(X_{n+1} = s \middle| X_{n_k} = \right. \tag{2}$$

$$P\left(X_{m+n} = s \middle| X_0 = x_0, X_1 = x_1, \cdots, X_m = x_m\right) = P\left(X_{m+n} = s \middle| X_m = x\right. \tag{3}$$

(b) Without loss of generality, we can denote the elements of S as 1, 2, . . . , k, although in some examples we may use the original labeling of the states to avoid confusion.
(c) The distribution vector of X_n, $\pi_n = (\pi_n(1), \pi_n(2), \cdots, \pi_n(k))$, is called the distribution at $t = n$ and π_0 the **initial distribution** of the Markov chain.

Definition 2. The Markov chain $\{X_n\}$ is called **homogeneous** if

$$P\left(X_{n+1}=i \mid X_n=j\right)=P\left(X_1=i \mid X_0=j\right)$$

for all n, i, j. The **transition matrix** $\mathbf{P}=(p_{ij})$ is the $k \times k$ matrix of **transition probabilities**

$$p_{ij}=P\left(X_{n+1}=i \mid X_n=j\right)$$

$$\begin{array}{ccc} & \text{State j at time } t=n & \\ & \downarrow & \\ \mathbf{P}= & \begin{bmatrix} & & \\ & p_{ij} & \\ & & \end{bmatrix} & \leftarrow \text{State i at time } t=n+1 \end{array}$$

From now on, all Markov chains are assumed homogeneous, with state space $S=\{1,2,\cdots,k\}$, transition matrix $\mathbf{P}\in\mathbb{R}^{k\times k}$, and probability distributions $\pi_n\in\mathbb{R}^k$, $n=0,1,2,\cdots$, where $\pi_n(i)=P(X_n=i)$, for $i=1,2,\cdots,k$.

Proposition 3. Let $\{X_n\}$ be a Markov chain. For all $m, n \geq 0$ and $i, j \in S$, we have

$$P\left(X_{m+n} = i \middle| X_m = j\right) = P\left(X_n = i \middle| X_0 = j\right).$$

The matrix $\mathbf{P}^{(n)} = \left(p_{ij}^{(n)}\right) \in \mathbb{R}^{k \times k}$, where $p_{ij}^{(n)} = P\left(X_n = i \middle| X_0 = j\right)$, is called the **n-step transition matrix** and the $p_{ij}^{(n)}$ are called the **n-step transition probabilities**. In particular, $\mathbf{P}$ is the 1-step transition matrix of the Markov chain.

$$\begin{array}{cc} & \text{State j at time } t = m \\ & \downarrow \\ \mathbf{P}^{(n)} = & \left[\quad p_{ij}^{(n)} \quad\right] \qquad \leftarrow \text{State i at time } t = m+n \end{array}$$

Theorem 4 (Chapman-Kolmogorov equations). Let $\{X_n\}$ be a Markov chain. For all $n \geq 0$ and $i, j \in S$, we have

$$P\left(X_{n+1} = i \middle| X_0 = j\right) = \sum_{\mathrm{l}=1}^{k} P\left(X_{n+1} = i \middle| X_n = \mathrm{l}\right) P\left(X_n = \mathrm{l} \middle| X_0 = j\right). \tag{4}$$

Proof

Since $P\left(X_{n+1} = i, X_0 = j\right) = \sum_{\mathrm{l}=1}^{k} P\left(X_{n+1} = i, X_n = \mathrm{l}, X_0 = j\right)$, and for each $\mathrm{l} = 1, 2, \cdots, k$,

$$\begin{aligned} P\left(X_{n+1} = i, X_n = \mathrm{l}, X_0 = j\right) &= P\left(X_{n+1} = i \middle| X_n = \mathrm{l}, X_0 = j\right) P\left(X_n = \mathrm{l}\right. \\ = P\left(X_{n+1} = i \middle| X_n = \mathrm{l}\right) P\left(X_n = \mathrm{l} \middle| X_0 = j\right) P\left(X_0 = j\right), \end{aligned}$$

we have

$$P\left(X_{n+1}=i \mid X_0=j\right) = \frac{P(X_{n+1}=i, X_0=j)}{P(X_0=j)} = \frac{1}{P(X_0=j)} \sum_{\mathrm{l}=1}^{k} P\left(X_{n+1}=i, X_n\right.$$
$$= \sum_{\mathrm{l}=1}^{k} P\left(X_{n+1}=i \mid X_n=\mathrm{l}\right) P\left(X_n=\mathrm{l} \mid X_0=j\right)$$

as desired.

Since $P\left(X_{n+1}=i \mid X_n=\mathrm{l}\right)=P\left(X_1=i \mid X_0=\mathrm{l}\right)=p_{i\mathrm{l}}$, the equation (4) can be rewritten as

$$p_{ij}^{(n+1)} = \sum_{\mathrm{l}=1}^{k} p_{i\mathrm{l}} p_{\mathrm{l}j}^{(n)},$$

where the right hand side is equal to the inner product of the i-th row of $\mathbf{P}$ with the j-th column of $\mathbf{P}^{(n)}$, and we arrive at

Corollary 5. For each $n \in \mathbb{N}$,

$$\mathbf{P}^{(n+1)} = \mathbf{P} \times \mathbf{P}^{(n)}.$$

In particular, $\mathbf{P}^{(n)} = \mathbf{P}^{n}$, i.e., the n-step transition matrix is equal to the n-th power of $\mathbf{P}$, the 1-step transition matrix. **Corollary 6**. Let $\{X_n\}$ be a Markov chain with initial distribution π_0. For all $n \in \mathbb{N}$, we have

$$\pi_n = P^n \pi_0.$$

Proof. For each $i = 1, 2, \cdots, k$,
$\pi_n(i) = P(X_n = i) = \sum_{j=1}^{k} P(X_n = i; X_0 = j)$, and for each $j = 1, 2, \cdots, k$,

$$P(X_n = i; X_0 = j) = P(X_n = i | X_0 = j) P(X_0 = j) = \mathbf{P}^{(n)}(i, j) \pi_0(j) =$$

i.e.,

$$\pi_n(i) = \sum_{j=1}^{k} \mathbf{P}^{n}(i, j) \pi_0(j).$$

Since the right hand side is the inner product of the i-th row of $\mathbf{P}^n$ with the (column) vector π_0, we conclude that $\pi_n = P^n \pi_0$ as desired.

Definition 7. Let $\{X_n\}$ be a Markov chain and let $\pi = (\pi(1), \pi(2), \cdots, \pi(k))$ be a distribution on S, i.e., $\pi(i) \geq 0$, for all $i = 1, 2, \cdots, k$, and $\sum_{i=1}^{k} \pi(i) = 1$.

(i) π is called the **limiting distribution** of $\{X_n\}$ if $\lim_{n \to \infty} \mathbf{P}^n(i, j) = \pi(i)$, for all $i, j = 1, 2, \cdots, k$.

(ii) π is called a **stationary distribution** of $\{X_n\}$ if

$$\pi = \mathbf{P}\pi.$$

Remarks. (a) By the uniqueness property of the limit, a Markov chain has at most one limiting distribution, whereas it may have more than one stationary distribution.

(b) If π is the limiting distribution, then $P(X_n = i | X_0 = j) \to \pi(i)$ as $n \to \infty$, for all $j = 1, 2, \cdots, k$, i.e., regardless of what the initial state j is, the probability that the chain is at the state i at time $t = n$, $P(X_n = i)$, can be approximated by $\pi(i)$ for large values of n.

(c) If the chain is at a stationary distribution π at time $t = n$, $\pi_n = \pi$, then at all subsequent times, the distribution of the state of the chain remains exactly the same as π, i.e., $\pi_{n+m} = \pi$, for all $m \in \mathbb{N}$.

Definition 8. The transition matrix $\mathbf{P}$ of a Markov chain is said to be **regular** if there exists $n \in \mathbb{N}$ such that all elements of $\mathbf{P}^n$ are positive.

Theorem 9. Let $\{X_n\}$ be a Markov chain with regular transition matrix $\mathbf{P}$. Then the limiting distribution of the chain exists and is equal to the unique stationary distribution.

Example. Let $\{X_n\}$ be a Markov chain with state space $S = \{1, 2, 3\}$, transition matrix

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.2 & 0 & 1 \\ 0.8 & 0.4 & 0 \\ 0 & 0.6 & 0 \end{bmatrix} \end{array}$$

and initial distribution $\pi_0 = (0.5, 0.5, 0)$.

The 3-step transition matrix is given by

$$\mathbf{P}^3 = \begin{bmatrix} 0.488 & 0.36 & 0.04 \\ 0.224 & 0.544 & 0.48 \\ 0.288 & 0.096 & 0.48 \end{bmatrix}$$

which gives the distribution of X_3,

$$\pi_3 = \mathbf{P}^3 \pi_0 = \begin{bmatrix} 0.424 \\ 0.384 \\ 0.192 \end{bmatrix}.$$

Therefore, we have

$$P\left(X_{10} = 2 \middle| X_7 = 1, X_6 = 3\right) = P\left(X_{10} = 2 \middle| X_7 = 1\right) = \mathbf{P}^3(2,1) = 0.224$$

$$P\left(X_4 = 2, X_3 = 1\right) = P\left(X_4 = 2 \middle| X_3 = 1\right) P\left(X_3 = 1\right) = \mathbf{P}(2,1) \cdot \pi_3(1) =$$

All elements of $\mathbf{P}^3$ are positive implies that $\mathbf{P}$ is regular and then the limiting distribution of the exists and is equal to its unique stationary distribution $\pi = (a, b, c)$, with $a, b, c \geq 0$, $a + b + c = 1$, and

$$\pi = \mathbf{P}\pi \Leftrightarrow (I_3 - \mathbf{P})\,\pi = \mathbf{0}. \tag{5}$$

By transforming

$$I_3 - \mathbf{P} = \begin{bmatrix} 0.8 & 0 & -1 \\ -0.8 & 0.6 & 0 \\ 0 & -0.6 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -\frac{5}{4} \\ 0 & 0.6 & -1 \\ 0 & -0.6 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -\frac{5}{4} \\ 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 \end{bmatrix}$$

and letting $c = \alpha \in \mathbb{R}$, we get the general solution $\pi = \left(\frac{5}{4}\alpha, \frac{5}{3}\alpha, \alpha\right)$, $\alpha \in \mathbb{R}$, of (5). The condition $a, b, c \geq 0$ implies that $\alpha \geq 0$, and the condition $1 = a + b + c = \frac{47}{12}\alpha$ gives $\alpha = \frac{12}{47}$, i.e.,

$$\pi = \left(\frac{15}{47}, \frac{20}{47}, \frac{12}{47}\right) \approx (0.3192, 0.4255, 0.2553)\,.$$

Thus in the long run, regardless of the initial distribution, the chain will be at the state 1, 2, 3, with probability 0.3192, 0.4255, 0.2553, respectively.