Markov chains – first steps

A not very exciting game

Consider a game with a playing board consisting of squares numbered 1-10 arranged in a circle. A player starts at square 1. At each turn, the player rolls a die and moves around the board by the number of spaces shown on the face of the die.
Let X_k be the position of the player after k moves (with $X_0 = 1$).
For example, assume that the player successively rolls 2,1, and 4.
The first four positions are:

- $X_0 = 1$
- $X_1 = 3$
- $X_2 = 4$
- $X_3 = 8$.

Given this information, what can be said about the player's next location?

A not very exciting game - cont'd

Even though we know the player's full past history of moves, the only information relevant for predicting a future position is the most recent location X_3. Since $X_3 = 8$, then necessarily

$$X_4 \in \{9, 10, 1, 2, 3, 4\},$$

with equal probabilities, so

$$\mathbb{P}\left(X_4 = j \middle| X_0 = 1, X_1 = 3, X_2 = 4, X_3 = 8\right) = \mathbb{P}\left(X_4 = j \middle| X_3 = 8\right) = \frac{1}{6}$$

for $j = 9, 10, 1, 2, 3, 4$. Given the player's most recent location X_3, her/his future position X_4 is independent of past history X_0, X_1, X_2.

Markov chains

Definition 1

Let S be a discrete set. A **(discrete-time) Markov chain** with the state space S is a sequence of random variables $X_0, X_1, \ldots$, denoted as $(X_n)_{n\in\mathbb{N}_0}$, taking values in S such that for all $n = 0, 1, 2, \ldots$, and for all $x_0, x_1, \ldots, x_n, x_{n+1} \in S$

$$\mathbb{P}\left(X_{n+1} = x_{n+1} \middle| X_0 = x_0, \ldots, X_n = x_n\right) \\ = \mathbb{P}\left(X_{n+1} = x_{n+1} \middle| X_n = x_n\right).$$

This is the so called **Markov property** - for a Markov chain

the future, given the present, is independent of the past.

Time-homogeneity

Definition 2

A Markov chain $(X_n)_n$ is *time-homogeneous* if

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

for all $i, j \in S$ and $n = 0, 1, 2, \ldots$.

Example 3

As an example of a **time-homogeneous Markov chain**, we can consider the following random process with two states $S = \{1, 2\}$. We assume that for $n = 0, 1, 2, \ldots$

- $\mathbb{P}\left(X_{n+1} = 1 \middle| X_n = 1\right) = 0.2$,
- $\mathbb{P}\left(X_{n+1} = 2 \middle| X_n = 1\right) = 0.8$,
- $\mathbb{P}\left(X_{n+1} = 1 \middle| X_n = 2\right) = 0.5$,
- $\mathbb{P}\left(X_{n+1} = 2 \middle| X_n = 2\right) = 0.5$.

Example 4

As an example of a **non-homogeneous Markov chain**, consider a random process on states $S = \{1, 2\}$, generated according the following rules: if $n = 0, 2, 4, 6, \ldots$, then

- $\mathbb{P}\left(X_{n+1} = 1 \middle| X_n = 1\right) = 0.2$,
- $\mathbb{P}\left(X_{n+1} = 2 \middle| X_n = 1\right) = 0.8$,
- $\mathbb{P}\left(X_{n+1} = 1 \middle| X_n = 2\right) = 0.5$,
- $\mathbb{P}\left(X_{n+1} = 2 \middle| X_n = 2\right) = 0.5$,

whereas if $n = 1, 3, 5, 7, \ldots$, then

- $\mathbb{P}\left(X_{n+1} = 1 \middle| X_n = 1\right) = 0.1$,
- $\mathbb{P}\left(X_{n+1} = 2 \middle| X_n = 1\right) = 0.9$,
- $\mathbb{P}\left(X_{n+1} = 1 \middle| X_n = 2\right) = 0.9$,
- $\mathbb{P}\left(X_{n+1} = 2 \middle| X_n = 2\right) = 0.1$.

Unless stated otherwise,

the Markov chains in this course will be time-homogeneous.

It makes sense to define:

$$p_{ij} := \mathbb{P}\left(X_{n+1} = j \middle| X_n = i\right) = \mathbb{P}\left(X_1 = j \middle| X_0 = i\right)$$

- **one-step transition probability** of moving from state i to state j.

The one-step transition probabilities can be arranged into a matrix $\mathbf{P}$, whose (i, j)-th entry is p_{ij}:

$$\mathbf{P} = (p_{ij})_{i,j\in S}.$$

This is the **transition matrix**, which contains the one-step transition probabilities of moving from state to state.

If the state space S has n elements, then $\mathbf{P}$ is a square $n \times n$ matrix. If S is countably infinite, $\mathbf{P}$ is an infinite matrix.

Definition 5

A **stochastic matrix** is a square matrix $\mathbf{P}$, which satisfies

- $\mathbf{P}_{ij} \geq 0$ for all i, j,
- For each row i,

$$\sum_j \mathbf{P}_{ij} = 1.$$

Fact 6

If $\mathbf{P}$ *is a one-step transition matrix of a Markov chain, then* $\mathbf{P}$ *is a stochastic matrix.*

Markov chains – basic computations

A powerful feature of Markov chains is the ability to use matrix algebra for computing probabilities. To use matrix methods, we consider probability distributions as vectors.

If X is a discrete random variable with $\mathbb{P}(X = j) = \alpha_j$, then $\alpha := (\alpha_1, \alpha_2, \ldots)$ is a **probability vector**. We say then that **the distribution of X is α**. For matrix computations we will identify discrete probability distributions with *row vectors*.

For a Markov chain $(X_n)_{n\in\mathbb{N}_0}$, the distribution of X_0 is called **the initial distribution** of the Markov chain $(X_n)_n$. Thus, if α is the initial distribution of $(X_n)_n$, then

$$\mathbb{P}(X_0 = j) = \alpha_j$$

for all $j \in S$.

n-step transition probabilities

For states i and j, and for $n \geq 1$,

$$\mathbb{P}\left(X_n = j \middle| X_0 = i\right)$$

is the probability that the chain started in i hits j in n steps.

The n-step transition probabilities can be arranged into a matrix, whose (i, j)-th entry is $\mathbb{P}\left(X_n = j \middle| X_0 = i\right)$ (this is the ***n*-step transition matrix** of the Markov chain). Clearly, for $n = 1$, this is just the usual (one-step) transition matrix $\mathbf{P}$.

Theorem 7

The n-step transition matrix is $\mathbf{P}^n$ *(the n-th matrix power of* $\mathbf{P}$*).*

Example 8

Let $(X_n)_n$ be a Markov chain with the state space $S = \{1, 2\}$ and transition matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}.$$

Compute $\mathbb{P}\left(X_2 = j \middle| X_0 = i\right)$ for $i, j \in \{1, 2\}$.

Chapman-Kolmogorov

Theorem 9

The Chapman-Kolmogorov equations:

$$\mathbb{P}\left(X_{n+m} = j \middle| X_0 = i\right) = \sum_{k \in S} \mathbb{P}\left(X_m = k \middle| X_0 = i\right) \mathbb{P}\left(X_{m+n} = j \middle| X_m = k\right)$$

Interpretation: transitioning from i to j in $m + n$ steps is equivalent to transitioning from i to some state k in m steps and then from that state to j in the remaining n steps.

Law of Total Probability

Theorem 10

Let $B_1, \ldots, B_k$ be a sequence of events that partition the sample space. That is, the B_i are mutually exclusive (disjoint) and their union is equal to Ω. Then, for any event A,

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}\left(A \middle| B_i\right) \mathbb{P}(B_i).$$

Example 11

There are nine fair coins and one two-headed coin. We pick one coin and random and make five tosses. What is the probability of obtaining three heads in this way?

Distribution of X_n

Example 12

Let $(X_n)_n$ be a Markov chain with the state space $S = \{1, 2\}$, initial distribution α and transition matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}.$$

- Find the distribution of X_1 if $\alpha = (1/2, 1/2)$.
- Find the distribution of X_1 if $\alpha = (1, 0)$.
- Find the distribution of X_1 if $\alpha = (0, 1)$.
- Find the distribution of X_2 if $\alpha = (0, 1)$.

Distribution of X_n - cont'd

Theorem 13

Let $(X_n)_n$ be a Markov chain with transition matrix $\mathbf{P}$ and initial distribution α. For all $n \geq 0$ the distribution of X_n is $\alpha\mathbf{P}^n$. That is

$$\mathbb{P}(X_n = j) = (\alpha\mathbf{P}^n)_j \quad \forall j.$$

Example 14

Let

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}.$$

- If $\alpha = (1/2, 1/2)$ then $\alpha\mathbf{P} = (1/4, 3/4)$.
- If $\alpha = (1, 0)$ then $\alpha\mathbf{P} = (1/2, 1/2)$.
- If $\alpha = (0, 1)$ then $\alpha\mathbf{P} = (0, 1)$.
- If $\alpha = (0, 1)$ then $\alpha\mathbf{P}^2 = (0, 1)$.

Present, future, and most recent past

The Markov property says that past and future are independent given the present. It is also true that past and future are independent, given *the most recent* past.

Example 15

If $(X_n)_n$ is a Markov chain then

$$\mathbb{P}\left(X_9 = x_9 \middle| X_6 = x_6, X_4 = x_4, X_3 = x_3, X_1 = x_1\right) \\ = \mathbb{P}\left(X_9 = x_9 \middle| X_6 = x_6\right).$$

Joint distribution

The marginal distributions of a Markov chain are determined by the initial distribution α and the one-step transition matrix $\mathbf{P}$.

However, a much stronger result is true. In fact, α and $\mathbf{P}$ determine all the joint distributions of a Markov chain, that is, the joint distribution of any finite subset of $X_0, X_1, X_2, \ldots$. In that sense, the initial distribution and transition matrix give a complete probabilistic description of a Markov chain.

Example 16

Express

$$\mathbb{P}(X_5 = i, X_3 = j)$$

in terms of α and $\mathbf{P}$.