Introduction to Markov Chains

Edmund Ruan

DRP Presentation, April 28, 2021
Mentor: Greyson

Outline

1. What is a Markov chain?
2. Why do we care?
3. How can we study them?

Outline

Basic Probability

Definition (Probability Measure)

A probability measure $\mathbb{P}$ is an $\mathbb{R}$-valued function on a set of "possible events."

For example if Z_k is a random variable recording the sum after rolling a die for k turns, then:

$$\mathbb{P}(Z_1 = 6) = \frac{1}{6}$$

Definition (Conditional Probability)

The probability of A given B is:

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

For example:

$$\mathbb{P}(Z_3 = 18 \mid Z_1 = 6) = \frac{1}{36}$$

Markov Chains

Definition (Markov Chain)

An $\mathbb{S}$-valued discrete-time stochastic process $(Z_n)_{n \in \mathbb{N}}$ is said to be Markov if, for all $n \geq 1$, the probability distribution Z_{n+1} is determined by the state Z_n of the process at time n and does not depend on the past values of Z_k for $k < n$. Here, $\mathbb{S}$ is a discrete state space, e.g. $\mathbb{S} = \mathbb{Z}$, $\mathbb{S} = \{0, 1\}$, etc.

In other words, for all $n \geq 1$ and all $i_0, i_1, \ldots, i_n, j \in \mathbb{S}$ we have:

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i_n, Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0) = \mathbb{P}(Z_{n+1} = j \mid Z_n = i_n)$$

Outline

Examples of Markov Chains

- **Weather** - the state of the weather right now can be predicted by yesterday's weather. We do not care too much about the weather, say, 3 years ago.
- **PageRank algorithm** for rating the relevance of web-pages for a Google search.
- **Text generation** - state space consists of different word sequences and next word(s) generated based on most recent word sequences.
- Modeling in **finance**, **insurance claims**, **genetics**, etc.

Outline

First-step analysis (e.g. Ruin probabilities in gambling)

Suppose you are gambling at a slot machine which costs \$1 to play and pays out \$2 with probability $0 \leq w \leq 1$. Let Z_i be how much money you have after playing for i rounds.

The **ruin probability** $\mathbb{P}(R \mid Z_0 = k)$ is defined as the probability that you lose all your money at some time, given that you start with \$k.

The key insight is that ruin probability depends on the data of the starting point and not on the starting time, i.e.:

$$\mathbb{P}(R \mid Z_1 = k \pm 1 \text{ and } Z_0 = k) = \mathbb{P}(R \mid Z_1 = k \pm 1) = \mathbb{P}(R \mid Z_0 = k \pm 1)$$

First-step analysis (e.g. Ruin probabilities in gambling)

From this, you can show that $\mathbb{P}(R \mid Z_0 = k)$ satisfies a recursive equation:

$$\mathbb{P}(R \mid Z_0 = k) = w \cdot \mathbb{P}(R \mid Z_0 = k+1) + (1-w) \cdot \mathbb{P}(R \mid Z_0 = k-1)$$
$$\mathbb{P}(R \mid Z_0 = 0) = 1$$
$$\mathbb{P}(R \mid Z_0 = M) = 0$$

where M is the amount of money you would need to not play at all or stop playing. Solving this recursive system, you can show that:

$$\mathbb{P}(R \mid Z_0 = k) = \frac{\left(\frac{w}{1-w}\right)^{M-k} - 1}{\left(\frac{w}{1-w}\right)^{M} - 1}$$

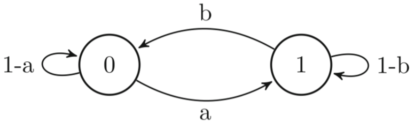
Transition Matrix (e.g. 2-state Markov Chain)

Since Markov chains only depend on the previous state, we can encode how they evolve with a **transition matrix** P, which is made up of transition probabilities:

$$P_{i,j} := \mathbb{P}(Z_{n+1} = j \mid Z_n = i), i, j \in \mathbb{S}$$

Since these probabilities are independent of n, they are referred to as "time homogeneous."

For example, consider the following two-state model, letting Z_i be either 0 or 1 at time i:



$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

Transition Matrix (e.g. 2-state Markov Chain)

Then, given a distribution at time 0, denoted by π_0:

$$\pi_0 = \begin{bmatrix} \mathbb{P}(Z_0 = 0) & \mathbb{P}(Z_0 = 1) \end{bmatrix}$$

You can compute the distribution at time 1 via:

$$\pi_1 = \pi_0 P = \begin{bmatrix} \mathbb{P}(Z_1 = 0) & \mathbb{P}(Z_1 = 1) \end{bmatrix}$$

In fact, the distribution at time k is given by:

$$\pi_k = \pi_0 P^k$$

P^k is called the **Higher-Order transition matrix**.

Transition Matrix (e.g. 2-state Markov Chain)

For example, let $a = 0.2$ and $b = 0.4$. Then, we can find an explicit formula for P^k by diagonalization:

$$P^k = \frac{1}{0.6}\begin{bmatrix} 0.4 + 0.2 \cdot (0.4)^k & 0.2 \cdot (1 - (0.4)^k) \\ 0.4 \cdot (1 - (0.4)^k) & 0.2 + 0.4 \cdot (0.4)^k \end{bmatrix}, k \in \mathbb{N}$$

Furthermore, given any initial distribution π_0, you can show that:

$$\lim_{k \to \infty} \pi_0 P^k = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

This is called a **limiting distribution** and tells us for example that:

$$\lim_{k \to \infty} \mathbb{P}(Z_k = 0 \mid Z_0 = 0) = \lim_{k \to \infty} \mathbb{P}(Z_k = 0 \mid Z_0 = 1) = \frac{2}{3}$$

References

Book: *Understanding Markov Chains: Examples and Applications*, Second Edition, Nicolas Privault

Questions?