IB Markov Chains

Ishan Nath, Michaelmas 2022

Based on Lectures by Dr. Perla Sousi

October 14, 2022

Contents

0 Introduction

Markov chains are random processes (sequence of random variables) that retain no memory of the past.

0.1 History

These were first studied by Markov in 1906. Before Markov, Poisson processes and branching processes were studied. The motivation was to extend the law of large numbers to a non-iid setting.

After Markov, Kolmogorov began studying continuous time Markov chains, also known as Markov processes. An important example is Brownian motion, which is a fundamental object in modern probability theory.

Markov chains are the simplest mathematical models for random phenomena evolving in time. They are **simple** in the sense they are amenable to tools from probability, analysis and combinatorics.

Applications of Markov chains include population growth, mathematical genetics, queueing networks and Monte Carlo simulation.

0.2 PageRank Algorithm

This is an algorithm used by Google Search to rank web pages. We model the web as a directed graph, $G = (V, E)$. Here, V is the set of vertices, which are associated to the website, and $(i, j) \in E$ if i contains a link to page j.

Let $L(i)$ be the number of outgoing edges from i, i.e. the outdegree, and let $|V| = n$. Then we define a set of probabilities

$$\hat{p}_{ij} = \begin{cases} \frac{1}{L(i)} & \text{if } L(i) > 0, (i,j) \in E, \\ \frac{1}{n} & \text{otherwise.} \end{cases}$$

Take $\alpha \in (0, 1)$, then we define $p_{ij} = \alpha \hat{p}_{ij} + (1 - \alpha)\frac{1}{n}$. Consider a random surfer, who tosses a coin with bias α, and either goes to $\hat{p}$, or chooses a website uniform at random.

We wish to find an invariant distribution $\pi = \pi P$. Then π_i is the proportion of time spent at webpage i by the surfer. We can then rank the pages by the values of π_i.

1 Formal Setup

We begin with a state space I, which is either finite or countable, and a σ-algebra $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1. A stochastic process $(X_n)_{n\geq 0}$ is called a **Markov chain** if for all $n \geq 0$, and $x_0, x_1, \ldots, x_{n+1} \in I$,

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

If $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ is independent of n for all x, y, then X is called **time-homogeneous**. Otherwise, it is **time-inhomogeneous**.

For a time-homogeneous Markov chain, define $P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x)$. P is called the **transition matrix** of the Markov chain. We have

$$\sum_{y \in I} P(x, y) = \sum_{y \in I} \mathbb{P}(X_1 = y \mid X_0 = x) = 1.$$

Such a matrix is called a **stochastic matrix**.

Definition 1.2. $(X_n)_{n\geq 0}$ with values in I is called Markov(λ, P) if $X_0 \sim \lambda$ and $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix P.

There are several equivalent definitions for Markov chains.

Theorem 1.1. *X is* Markov(λ, P) *if for all $n \geq 0$, $x_0, x_1, \ldots, x_n \in I$,*

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \lambda(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Proof: If X is Markov(λ, P), then

$$\begin{aligned}\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0) &= \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \times \cdots \\ &\quad \times \mathbb{P}(X_1 = x_1 \mid X_0 = x_0)\mathbb{P}(X_0 = x_0) \\ &= \lambda(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n).\end{aligned}$$

For the other direction, note for $n = 0$, we have $\mathbb{P}(X_0 = x_0) = \lambda(x_0)$, so $X_0 \sim \lambda$, and

$$\begin{aligned}\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) &= \frac{\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0)}{\mathbb{P}(X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)} \\ &= P(x_{n-1}, x_n).\end{aligned}$$

Definition 1.3. Let $i \in I$. The δ_i-mass at i is defined by

$$\delta_{ij} = \mathbb{1}(i = j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.4. Let $X_1, \ldots, X_n$ be discrete random variables with values in I. They are independent if for all $x_1, \ldots, x_n \in I$,

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

Let $(X_n)_{n \geq 0}$ be a sequence of random variables in I.They are independent if for all $i_1 < i_2 < \cdots < i_k$, and for all $x_1, \ldots, x_k \in I$,

$$\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = i_k) = \prod_{j=1}^{k} \mathbb{P}(X_{i_j} = x_j).$$

Let $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ be two sequences. We say $X \perp Y$, or X independent to Y, if for all $k, m \in \mathbb{N}$, $i_1 < \cdots < i_k$, $j_1 < \cdots < j_m$, $x_1, \ldots, x_k, y_1, \ldots, y_m$,

$$\begin{aligned} &\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k, Y_{j_1} = y_1, \ldots, Y_{j_m} = y_m) \\ &\quad = \mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k)\mathbb{P}(Y_{j_1} = y_1, \ldots, Y_{j_m} = y_m). \end{aligned}$$

Theorem 1.2 (Simple Markov Property)**.** *Suppose X is* Markov(λ, P) *with values in I. Let $m \in \mathbb{N}$ and $i \in I$. Then conditional on $X_m = i$, the process $(X_{m+n})_{n \geq 0}$ is* Markov(δ_i, P) *and it is independent of $X_0, \ldots, X_m$.*

Proof: Let $x_0, \ldots, x_n \in I$. Then

$$\begin{aligned} &\mathbb{P}(X_m = x_0, \ldots, X_{m+n} = x_n \mid X_m = i) \\ &\qquad = \delta_{ix_0} \frac{\mathbb{P}(X_m = x_0, \ldots, X_{m+n} = x_n)}{\mathbb{P}(X_m = i)}, \\ &\mathbb{P}(X_m = x_0, \ldots, X_{m+n} = x_n) \\ &\qquad = \sum_{y_0, \ldots, y_{m-1}} \mathbb{P}(X_0 = y_0, \ldots, X_m = x_0, \ldots, X_{m+n} = x_0) \\ &\qquad = \sum_{y_0, \ldots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, x_0) \cdots P(x_{n-1}, x_n) \\ &\qquad = P(x_0, x_1) \cdots P(x_{n-1}, x_n) \sum_{y_0, \ldots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, x_0), \\ &\mathbb{P}(X_m = i) = \sum_{y_0, \ldots, y_{m-1}} \lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, i). \end{aligned}$$

Putting this together, we get the probability is

$$\delta_{ix_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n) \implies \text{Markov}(\delta_i, P).$$

Now we show independence. Let $m \le i_1 < \cdots < i_k$. Then,

$$\begin{aligned}
&\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k, X_0 = y_0, \ldots, X_m = y_m \mid X_m = i) \\
&\qquad = \frac{\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k, X_0 = y_0, \ldots, X_m = y_m)}{\mathbb{P}(X_m = i)} \\
&\qquad = \frac{\lambda(y_0) P(y_0, y_1) \cdots P(y_{m-1}, y_m)}{\mathbb{P}(X_m = i)} \mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k \mid X_m = i) \\
&\qquad = \mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k \mid X_m = i) \mathbb{P}(X_0 = y_0, \ldots \mid X_m = i).
\end{aligned}$$

Let $X \sim \text{Markov}(\lambda, P)$. How can we find $\mathbb{P}(X_n = x)$? Evaluating,

$$\begin{aligned}
\mathbb{P}(X_n = x) &= \sum_{x_0, \ldots, x_{n-1}} \mathbb{P}(X_0 = x_0, \ldots, X_n = x) \\
&= \sum_{x_0, \ldots, x_{n-1}} \lambda(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x) = (\lambda P^n) x.
\end{aligned}$$

Here, λ is a row vector, and P^n is the n'th power of the transition matrix. By convention, $P^0 = I$.

Consider the related problem of finding $\mathbb{P}(X_{n+m} = y \mid X_m = x)$. From the simple Markov property, $(X_{m+n})_{n \ge 0}$ is $\text{Markov}(\delta_x, P)$. So

$$\mathbb{P}(X_{n+m} = y \mid X_m = x) = (\delta_x P^n) y = (P^n) xy.$$

Example 1.1. Take the transition matrix

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

Then $p_{11}(n+1) = (1 - \alpha) p_{11}(n) + \beta p_{12}(n)$. Since $p_{11}(n) + p_{12}(n) = 1$, we get the general form

$$p_{11}(n) = \begin{cases} \dfrac{\alpha}{\alpha + \beta} + \dfrac{\alpha}{\alpha + \beta}(1 - \alpha - \beta)^n & \alpha + \beta > 0, \\ 1 & \alpha + \beta = 0. \end{cases}$$

Suppose P is $k \times k$ stochastic, and let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of P.

If $\lambda_1, \ldots, \lambda_k$ are all distinct, then P is diagonalisable, so we can write

$$P = U \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix} U^{-1}.$$

Then we get

$$P^n = U \begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^n \end{pmatrix}.$$

Hence $p_{11}(n) = \alpha_1\lambda_1^n + \cdots + \alpha_k\lambda_k^n$.

If one of the eigenvalues is complex, say λ_{k-1}, then also its conjugate is an eigenvalue. Say $\lambda_k = \overline{\lambda_{k-1}}$. If $\lambda_{k-1} = re^{i\theta} = r\cos\theta + ir\sin\theta$, $\lambda_k = r\cos\theta - ir\sin\theta$, then we can write the general form as

$$p_{11}(n) = \alpha_1\lambda_1^n + \cdots + \alpha_{k-2}\lambda_{k-2}^n + \alpha_{k-1}r^n\cos n\theta + \alpha_k r^n \sin n\theta.$$

If an eigenvalue λ has multiplicity r, then we must include the term $(a_{r-1}n^{r-1} + \cdots + a_1 n + a_0)\lambda^n$, by Jordan Normal Form.

Index