# Markov Chains

Adam Kelly (`ak2316@cam.ac.uk`)

October 14, 2021

This article constitutes my notes for the 'Markov Chains' course, held in Michaelmas 2021 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered.

## Contents

## §1 Introduction

For this whole course, $I$ will be a finite or countable set. All of our random variables will also be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

> **Definition 1.1** (Markov Chain)
>
> A **stochastic process** $(X_n)_{n \geq 0}$ is called a **Markov chain** if for all $n \geq 0$ and all $x_0, \ldots, x_{n+1} \in I$, we have[a]
>
> $$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$
>
> ---
> [a]We assume here that we are not conditioning on a zero probability event.

**Remark.** This definition gives a *discrete time* Markov chain. It is possible to define a continuous time Markov chain, but we won't worry about that for now.

If $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ for all $x, y \in I$ is independent of $n$, then $X$ is called **time-homogeneous**. Otherwise, it is **time-inhomogeneous**. In this course, we will only study time-homogeneous Markov chains.

We will write $P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x)$, where $x, y \in I$. We call $P$ a **stochastic matrix**, because

$$\sum_{y \in I} P(x, y) = 1,$$

that is, the sum of each row is 1.

**Remark.** The index set does not have to be $\mathbb{N}$, it could be say $\{0, 1, \ldots, N\}$ for $N \in \mathbb{N}$.

So to characterize a Markov chain, we need this matrix $P$, giving the probability of passing from a state $x$ to a state $y$. We call this matrix the **transition matrix** of $X$.

**Definition 1.2** (Markov)

We say that $X$ is Markov$(\lambda, P)$ if $X_0$ has distribution $\lambda$ and $P$ is the transition matrix. That is,

(i) $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$, $x_0 \in I$,

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

We usually represent a Markov chain by its diagram corresponding to the allowed transitions.

**Example 1.3** (Diagram of a Markov Chain)

Let $\alpha, \beta \in (0, 1)$. We consider the matrix

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

This is a transition matrix on two states which we can call 1 and 2. Here $\alpha$ is the probability of staying at 1, and $1 - \alpha$ is the probability of moving from state 2 when at state 1.
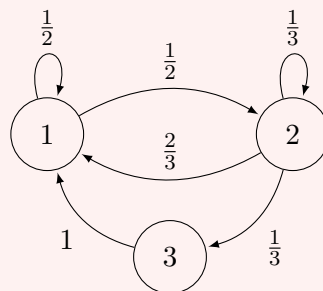
A diagram of this is given below. This is a directed graph with the relevant probabilities labelling each edge.
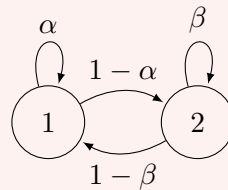


**Example 1.4**

Suppose that we have the transition matrix

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

This is a transition matrix on three states and corresponds with the diagram below.

**Theorem 1.5**

The process $X$ is Markov$(\lambda, P)$ if and only if for all $n \geq 0$ and all $x_0, \ldots, x_n \in I$ we have
$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \lambda_{x_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

*Proof.* First suppose that $X$ is Markov$(\lambda, P)$. Then

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

as required.

Now suppose that the property holds. Then $n = 0$ gives $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$, so our base case holds. Then

$$\begin{aligned}
\mathbb{P}(X_n = x_\rangle n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) &= \frac{\lambda_{x_0} P(x_0, x_1) \cdots P(x_{n-1}, x_n)}{\lambda_{x_0} P(x_0, x_1) \cdots P(x_{n-2}, x_{n-1})} \\
&= P(x_{n-1}, x_n)
\end{aligned}$$

$\square$

Now we are going to define some useful notation.

**Definition 1.6** ($\delta_i$-mass)

For $i \in I$, the $\delta_i$-**mass** of $i$ is defined as $\delta_{ij} = \mathbb{1}(i = j)$.

Recall the notion of independence for random variables. Let $X_1, \ldots, X_n$ be discrete random variables. They are *independent* if for all $x_1, \ldots, x_n \in I$, we have

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

We have a similar notion for sequences of random variables. We say a sequence $(X_n)_{n \geq 0}$ is *independent* if for all $i_1 < i_2 < \cdots < i_k$ and all $x_1, \ldots, x_k$,

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

If $X = (X_n)_{n \geq 0}$ and $Y = (Y_n)_{n \geq 0}$ are two sequences of random variables, they are independent if for all $k, m$ and $i_1 < \cdots < i_k$, $j_1 < \cdots j_m$ we have

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

**Theorem 1.7** (Simple Markov Property)

Suppose that $X \sim \text{Markov}(\lambda, P)$. Fix $m \in \mathbb{N}$ and $i \in I$. Conditional on $X_m = i$, the process $(X_{m+n})_{n \geq 0}$ is $\text{Markov}(\delta_i, P)$ and it is independent of $X_0, \ldots, X_m$.

*Proof.* We have

$$\mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m \mid X_m = i\right) = \frac{\mathbb{P}(X_{m+n} = x_{m+n}, \ldots, X_m = x_m)\delta_{ix_m}}{\mathbb{P}(X_m = i)}.$$

We can rewrite the numerator as

$$
\begin{aligned}
&\mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m\right) \\
&= \sum_{x_0, \ldots, x_{m-1} \in I} \mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m, X_{m-1} = x_{m-1}, \ldots, X_0 = x_0\right) \\
&= \sum_{x_0, \cdots, x_{m-1}} \lambda_{x_0} P\left(x_0, x_1\right) \cdots P(x_{m-1}, x_m) P(x_m, x_{m+1}) \cdots P(x_{m+n-1}, x_{m+n}) \\
&= P\left(x_m, x_{m+1}\right) \cdots P\left(x_{m+n-1}, x_{m+n}\right) \mathbb{P}\left(X_m = x_m\right).
\end{aligned}
$$

Substituting this back into our original expression, we get

$$\mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m \mid X_m = i\right) = \delta_{ix_m} P(x_m, x_{m+1}) \cdots P(x_{m+n-1}, x_{m+n}),$$

showing that $(X_{m+n})_{n \geq 0}$ is $\text{Markov}(\delta_i, P)$ conditional on $X_m = i$. $\qquad\square$

**Remark.** Informally, this theorem says 'past and future are independent given the present'.