1 Discrete Time Markov Chains – Basic Definitions

We will define a class of stochastic processes with possible outcomes that only depend on its current state.

Definition 1.1. A stochastic process $X = \{X_n\}_{n\geq 0}$ taking values in a state space S is called a **discrete-time Markov Chain (DTMC)** if

1. The state space S is **countable**.
2. The **Markov property** holds: for all $n \in \{0, 1, 2, \cdots\} = \mathbb{N}$ and $x_0, \dots, x_{n+1} \in S$, we have

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

Intuitively, the Markov property states that the probability that the Markov chain moves to state x_{n+1} at time $n+1$ depends only on the state x_n in which it is in at time n and not on any states in which it was before time n. The state space is often taken to be subsets of the integers $\mathbb{Z}$.

Example 1.2. Let $Y_1, Y_2, \dots$ be an i.i.d. (independent and identically distributed) sequence of random variables taking the values -1 and $+1$ with equal probability, i.e.,

$$\mathbb{P}(Y_i = -1) = \mathbb{P}(Y_i = +1) = \frac{1}{2}.$$

For some initial value $x \in \mathbb{Z}$, we let

$$S_0 := x \qquad \text{and} \qquad S_n := x + Y_1 + \cdots + Y_n \quad \text{for } n \geq 1.$$

Then the stochastic process $\{S_n\}_{n=0,1,\dots}$ is a Markov chain and called the **simple random walk** starting from x.

Definition 1.3. A DTMC X is **homogeneous** if

$$\mathbb{P}\left(X_{n+1} = j \middle| X_n = i \right) = \mathbb{P}\left(X_1 = j \middle| X_0 = i \right), \quad \text{for any } n \in I \text{ and } i, j \in S.$$

Unless otherwise stated, we will assume that all DTMCs are homogeneous DTMCs from now on. This assumption essentially implies that the state space and the transition probabilities do not change over time.

Example 1.4. The simple random walk is a homogeneous DTMC.

1.1 The Distribution of DTMCs

We want to describe the distribution of (homogeneous) DTMCs. We first define the transition probabilities, which determine the probability to go to each state. This is encoded by a single matrix for homogeneous DTMCs. We first define the probabilities to go from state i to state j in 1 step.

Definition 1.5. The **1-step transition matrix** is

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

where

$$p_{ij} := \mathbb{P}\left(X_1 = j \middle| X_0 = i \right) \qquad \text{for } i, j \in S.$$

Since the transition matrix is a matrix of probabilities, its entries must be non-negative and conditionally on $X_0 = i$, the sum over the row must sum to 1.

Proposition 1.6

The transition matrix is a ***right stochastic matrix*** *i.e.*

1. $p_{ij} \geq 0$ *for any* $i, j \in S$.
2. $\sum_{j \in S} p_{ij} = 1$ *for any* $i \in S$.

Example 1.7. For the simple random walk, we have $S = \mathbb{Z}$ and

$$p_{i,i+1} = p_{i,i-1} = \frac{1}{2} \qquad \text{and} \qquad p_{i,j} = 0 \text{ otherwise.}$$

Next, we want to compute the probabilities to go from state i to state j in n steps.

Definition 1.8. More generally, we want to define the n**-step transition probabilities**

$$p_{ij}^{(n)} = \mathbb{P}\left(X_n = j \middle| X_0 = i \right) \text{ for } n \in \mathbb{N} \text{ and } i, j \in S,$$

and the corresponding n**-step transition matrix**

$$\boldsymbol{P}^{(n)} = \left(p_{ij}^{(n)} \right)_{i,j \in S}.$$

The next result shows that the n-step transition matrix can be deduced from the (1-step) transition matrix through matrix multiplication.

Proposition 1.9

For any $n \in \mathbb{N}$,

$$\boldsymbol{P}^{(n)} = \boldsymbol{P}^n.$$

That is, the n*-step transition matrix is equal to the* n^{th} *matrix power of the 1-step transition matrix.*

A direct consequence of this result means that we can get the $(n+m)$-step transition matrix by multiplying n-step and m-step transition matrices. Intuitively, the probability to go from state i to j in $n+m$ steps is equal to the sum over all intermediate states k of the probability to go from i to k in n steps and from k to j in m steps.

Corollary 1.10 (*Chapman-Kolmogorov equation*)

For any $m, n \in N$ *and* $i, j \in S$,

$$\boldsymbol{P}^{(n+m)} = \boldsymbol{P}^{(n)} \boldsymbol{P}^{(m)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(n)}$$

or, equivalently,

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}.$$

We are now interested in the marginal distribution of the stochastic process at time n. Let

$$\boldsymbol{\nu}^{(n)} = (\mathbb{P}(X_n = k))_{k \in S} = (\nu_k^{(n)})_{k \in S}$$

be the probability mass function of X_n. Notice that $\boldsymbol{\nu}^{(n)}$ is a $|S|$ dimensional row vector.

Definition 1.11. The **initial distribution** of the Markov chain $X = \{X_n\}_{n \geq 0}$ is the distribution of X_0,

$$\boldsymbol{\nu}^{(0)} = (\mathbb{P}(X_0 = k))_{k \in S} = (\nu_k^{(0)})_{k \in S}$$

We have the following formula for the ν_k.

Proposition 1.12 (***Marginal Distribution***)

For any $n \in \mathbb{N}$,

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)}\boldsymbol{P}^n.$$

Remark 1.13. It is important that we multiply $\nu^{(0)}$ on the right by $\boldsymbol{P}^n$. Since $\boldsymbol{P}$ is not necessarily symmetric, the vector $\boldsymbol{P}^{(n)}(\boldsymbol{\nu}^{(0)})^\intercal$ does not have the same meaning in terms of marginal probabilities. Furthermore, we also get that (see Problem 1.5)

$$\nu_k^{(n)} \geq 0 \text{ and } \sum_{k \in S} \nu_k^{(n)} = 1$$

which is consistent with the definition of a probability mass function.

The joint distribution of the vector $(X_0, \dots, X_n)$ is given by a similar formula through an application of the chain rule for conditional probabilities. That is, the probability you visit states $x_0, x_1, \dots x_n$ in that order is equal to the product of the probability you start at x_0, the probability you go from state x_1 to state x_2, etc.

Corollary 1.14 (***Joint Distribution***)

For any $x_0, x_1 \dots, x_n \in S$, we have

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n) = \nu_{x_0}^{(0)} p_{x_0x_1} p_{x_1x_2} \cdots p_{x_{n-1}x_n}$$

Remark 1.15. We have shown that the law of the entire DTMC is determined by its initial distribution and transition matrix.

1.2 Example Problems

1.2.1 Proofs of Results

Problem 1.1. Prove Proposition 1.9

Solution 1.1. We prove it by induction on n. First, by the definition of the 1-step transition matrix, we have $\boldsymbol{P}^{(1)} = \boldsymbol{P}$. Next, assume that $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$, we have

$$\begin{aligned}
p_{ij}^{(n+1)} &= \mathbb{P}(X_{n+1} = j \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j, X_n = k \mid X_0 = i) && \text{law of total probability} \\
&= \sum_{k \in S} \frac{\mathbb{P}(X_{n+1} = j, X_n = k, X_0 = i)}{\mathbb{P}(X_0 = i)} && \text{definition of cond. prob.} \\
&= \sum_{k \in S} \frac{\mathbb{P}(X_{n+1} = j, X_n = k, X_0 = i)}{\mathbb{P}(X_n = k, X_0 = i)} \cdot \frac{\mathbb{P}(X_n = k, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j \mid X_n = k, X_0 = i)\,\mathbb{P}(X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} \underbrace{\mathbb{P}(X_{n+1} = j \mid X_n = k)}_{=p_{kj}}\underbrace{\mathbb{P}(X_n = k \mid X_0 = i)}_{=p_{ik}^{(n)}} && \text{Markov property.} \\
&= \sum_{k \in S} p_{kj} p_{ik}^{(n)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}
\end{aligned}$$

The final expression is equal to the entry i, j of the matrix product $\boldsymbol{P}^{(n)}\boldsymbol{P}$. Using the induction hypothesis, $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$, we conclude that

$$p_{ij}^{(n+1)} = (\boldsymbol{P}^{(n)}\boldsymbol{P})_{ij} = (\boldsymbol{P}^n\boldsymbol{P})_{ij} = (\boldsymbol{P}^{n+1})_{ij}.$$

Problem 1.2. Prove the Chapman–Kolmogorov equation (Corollary 1.10)

Solution 1.2. From Proposition 1.9, we get

$$\boldsymbol{P}^{(n+m)} = \boldsymbol{P}^{n+m} = \boldsymbol{P}^n\boldsymbol{P}^m = \boldsymbol{P}^{(n)}\boldsymbol{P}^{(m)}.$$

Clearly, we also have $\boldsymbol{P}^m\boldsymbol{P}^n = \boldsymbol{P}^{n+m} = \boldsymbol{P}^n \cdot \boldsymbol{P}^m$ by the properties of matrix multiplication, so $\boldsymbol{P}^{(n+m)} = \boldsymbol{P}^{(m)}\boldsymbol{P}^{(n)}$. The formula

$$p_{ij}^{(m+n)} = \sum_{k\in S} p_{ik}^{(n)} p_{kj}^{(m)} = \sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(n)}$$

is the entrywise definition of the multiplication of matrices $\boldsymbol{P}^{(m)}$ and $\boldsymbol{P}^{(n)}$.

Problem 1.3. Prove the formula for the marginals in Proposition 1.12.

Solution 1.3. We have

$$\begin{aligned}
\nu_k^{(n)} = \mathbb{P}(X_n = k) &= \sum_{j\in S} \mathbb{P}(X_n = k \mid X_0 = j)\,\mathbb{P}(X_0 = j) && \text{Law of total probability} \\
&= \sum_{j\in S} p_{jk}^{(n)} \nu_j^{(0)} \\
&= \sum_{j\in S} \nu_j^{(0)} (\boldsymbol{P}^n)_{jk} && \text{Proposition 1.9}\ ,
\end{aligned}$$

so

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)}\boldsymbol{P}^n.$$

That is, $\nu^{(n)}$ is obtained by the multiplying the row vector $\nu^{(0)}$ on the right with the matrix $\boldsymbol{P}^n$.

Problem 1.4. Prove the formula for the joint distribution Corollary 1.14.

Solution 1.4. By the chain rule for conditional probabilities,

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) \\
&= \mathbb{P}(X_0 = x_0)\,\mathbb{P}(X_1 = x_1 | X_0 = x_0) \cdots \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \\
&= \nu_{x_0}^{(0)} p_{x_0x_1} p_{x_1x_2} \cdots p_{x_{n-1}x_n}.
\end{aligned}$$

Problem 1.5. If $\boldsymbol{P}$ is a transition matrix, define the row vector

$$\mathbf{x}^\intercal = \nu^{(0)}\boldsymbol{P}^n.$$

Show that

$$x_i \geq 0 \text{ and } \sum_{i\in S} x_i = 1.$$

That is, the row vector $\mathbf{x}^\intercal \in \mathbb{R}^{|S|}$ encodes a probability mass function.

Solution 1.5. Since $\nu^{(0)}$ and $\boldsymbol{P}$ are non-negative the product will also have non-negative entries. Furthermore, since $\boldsymbol{P}$ is a right stochastic matrix if $\mathbf{1} = (1, \dots, 1)^\intercal$ then

$$\boldsymbol{P}\mathbf{1} = \mathbf{1}.$$

Therefore,

$$\sum_{i \in S} x_i = \mathbf{x}^\intercal \mathbf{1} = \boldsymbol{\nu}^{(0)} \boldsymbol{P}^n \mathbf{1} = \boldsymbol{\nu}^{(0)} \mathbf{1} = \sum_{k \in S} \nu_k^{(0)} = 1 = 1$$

since $\nu^{(0)}$ is a probability vector so $\sum_{k \in S} \nu_k^{(0)} = 1$.

1.2.2 Applications

Problem 1.6. Consider a model for the state of a phone where $X_n = 1$ means the phone is free at time n and $X_n = 2$ means the phone is busy. If the phone is free, it will be busy during the next interval with probability p. If the phone is busy, it will be free during the next interval with probability q. Find the transition matrix for this DTMC.

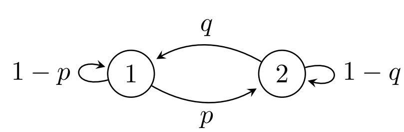
Solution 1.6. We model this with a DTMC with state space $S = \{1, 2\}$ where 1 corresponds to a free state, and 2 corresponds to a busy state. From the problem description, we have

$$p_{11} = 1 - p, \quad p_{12} = p, \quad p_{21} = q, \quad p_{22} = 1 - q,$$

so that the transition matrix becomes

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

The transition diagram is displayed below:



Problem 1.7. In the same setting as Problem 1.6 assume that the phone is free at time 0, and that $p = 1/4$ and $q = 1/6$. What is the probability that the phone is busy at time 6?

Solution 1.7. We have by Proposition 1.9

$$\boldsymbol{P} = \begin{pmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{pmatrix} \text{ and hence } \boldsymbol{P}^{(6)} = \boldsymbol{P}^6 = \begin{pmatrix} 0.424 & 0.576 \\ 0.384 & 0.616 \end{pmatrix},$$

with the initial distribution $\boldsymbol{\nu}^{(0)} = (1, 0)$ since the phone is free at time 0. By Proposition 1.12, we have

$$\boldsymbol{\nu}^{(6)} = \boldsymbol{\nu}^{(0)} \boldsymbol{P}^6 = (0.424, 0.576).$$

Hence, the probability that the phone is busy at time 6 equals 0.576.