Discrete time Markov chains

Susana Gomes

Last updated October 1, 2023

Outline

1. Definition of stochastic processes and Markov chains
2. Some examples
3. Some properties of Markov chains
4. Stationary and reversible distributions
5. Absorbing states

Discrete Time Markov Chains

Discrete-time stochastic processes, and the Markov property

We want to have an efficient description for processes such as the simple random walk from yesterday. We will start with some definitions and then will look at some useful examples and properties.

A **discrete-time stochastic process** with **state space** S is a sequence

$$Y_0, Y_1, \ldots = (Y_n : n \in \mathbb{N}_0)$$

of random variables taking values in S.

The process is called **Markov**, if for all $A \subset S$, $n \in \mathbb{N}_0$ and $s_0, \ldots, s_n \in S$ we have

$$\mathbb{P}(Y_{n+1} \in A | Y_n = s_n, \ldots, Y_0 = s_0) = \mathbb{P}(Y_{n+1} \in A | Y_n = s_n).$$

A Markov process (MP) is called **homogeneous** if for all $A \subset S$, $n \in \mathbb{N}_0$ and $s \in S$

$$\mathbb{P}(Y_{n+1} \in A | Y_n = s) = \mathbb{P}(Y_1 \in A | Y_0 = s).$$

If S is discrete, the MP is called a **Markov chain (MC)**.

Some notation

We will be working with the generic probability space Ω, which is the **path space**

$$\Omega = D(\mathbb{N}_0, S) := S^{\mathbb{N}_0} = S \times S \times \dots$$

Note that Ω is uncountable, even when S is finite.

For a given $\omega \in \Omega$ the function $n \mapsto Y_n(\omega)$ is called a **sample path**.

Up to finite time N and with finite S, $\Omega_N = S^{N+1}$ is finite.

Example 1

Consider the simple random walk from yesterday, with a random walker starting at $X_0 = 0$. We have

- $Y_0 = X_0 = 0$.
- The state space $S = \mathbb{Z}$.
- Up to time N, $\mathbb{P}$ is a distribution on the finite path space Ω_N with

$$\mathbb{P}(\omega) = \begin{cases} p^{\text{\# of right-steps}} q^{\text{\# of left-steps}} & \text{, path } \omega \text{ possible} \\ 0 & \text{, path } \omega \text{ not possible} \end{cases}$$

- There are only 2^N paths in Ω_N with non-zero probability.
- If $p = q = 1/2$ all paths have the same probability $(1/2)^N$.

More examples

- A **generalised random walk** is a random walk with $Y_0 = 0$ and increments $X_{n+1} = Y_{n+1} - Y_n \in \mathbb{R}$. This is a Markov process with $S = \mathbb{R}$ and $\Omega_N = \mathbb{R}^N$. It has an uncountable number of possible paths.

- A sequence $Y_0, Y_1, \ldots \in S$ of iid rv's is also a Markov process with state space S.

- Let $S = \{1, \ldots, 52\}$ be a deck of cards, and $Y_1, \ldots, Y_{52}$ be the cards drawn at random without replacement. Is this a Markov process?

Homogeneous MCs and the Chapman-Kolmogorov equations

Let $(X_n : n \in \mathbb{N}_0)$ be a homogeneous DTMC with **discrete** state space S.

Then we can define the **transition function**

$$p_n(x, y) := \mathbb{P}[X_n = y | X_0 = x] = \mathbb{P}[X_{k+n} = y | X_k = x] \quad \text{for all } k \geq 0.$$

Proposition: Chapman-Kolmogorov equations

If $(X_n : n \in \mathbb{N}_0)$ is a homogeneous DTMC, $p_n(x, y)$ is well defined and fulfills the **Chapman Kolmogorov equations**:

$$p_{k+n}(x, y) = \sum_{z \in S} p_k(x, z)\, p_n(z, y) \quad \text{for all } k, n \geq 0,\ x, y \in S.$$

Proof.

We use the law of total probability, the Markov property and homogeneity

Transition matrices

It is convenient to write all of this in matrix form. For this, we define the matrix $P_n = \big(p_n(x,y) : x, y \in S\big)$, and can rewrite the Chapman-Kolmogotov equations as:

$$P_{n+k} = P_n \, P_k$$

and, in particular,

$$P_{n+1} = P_n \, P_1 .$$

which gives us a recursion relation to the matrices P_n.

With $P_0 = \mathbb{I}$, the obvious solution to this recursion is $P_n = P^n$, where we define the **transition matrix**

$$P = P_1 = \big(p(x,y) : x, y \in S\big).$$

The transition matrix P and the initial condition $X_0 \in S$ completely determine a homogeneous DTMC, since for all $k \geq 1$ and all events $A_1, \ldots, A_k \subset S$

$$\mathbb{P}[X_1 \in A_1, \ldots, X_k \in A_k] = \sum_{s_1 \in A_1} \cdots \sum_{s_k \in A_k} p(X_0, s_1) p(s_1, s_2) \cdots p(s_{k-1}, s_k).$$

Properties of transition matrices and some notation

Note that there is no reason to have a fixed X_0 and instead we can work with an **initial distribution**

$$\pi_0(x) := \mathbb{P}[X_0 = x].$$

The distribution at time n is then

$$\pi_n(x) = \sum_{y \in S} \sum_{s_1 \in S} \cdots \sum_{s_{n-1} \in S} \pi_0(y) p(y, s_1) \cdots p(s_{n-1}, x)$$

In this case, we can also write

$$\langle \pi_n | = \langle \pi_0 | P^n,$$

where $\langle \cdot |$ denotes a row vector.

Finally, the transition matrix P is **stochastic**, i.e.

$$p(x, y) \in [0, 1] \quad \text{and} \quad \sum_{y \in S} p(x, y) = 1 \ ,$$

or equivalently, the column vector $|1\rangle = (1, \ldots, 1)^T$ is an **eigenvector** of P with **eigenvalue** 1: $\quad P|1\rangle = |1\rangle$

Quick example

Back to the simple random walk...

Example: Random walk with boundaries

To make our lives simpler, we can look at MCs with finite state space. A good example of this are random walks with boundaries.

Let $(X_n : n \in \mathbb{N}_0)$ be a simple random walk on $S = \{1, \ldots, L\}$ with

$$p(x, y) = p\delta_{y,x+1} + q\delta_{y,x-1}.$$

In this case, we need to tell it what happens once we reach the boundary (1 or L). We can have the following boundary conditions:

- **periodic** if $\quad p(L, 1) = p \,, \quad p(1, L) = q,$
- **absorbing** if $\quad p(L, L) = 1 \,, \quad p(1, 1) = 1,$
- **closed** if $\quad p(1, 1) = q \,, \quad p(L, L) = p,$
- **reflecting** if $\quad p(1, 2) = 1 \,, \quad p(L, L-1) = 1.$

Stationary and reversible distributions

Let $(X_n : n \in \mathbb{N}_0)$ be a homogeneous DTMC with state space S.

We say that the distribution $\pi(x),\ x \in S$ is **stationary** if for all $y \in S$

$$\sum_{x \in S} \pi(x) p(x, y) = \pi(y) \quad \text{or} \quad \langle \pi | P = \langle \pi |.$$

π is called **reversible** if it fulfills the **detailed balance** conditions

$$\pi(x) p(x, y) = \pi(y) p(y, x) \quad \text{for all } x, y \in S.$$

Note that if a distribution π is reversible then it is stationary:

$$\sum_{x \in S} \pi(x) p(x, y) = \sum_{x \in S} \pi(y) p(y, x) = \pi(y).$$

Existence of stationary distributions

If we see a stationary distribution as a row vector $\langle\pi| = (\pi(x) : x \in S)$, then we can see that stationary distributions are **left eigenvectors** of P with **eigenvalue** 1:

$$\langle\pi| = \langle\pi|P.$$

This immediately implies the following:

Existence

Every DTMC **with finite state space** has at least one stationary distribution.

Proof. Since P is stichastic, we have $P|\mathbf{1}\rangle = |\mathbf{1}\rangle$.
So 1 is an eigenvalue of P, and its corresponding left eigenvector(s) can be shown to have non-negative entries and thus can be normalized to be stationary distribution(s) $\langle\pi|$.

This last part relies on the Perron-Frobenius theorem, which we will look at in more detail later.

Example

However, there is no reason for a stationary distribution to be unique...

Irreducible DTMCs

A DTMC is called **irreducible**, if for all $x, y \in S$ we have

$$p_n(x, y) > 0 \text{ for some } n \in \mathbb{N}.$$

Using the Perron-Frobenius theorem again, we will be able to prove (in a week or so) that irreducible DTMCs have a **unique** stationary distribution.

Absorbing states

A state $s \in S$ is called **absorbing** for a DTMC with transition matrix $p(x, y)$, if

$$p(s, y) = \delta_{s,y} \quad \text{for all } y \in S.$$

In other words, once the chain reached the state s, it won't leave it.

A good example of a DTMC with an absorbing state is a **simple Random Walk with absorbing boundary conditions.**

Recall:

A simple random walk in $S = \{1, \dots, L\}$ with absorbing BCs verifies:

- $p(x, y) = p\delta_{y,x+1} + q\delta_{y,x-1}$
- $p(L, L) = 1$, $p(1, 1) = 1$.

We can use this to compute interesting things, like the **absorption probability**. In practice, we can also have absorbing "cycles", or other interesting behaviours without absorbing states. We will say more about this later.

Absorption probability

Let h_k be the **absorption probability** for $X_0 = k \in S = \{1, \dots, L\}$, i.e.

$$h_k = \mathbb{P}(\text{absorption}|X_0 = k) = \mathbb{P}(X_n \in \{1, L\} \text{ for some } n \geq 0 | X_0 = k).$$

First of all, we can clearly see that $h_1 = h_L = 1$.

Now, for the other starting points, we condition on the first jump and use the Markov property:

$$h_k = ph_{k+1} + qh_{k-1} \quad \text{for } k = 2, \dots, L-1,$$

and this gives us a recursive relation.

Assume that the solution is of the form $\quad h_k = \lambda^k\,, \quad \lambda \in \mathbb{C}$.

Distribution at time n

The last topic we will consider for DTMCs is their distribution at time n.

Consider a DTMC on a finite state space with $|S| = L$, and let $\lambda_1, \ldots, \lambda_L \in \mathbb{C}$ be the **eigenvalues** of the transition matrix P with corresponding

left (row) eigenvectors $\langle u_i|$ and **right (column) eigenvectors** $|v_i\rangle$.

If we assume that **all eigenvalues are distinct** we can always write

$$P = \sum_{i=1}^{L} \lambda_i |v_i\rangle\langle u_i| \quad \text{and} \quad P^n = \sum_{i=1}^{L} {\lambda_i}^n |v_i\rangle\langle u_i|,$$

since eigenvectors can be chosen **orthonormal** $\langle u_i|v_j\rangle = \delta_{i,j}$.

Since $\langle \pi_n| = \langle \pi_0|P^n$ we get

$$\langle \pi_n| = \langle \pi_0|v_1\rangle {\lambda_1}^n \langle u_1| + \ldots + \langle \pi_0|v_L\rangle {\lambda_L}^n \langle u_L|.$$

Distribution at time *n* - some observations

The **Gershgorin theorem** (see handout 1) implies that $|\lambda_i| \leq 1$.

This also means that contributions with $|\lambda_i| < 1$ decay exponentially.

$\lambda_1 = 1$ corresponds to the **stationary distribution** $\langle\pi| = \langle u_1|$ and $|v_1\rangle = |1\rangle$.

Other $\mathbb{C} \ni \lambda_i \neq 1$ with $|\lambda_i| = 1$ correspond to **persistent oscillations**.