WARWICK
THE UNIVERSITY OF WARWICK

Lecture 3:
Discrete time Markov chains (continued) and intro to continuous time Markov chains

Susana Gomes

October 14^{th}, 2021

Plan for today

1. Stationary and reversible distributions
2. Absorbing states
3. Basics of Continuous time Markov chains

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...

$$p_n(x,y) = \mathbb{P}(X_n = y \mid X_0 = x) > 0$$

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 $S = \{1, \ldots, L\}$ 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:

$$h_k = p h_{k+1} + q h_{k-1} \quad \text{for } k = 2, \dots, L-1,$$

and this gives us a recursive relation.

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

$\lambda^k = p\lambda^{k+1} + q\lambda^{k-1} \Rightarrow \lambda = p\lambda^2 + q$

$\lambda_1 = 1$ or $\lambda_2 = q/p$

$\Rightarrow h_k = a\lambda_1^k + b\lambda_2^k = a + b\left(\frac{q}{p}\right)^k$

$h_1 = 1 \Rightarrow a + b\frac{q}{p} = 1$ $\quad a = 1 \Rightarrow h_k = 1$

$h_L = 1 \Rightarrow a + b\left(\frac{q}{p}\right)^L = 1$ $\quad b = 0$ $\quad \forall k$

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**.

WARWICK
THE UNIVERSITY OF WARWICK

Continuous time Markov chains

Continuous-time Markov chains

We'll spend the rest of today (and probably tomorrow) generalising the previous results for the case of **continuous time**. So... as before:

- A **continuous-time stochastic process** with state space S is a family $(X_t : t \geq 0)$ of random variables taking values in S.
- The process is called **Markov** if, for all $A \subset S$, $n \in \mathbb{N}$, $t_1 < \ldots < t_{n+1} \in [0, \infty)$ and $s_1, \ldots, s_n \in S$, we have

$$\mathbb{P}(X_{t_{n+1}} \in A | X_{t_n} = s_n, \ldots, X_{t_1} = s_1) = \mathbb{P}(X_{t_{n+1}} \in A | X_{t_n} = s_n).$$

- A Markov process (MP) is called **homogeneous** if for all $A \subset S$, $t, u > 0$ and $s \in S$

$$\mathbb{P}(X_{t+u} \in A | X_u = s) = \mathbb{P}(X_t \in A | X_0 = s).$$

- If S is discrete, the MP is called a continuous-time **Markov chain (CTMC)**.

A couple of technical details

The generic probability space Ω of a CTMC is the space of **right-continuous paths**

$$\Omega = D([0,\infty), S) := \left\{ X : [0,\infty) \to S \,\middle|\, X_t = \lim_{u \searrow t} X_u \right\}$$

$\mathbb{P}$ is a probability distribution on Ω.
By **Kolmogorov's extension theorem** $\mathbb{P}$ is fully specified by its **finite dimensional distributions (FDDs)**. These are the distributions of the form

$$\mathbb{P}[X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n] \,, \quad n \in \mathbb{N},\ t_i \in [0,\infty),\ A_i \subset S \,.$$

Chapman-Kolmogorov equations for CTMCs

We can still define a transition function as before and the Chapman-Kolmogorov equations are valid.

Let $(X_t : t \geq 0)$ by a homogeneous CTMC with state space S. Then for all $t \geq 0$ the **transition function** is given by

$$p_t(x, y) := \mathbb{P}[X_t = y | X_0 = x] = \mathbb{P}[X_{t+u} = y | X_u = x] \quad \text{for all } u \geq 0.$$

Proposition: Chapman-Kolmogorov equations

The transition function is well defined and fulfills the **Chapman Kolmogorov equations**

$$p_{t+u}(x, y) = \sum_{z \in S} p_t(x, z)\, p_u(z, y) \quad \text{for all } t, u \geq 0,\ x, y \in S.$$

Generator of a CTMC

As before, we can write this in matrix notation:

We define $P_t = \big(p_t(x,y) : x,y \in S\big)$ and we can write

$$P_{t+u} = P_t\, P_u \quad \text{with} \quad P_0 = \mathbb{I}.$$

In particular,

$$\frac{P_{t+\Delta t} - P_t}{\Delta t} = P_t \frac{P_{\Delta t} - \mathbb{I}}{\Delta t} = \frac{P_{\Delta t} - \mathbb{I}}{\Delta t} P_t.$$

We now take $\Delta t \searrow 0$ and get the so-called **forward and backward equations**

$$\frac{d}{dt} P_t = P_t G = G P_t \,, \quad \text{where} \quad G = \frac{dP_t}{dt}\bigg|_{t=0}$$

is called the **generator** of the process (sometimes also Q-matrix).

Forward and backward equations for CTMCs

The solution for these equations is given by the matrix exponential

$$P_t = \exp(tG) = \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = \mathbb{I} + tG + \frac{t^2}{2} G^2 + \dots$$

And from this we can obtain the **distribution π_t at time $t > 0$**:

$$\langle \pi_t | = \langle \pi_0 | \exp(tG) \quad \text{which solves} \quad \frac{d}{dt} \langle \pi_t | = \langle \pi_t | G.$$

Note that, just like before, if S is finite, we can compute the eigenvalues of G, $\lambda_1, \dots, \lambda_L \in \mathbb{C}$. Then, **$P_t$ has eigenvalues $\exp(t\lambda_i)$** with the same eigenvectors $\langle v_i |$, $| u_i \rangle$.

If the λ_i are distinct, we can still expand the initial condition in the eigenvector basis

$$\langle \pi_0 | = \alpha_1 \langle v_1 | + \dots + \alpha_L \langle v_L |,$$

where $\alpha_i = \langle \pi_0 | u_i \rangle$. This leads to

$$\langle \pi_t | = \alpha_1 \langle v_1 | e^{\lambda_1 t} + \dots + \alpha_L \langle v_L | e^{\lambda_L t} \, .$$

Transition rates

Using the expression for P_t we have, for $G = \big(g(x,y) : x,y \in S\big)$,

$$p_{\Delta t}(x,y) = g(x,y)\Delta t + o(\Delta t) \quad \text{for all } x \neq y \in S,$$

so the $g(x,y) \geq 0$ can be interpreted as **transition rates**.

We also have

$$p_{\Delta t}(x,x) = 1 + g(x,x)\Delta t + o(\Delta t) \quad \text{for all } x \in S.$$

Since $\sum_y p_{\Delta t}(x,y) = 1$, this implies that

$$g(x,x) = -\sum_{y \neq x} g(x,y) \leq 0 \quad \text{for all } x \in S.$$

The Master equation

Using the results from the previous slide, we can rewrite the equation for the distribution at time t:

$$\langle \pi_t | = \langle \pi_0 | \exp(tG) \quad \text{which solves} \quad \frac{d}{dt}\langle \pi_t | = \langle \pi_t | G,$$

as the **Master equation**

$$\frac{d}{dt}\pi_t(x) = \underbrace{\sum_{y \neq x} \pi_t(y) g(y,x)}_{\text{gain term}} - \underbrace{\sum_{y \neq x} \pi_t(x) g(x,y)}_{\text{loss term}} \quad \text{for all } x \in S\,.$$

Note that: the Gershgorin theorem now implies that either $\lambda_i = 0$ or $\mathrm{Re}(\lambda_i) < 0$ for the eigenvalues of G, so there are **no persistent oscillations for CTMCs**.