

Randomness and Computation

or, “Randomized Algorithms”

Mary Cryan

School of Informatics
University of Edinburgh



RC (2018/19) – Lecture 14 – slide 1

Markov chains

Definition (Definition 7.1)

A discrete-time stochastic process on the state space Ω is said to be a *Markov chain* if

$$\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \dots, X_0 = a_0] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}].$$

We will often denote the Markov chain by the name M , and write $M[a_{t-1}, a_t]$ to denote the probability $\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}]$.



RC (2018/19) – Lecture 14 – slide 3

Markov processes

- ▶ A *Markov process* is a r.v $\mathbf{X} = \{X(t) : t \in T\}$, usually $T = \mathbb{N}^0$.
- ▶ $X(t)$ (sometimes written as X_t) is the state of the process at time $t \in T$, this is an element of some state set Ω , usually a discrete finite set, sometimes countably infinite.
- ▶ A Markov process with the *memoryless property* is one where X_{t+1} will depend on the previous state X_t , but on none of the previous states.



RC (2018/19) – Lecture 14 – slide 2

Markov chains

We can think about the *Markov chain* M in terms of a matrix of dimensions $|\Omega| \times |\Omega|$ (if Ω is finite) or of infinite dimension if Ω is countably infinite.

$$\begin{bmatrix} M[a_1, a_1] & M[a_1, a_2] & \dots & M[a_1, a_j] & \dots \\ M[a_2, a_1] & M[a_2, a_2] & \dots & M[a_2, a_j] & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ M[a_j, a_1] & M[a_j, a_2] & \dots & M[a_j, a_j] & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

This is often called the *transition matrix* of the Markov chain.



RC (2018/19) – Lecture 14 – slide 4

Transition matrix

This representation assumes that the *states* of the Markov chain have been put in 1-1 correspondance with \mathbb{N} . This is certainly possible (as we assume Ω is countably infinite at worst) but it's not *always* the case that this ordering is particularly *natural*. In many situations (especially when Ω is finite but exponentially sized), we'll avoid writing down the *transition matrix* at all - as an example, see the description of the contingency tables 2×2 chain.

Note that for every $a \in \Omega$, $\sum_{b \in \Omega} M[a, b] = 1$ (so *a*'s-row sums to 1).

Iterations of the Markov chain

- ▶ If we then carry out a second step of the Markov chain, the random variable $X(2)$ will then be distributed according to $\bar{p}(2)$, where

$$\bar{p}(2) = \bar{p}(1) \cdot M = \bar{p}(0) \cdot M \cdot M = \bar{p}(0) \cdot M^2.$$

- ▶ And so on ...

After t steps of the Markov chain M , the random variable $X(t)$ will then be distributed according to $\bar{p}(t)$, where

$$\bar{p}(t) = \bar{p}(0) \cdot M^t.$$

Iterations of the Markov chain

Suppose we start our Markov process with the initial state $X(0)$ being some fixed $a \in \Omega$.

- ▶ The “next state” $X(1)$ is distributed according to *a*'s row of the transition matrix M (with probability $M[a, b]$ of $X(1)$ becoming b).
- ▶ If we define $\bar{p}(0)$ to be the row vector with $\bar{p}_a(0) = 1$ and all other entries 0, then we can define the probability distribution $\bar{p}(1)$ by

$$\bar{p}(1) = \bar{p}(0) \cdot M,$$

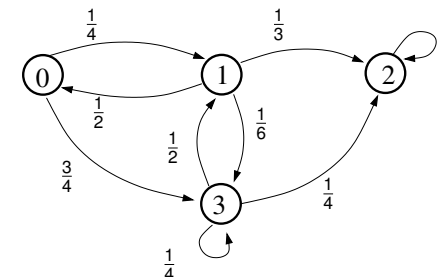
and the probability that $X(1)$ is b at the next step is $p_b(1)$ (which equals $M[a, b]$), for every $b \in \Omega$.

- ▶ The “next state” $X(1)$ is a *random variable*, and its distribution is the vector $\bar{p}(1)$ above (with values summing to 1).

Example Markov chain

Let's look at a concrete example of a Markov chain on the *state set* $\Omega = \{0, 1, 2, 3\}$, with probabilities given by the following *transition matrix*:

$$M = \begin{bmatrix} 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & 0 & \frac{1}{3} & \frac{1}{6} \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$



Example Markov chain

- ▶ Notice that every row of the matrix sums to 1, as required.
 - ▶ Not the case for the columns (and doesn't have to be).
 - ▶ In the graph representation, the "states" are vertices and the "transitions" of the chain are directed edges (labelled with the appropriate probability).
 - ▶ Consider running this Markov chain from $3 \in \Omega$. Then we set $\bar{p}(0) = (0 \ 0 \ 0 \ 1)$.
 - ▶ Evaluating $\bar{p}(1) = \bar{p}(0) \cdot M$, we note it will be $(0 \ \frac{1}{2} \ \frac{1}{4} \ \frac{1}{4})$.
 - ▶ Evaluating $\bar{p}(2) = \bar{p}(0) \cdot M^2$, we can compute it as $(\frac{1}{4} \ \frac{1}{8} \ \frac{23}{48} \ \frac{7}{48})$.
 - ▶ And so on ...
 - ▶ We would expect these $\bar{p}(t)$ values to be different if we had started with a different initial state.
- However, these alternate t -step distributions can be easily calculated/compared, if we have pre-computed M^t .

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

RC (2018/19) – Lecture 14 – slide 9

Markov chains for Randomized Algorithms

Another application of Markov chains arises when we want to *randomly sample* (or *approximately count*) the number of combinatorial structures satisfying some constraint.

- ▶ In these cases the elements of the state space Ω are individual combinatorial structures themselves, and this space is often very large (potentially exponential in the size of the input).
- ▶ One example we can imagine is the set of *contingency tables* $\Sigma_{r,c}$ for given row sums $r = (r_1, \dots, r_m)$ and column sums $c = (c_1, \dots, c_n)$.
- ▶ With large state spaces of combinatorial elements, we don't explicitly "write down" the transition matrix of a Markov chain on the space - we just give (randomized) rules explaining how we move (randomly) from one element to another.
- ▶ In Coursework 2, you are asked to prove that the " 2×2 " Markov chain converges to the uniform distribution on $\Omega = \Sigma_{r,c}$.

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

RC (2018/19) – Lecture 14 – slide 11

Markov chains for Randomized Algorithms

The graph (and transition matrix) on slide is just a "toy" example. In our world, we will want to exploit Markov chains to obtain randomized algorithms.

- ▶ In the 2SAT example of Section 7.1, [MU], we are concerned with *assignments* (functions) A of boolean values to the n logical variables (and which will achieve the max number of satisfied clauses).
- ▶ The transitions of the modified Markov chain for 2SAT (Y) are designed to model transitions back and forth between "better" and "worse" assignments.
- ▶ However, the *states* of the Markov chain (Y) are the natural numbers $0, 1, \dots, n$, with n being the number of variables. This is a fairly small state space.
- ▶ Next Tuesday we will prove that for a 2SAT instance which *does* have a satisfying assignment, the *expected number of steps* before the Markov chain finds one is at most n^2 . This is expected "hitting time".

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

RC (2018/19) – Lecture 14 – slide 10

Irreducibility, recurrence, aperiodicity

To analyse the behaviour (both temporary and long-term) of a Markov chain M on the state space Ω , we often work from the *graph representation* of M 's transitions.

Definition (Definition 7.2)

For $a, b \in \Omega$, we say that state b is accessible from state a if there is some $n \in \mathbb{N}$ such that $M^n[a, b] > 0$. If a, b are both accessible from one another, we say that they *communicate*, and write $a \leftrightarrow b$.

Definition (Definition 7.3)

A Markov chain M on the finite state space Ω is said to be *irreducible* if all states of Ω belong to the same communicating class.

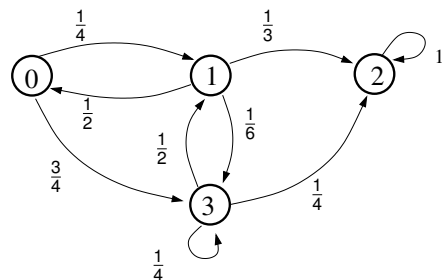
Or *equivalently*, if we can show that for every $X, Y \in \Omega$, there is some path $Z^0 = X, Z^1, \dots, Z^\ell = Y$ of states from Ω connecting X and Y , such that $M[Z^j, Z^{j+1}] > 0$ for every $0 \leq j \leq \ell - 1$.

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

RC (2018/19) – Lecture 14 – slide 12

Irreducibility, recurrence, aperiodicity

Let's consider our example chain:



- ▶ $\{0, 1, 3\}$ forms a maximal communicating class in the graph.
- ▶ $\{2\}$ is an isolated communicating class - we can reach 2 from $\{0, 1, 3\}$, but 2 has no outgoing transitions.
- ▶ Our example is *not* irreducible.

RC (2018/19) – Lecture 14 – slide 13

Irreducibility, recurrence, aperiodicity

The expected “time to travel” will be important for some analyses. For a pair of given states $Z, W \in \Omega$, we define this value as

$$h_{Z,W} =_{\text{def}} \sum_{t=0}^{\infty} t \cdot r_{Z,W}^t.$$

Definition (Definition 7.5)

A recurrent state Z of a Markov chain M is *positive recurrent* if $h_{Z,Z} < \infty$. Otherwise, it is *null recurrent*.

Lemma (Lemma 7.5)

In a Markov chain M on a finite state space Ω ,

1. At least one state is recurrent;
2. All recurrent states are positive recurrent.

RC (2018/19) – Lecture 14 – slide 15

Irreducibility, recurrence, aperiodicity

We define the concept of “recurrence” in terms of a parameter $r_{Z,W}^t$, where

$$r_{Z,W}^t =_{\text{def}} \Pr[X(t) = W \text{ and for all } 1 \leq s \leq t-1, X(s) \neq W \mid X(0) = Z].$$

Definition (Definition 7.4)

If M is a Markov chain and $Z \in \Omega$ a state of that chain, we say Z is *recurrent* if $\sum_{t=0}^{\infty} r_{Z,Z}^t = 1$, and it is *transient* if $\sum_{t=0}^{\infty} r_{Z,Z}^t < 1$.

A Markov chain is recurrent if every state is recurrent.

In our example, the only recurrent state is 2.

RC (2018/19) – Lecture 14 – slide 14

Irreducibility, recurrence, aperiodicity

Definition (Definition 7.6)

A state Z of a Markov chain M is *periodic* if there exists an integer $k \geq 2$ such that $\Pr[X(t+k) = Z \mid X(t) = Z]$ is non-zero *if and only if* k is divisible by k .

A discrete Markov chain is *periodic* if any of its states are periodic. Otherwise (all states aperiodic), we say that M is *aperiodic*.

Definition (Definition 7.7)

An aperiodic, positive recurrent state $Z \in \Omega$ of a Markov chain M is said to be an *ergodic* state.

The Markov chain is said to be *ergodic* if *all* its states are ergodic.

Corollary (Corollary 7.6)

Any finite, irreducible and aperiodic Markov chain is an ergodic chain.

RC (2018/19) – Lecture 14 – slide 16

Ergodic Markov chains

Corollary (Corollary 7.6)

Any finite, irreducible and aperiodic Markov chain is an ergodic chain.

Proof.

By Lemma 7.5 every chain with a finite state space has *at least one* recurrent state.

If the chain is irreducible, then all states can be reached from one another (with positive probability), so *all states are hence recurrent*.

By Lemma 7.5, this means all states are positive recurrent.

Hence by Definition 7.7 (we have positive recurrence, irreducibility and aperiodicity) the chain is ergodic. \square

Take-away: For Markov chains over a *finite* state space, we only need to check *aperiodicity* and *irreducibility* (for all pairs of states) to show *ergodicity*.



RC (2018/19) – Lecture 14 – slide 17

Ergodic Markov chains

Theorem (Theorem 7.7)

Any finite, irreducible and aperiodic Markov chain has the following properties:

1. The chain has a unique stationary distribution $\bar{\pi} = (\pi_1, \dots, \pi_{|\Omega|})$.
2. For all $X, Y \in \Omega$, $\lim_{t \rightarrow \infty} M^t[X, Y]$ exists and is independent of X .
3. $\pi_Y = \lim_{t \rightarrow \infty} M^t[X, Y] = \frac{1}{h_{Y,Y}}$.

We are not going to prove Theorem 7.7 in this course.

We will later see some ways of *bounding the number of steps needed* for the chain to approach its stationary distribution.



RC (2018/19) – Lecture 14 – slide 19

Ergodic Markov chains

Why do we care about states/Markov chains being ergodic?

Because an ergodic Markov chain has a *unique stationary distribution*.

Definition (Definition 7.8)

A *stationary distribution* of a Markov chain M on Ω is a probability distribution $\bar{\pi}$ on Ω such that

$$\bar{\pi} = \bar{\pi} \cdot M.$$

Stationary distributions are essentially “settling-down” distributions (in the limit). The “settling-down” is to a *distribution*, not a particular state.



RC (2018/19) – Lecture 14 – slide 18

Reading and Doing

Reading

- ▶ You will want to read Sections 7.1, 7.2, and 7.3 from the book. I have used M (name of the Markov chain) to refer to the transition matrix, as opposed to the book's P .

Doing

- ▶ Consider an example of contingency tables where we have $r = (2, 2, 4)$, $c = (2, 3, 3)$. Suppose that we take the following *state* as our starting state X :

$$X = \begin{array}{ccc|c} 2 & 0 & 0 & 2 \\ 0 & 2 & 0 & 2 \\ 0 & 1 & 3 & 4 \\ \hline 2 & 3 & 3 & \end{array}$$

Work out the subset of contingency tables which can be reached from X in *one transition* of the Markov chain. Also work out the probability of each such transition.

- ▶ Tutorial sheet for week 8 (which includes the above question).



RC (2018/19) – Lecture 14 – slide 20