Randomness and Computation
or, “Randomized Algorithms”

Mary Cryan
School of Informatics
University of Edinburgh

Markov processes

- A Markov process is a r.v $X = \{X(t) : t \in T\}$, 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 $\Omega$, usually a discrete finite set, sometimes countably infinite.
- A Markov process with the memoryless property is one where $X_{t+1}$ will depends on the previous state $X_t$, but on none of the previous states.

Markov chains

Definition (Definition 7.1)
A discrete-time stochastic process on the state space $\Omega$ is said to be a Markov chain if

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

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

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

This representation assumes that the states of the Markov chain have been put in 1-1 correspondence with \( \mathbb{N} \). This is certainly possible (as we assume \( \Omega \) is countably infinite at worst) but it’s not always the case that this ordering is particularly natural. In many situations (especially when \( \Omega \) is finite but exponentially sized), we will 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).

Example Markov chain

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

\[
M = \begin{bmatrix}
0 & 1/4 & 0 & 3/4 \\
1/2 & 0 & 1/2 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1/4 & 1/4 & 1/4 \\
\end{bmatrix}
\]

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

\[
\bar{p}(t) = \bar{p}(0) \cdot M^t.
\]
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 $\hat{p}(0) = (0 \ 0 \ 0 \ 1)$. 
- Evaluating $\hat{p}(1) = \hat{p}(0) \cdot M$, we note it will be $(0 \ \frac{1}{2} \ \frac{1}{4} \ \frac{1}{4})$.
- Evaluating $\hat{p}(2) = \hat{p}(0) \cdot M^2$, we can compute it as $(\frac{1}{4} \ \frac{1}{8} \ \frac{33}{8} \ \frac{7}{8})$.
- And so on ...
- We would expect these $\hat{p}(t)$ 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 $M^t$.

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 $\Omega$ 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 $r = (r_1, \ldots, r_m)$ and column sums $c = (c_1, \ldots, c_n)$.
- 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 this week’s tutorial, you are asked to prove that the “$2 \times 2$” Markov chain converges to the uniform distribution on $\Omega = \Sigma_{r,c}$.

Irreducibility, recurrence, aperiodicity

To analyse the behaviour (both temporary and long-term) of a Markov chain $M$ on the state space $\Omega$, 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 $\Omega$ is said to be irreducible if all states of $\Omega$ belong to the same communicating class. Or equivalently, if we can show that for every $X, Y \in \Omega$, there is some path $Z^0 = X, Z^1, \ldots, Z^t = Y$ of states from $\Omega$ connecting $X$ and $Y$, such that $M[Z^j, Z^{j+1}] > 0$ for every $0 \leq j \leq t - 1$. 

RC (2017/18) – Lecture 13 – slide 10

RC (2017/18) – Lecture 13 – slide 9

RC (2017/18) – Lecture 13 – slide 12

RC (2017/18) – Lecture 13 – slide 11

RC (2017/18) – Lecture 13 – slide 10
Irreducibility, recurrence, aperiodicity

Let's consider our example chain:

0 -> 1 -> 2

- \{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 (2017/18) – Lecture 13 – slide 13

Irreducibility, recurrence, aperiodicity

We define the concept of "recurrence" in terms of a parameter \( r_{Z,W}^t \), where

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

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 (2017/18) – Lecture 13 – slide 14

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

\[
h_{Z,W} = \sum_{t=0}^{\infty} t \cdot r_{Z,W}^t.
\]

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 \( \Omega \),

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

RC (2017/18) – Lecture 13 – slide 15

Definition (Definition 7.6)

A state \( Z \) of a Markov chain \( M \) is periodic if there exists an integer \( k \geq 2 \) such that \( \Pr[X(t + s) = Z \mid X(t) = Z] \) is non-zero if and only if \( s \) 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 (2017/18) – Lecture 13 – 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.

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.

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 $\Omega$ is a probability distribution $\pi$ on $\Omega$ such that

$$\pi = \pi \cdot M.$$ 

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

Theorem (Theorem 7.7)
Any finite, irreducible and aperiodic Markov chain has the following properties:
1. The chain has a unique stationary distribution $\pi = (\pi_1, \ldots, \pi_\Omega)$.
2. For all $X, Y \in \Omega$, $\lim_{t \to \infty} M^t[X, Y]$ exists and is independent of $X$.
3. $\pi_Y = \lim_{t \to \infty} M^t[X, Y] = \frac{1}{\pi_X}$.

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.

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 $X$:

$$X = \begin{pmatrix} 2 & 0 & 0 & 2 \\ 0 & 2 & 0 & 2 \\ 0 & 1 & 3 & 4 \\ 2 & 3 & 3 & 2 \end{pmatrix}$$

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.
- Exercises 7.1 and 7.2 from the [MU] book.