Long Term Behaviour of Markov Chains

1 Summary of known facts pertinent to the topic

Let $X_n,\ n = 0, 1, 2, ...$ be a MC with the state space $S = (1, 2, ...m)$, transition probabilities $p_{ij} \stackrel{\text{def}}{=} P\{X_{n+1} = j \,|\, X_n = i\}$, and the transition matrix

$$\mathbb{P} = (p_{ij}) = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}$$

Theorem 1.1
Put $p_{ij}^{(n)} \stackrel{\text{def}}{=} P\{X_n = j \,|\, X_0 = i\}$ *and let* $\mathbb{P}^{(n)} \stackrel{\text{def}}{=} (p_{ij}^{(n)})$, $1 \le i, j \le m$, *be the* $m \times m$ *matrix of these probabilities. Then*

$$\mathbb{P}^{(n)} = \mathbb{P}^n. \tag{1}$$

Formula (1) is equivalent to saying that the probability for X_n to reach state j after n transitions starting from i is equal to the (i, j)-th element of the n-th power of $\mathbb{P}$.

Suppose that the distribution of X_0 is given by $\underline{\mu^{(0)}} = (\mu_1^{(0)}, \mu_2^{(0)}, ..., \mu_m^{(0)})$, where $\mu_0^i \stackrel{\text{def}}{=} P\{X_0 = i\}$. Let $\mu_i^{(n)} \stackrel{\text{def}}{=} P\{X_n = i\}$.

Lemma 1.2
$\underline{\mu^{(n)}} = \underline{\mu^{(0)}}\mathbb{P}^n$

Proof. $P\{X_n = j\} \stackrel{\text{by Tot. Prob. Law}}{=} \sum_{i=1}^m \mu_i^{(0)} P\{X_n = j \,|\, X_0 = i\} = (\underline{\mu^{(0)}}\mathbb{P}^n)_j$. □

We say that $\underline{\pi} = (p_1, ..., p_m)$ is a *probability vector* if all its components $p_i \ge 0$ and their sum is 1: $p_1 + ... + p_m = 1$.

Definition 1.3
A probability vector $\underline{\pi}$ *is an equilibrium distribution of a MC if* $\underline{\pi}\mathbb{P} = \underline{\pi}$.

The probabilistic meaning of this notion is explained by the following

Theorem 1.4
If the initial distribution of the MC X_t *coincides with its equilibrium distribution then* $P\{X_n = j\} = p_j$ *for all* n *(and* j*).*

Proof. We have to show that $\underline{\mu^{(n)}} = \underline{\pi}$. By the assumption of our theorem, $\underline{\mu^{(0)}} = \underline{\pi}$ and hence, by Lemma 1.2

$$\underline{\mu^{(n)}} = \underline{\mu^{(0)}}\mathbb{P}^n = \underline{\pi}\mathbb{P}^n.$$

But

$$\underline{\pi}\mathbb{P}^n = \underline{\pi}\mathbb{P}\mathbb{P}^{n-1} = \underline{\pi}\mathbb{P}^{n-1} = \underline{\pi}\mathbb{P}^{n-2} = ... = \underline{\pi}.$$

□

2 The questions we want to study

There is a wide variety of questions which are often asked about the long term behaviour of Markov chains. In fact the larger part of the theory of Markov chains is the one studying different aspects of their long term behaviour. Here are examples of such questions and these are the ones we are going to discuss in this course.

1. Suppose that $X_0 = i$.

   What can be said about $P\{X_n = j \mid X_0 = i\}$ as n is increasing? More precisely, is there a limit $\lim_{n\to\infty} P\{X_n = j \mid X_0 = i\}$?

   If yes, does this limit depend on i? and can it be found?

2. For large n, what is the proportion of time the chain would spend in state i?

   It is meant that every time the MC reaches i it spends one unit of time in i. This
   $$n_i \stackrel{\text{def}}{=} \#\{k \text{ such that } X_k = i \text{ and } k \le n\ \}$$
   and the proportion of time spent by the chain in i by time n is $\frac{n_i}{n}$. We thus ask:

   Is there a limit $\lim_{n\to\infty} \frac{n_i}{n}$? And if yes, can it be found?

3. Let $f(i)$ be a real valued function on S. Is there a limit
   $$\lim_{n\to\infty} \frac{1}{n}(f(X_0) + f(X_1) + ... + f(X_{n-1})$$
   We shall show that the second question is a particular case of this one.

3 Example: the two state chain

If $S = \{1, 2\}$ then the transition matrix is

$$\mathbb{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

where $0 \le \alpha, \beta \le 1$.

Lemma 3.1
Suppose that $\alpha + \beta \neq 0$. Then

$$\mathbb{P}^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \tag{2}$$

Proof. We use induction. Base of induction: $n = 1$. We have to check that

$$\mathbb{P} = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{1-\alpha-\beta}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$

which is a straightforward computation. Suppose that the statement holds for some $n \geq 1$; we have to deduce from here that it holds for $n+1$ (the main step of induction). Indeed

$$\begin{aligned}
\mathbb{P}^{n+1} &= \mathbb{P}\mathbb{P}^n = \mathbb{P}\left[\frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}\right] \\
&= \frac{1}{\alpha+\beta}\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \\
&= \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^{n+1}}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}
\end{aligned}$$

Once again, the last step follows from a direct computation. $\square$

We shall now compute the $\lim_{n\to\infty} \mathbb{P}^n$. But we start with the two simple cases where (2) will not be used.

1. Suppose that $\alpha + \beta = 0$. This is equivalent to saying that $\alpha = \beta = 0$. Then

$$\mathbb{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and thus } \lim_{n\to\infty} \mathbb{P}^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The probabilistic meaning of this case is obvious: both states are absorbing and once thechain starts at i it remains in i forever.

2. Suppose that $\alpha + \beta = 2$. This is equivalent to saying that $\alpha = \beta = 1$. Then

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and thus } \mathbb{P}^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \mathbb{P}^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The sequence of matrices $\mathbb{P}^n$ does not have a limit. The probabilistic meaning of this case is also simple: the chain moves from the state it is in to another one at each transition, e. g. if $X_0 = 1$ then the trajectory of the chain is 1212121... In fact, this is a deterministic movement with the initial state being the only random state of the chain.

Exercises. (a) Suppose that $P\{X_0 = 1\} = 0.3$. What is then $P\{X_{27} = 2\}$? (b) Prove that the only equilibrium distribution of this chain is given by $\underline{\pi} = (0.5, 0.5)$.

3. Finally, if $0 < \alpha + \beta < 2$ then $|1 - \alpha - \beta| < 1$. Hence $\lim_{n\to\infty}(1-\alpha-\beta)^n = 0$ and it follows from (2) that

$$\lim_{n\to\infty} \mathbb{P}^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \equiv \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} \tag{3}$$

We see that $\lim_{n\to\infty} P\{X_n = j \,|\, X_0 = i\} = \begin{cases} \frac{\beta}{\alpha+\beta} & \text{if } j = 1 \\ \frac{\beta}{\alpha+\beta} & \text{if } j = 2 \end{cases}$ This answers the first question from section 2. Moreover, we also see that these limits do not depend on i: the chain forgets its starting point!

4 General results

Definition 4.1
A MC is irreducible if for all $i, j \in S$ there is some $k \geq 1$ with $p_{ij}^{(k)} > 0$. A MC is regular if there is some $k \geq 1$ such that $p_{ij}^{(k)} > 0$ for all $i, j \in S$.

Remarks.
1. Note that a MC is regular if and only if there is a k such that $\mathbb{P}^k > 0$. (We use here the following convention: if $B = (b_{ij})$ is a matrix then the inequality $B > 0$ means that all matrix elements are strictly positive: $b_{ij} > 0$.
2. Irreducible chains do not have absorbing states (explain this, it is very simple).
3. A regular chain is irreducible but the reverse is false. Example: if $\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ then $\mathbb{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbb{P}^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$... This chain is irreducible but not regular. However, here is a sufficient condition for regularity of irreducible chains.

Theorem 4.2
If a finite MC is irreducible and there is $i \in S$ such that $p_{ii} > 0$ then this chain is regular.

(Remember that the term "finite MC" means that the number of states, $\#(S)$, is finite.)

Definition 4.3
If $\underline{\mu^{(n)}} \to \underline{w}$ as $n \to \infty$ then the probability vector $\underline{w}$ is called the limiting distribution of the MC X_n.

Theorem 4.4
Let $\mathbb{P}$ be a transition matrix of a finite regular MC. Then
1. $\mathbb{P}^n \to W$ as $n \to \infty$, where W is a matrix with all rows being equal to the same probability vector $\underline{w} = (w_1, ..., w_m)$.
2. $\underline{w}$ is the unique solution of

$$\underline{w}\mathbb{P} = \underline{w} \quad \textit{with} \quad \sum_{i=1}^{m} w_i = 1. \tag{4}$$

Proof will not be given and you are not required to know it. However you are required to know the statement of this theorem and to be able to use it. This applies also to Theorems 4.5 and 4.6 stated below.

Corollary. $\underline{w}$ is the limiting distribution of the MC X_n. Moreover, it does not depend on $\underline{\mu_0}$. Indeed, the first statement of the Theorem means that $\lim_{n\to\infty} P\{X_n = j \,|\, X_0 = i\} = w_j$ does not depend on i. Hence

$$\lim_{n\to\infty} P\{X_n = j\} = \lim_{n\to\infty} (\sum_{i=1}^{m} \mu_i P\{X_n = j \,|\, X_0 = i\}) = \sum_{i=1}^{m} \mu_i w_j = w_j.$$

By now, you must have noticed that Theorem 4.4 provides a complete answer to the first question stated in section 2.

Remarks.
1. It is useful to rewrite equations (4) in a more explicit form:

$$\begin{cases} \sum_{i=1}^{m} w_i = 1, \\ \sum_{i=1}^{m} w_i p_{ij} = w_j, \quad 1 \le j \le m \end{cases} \tag{5}$$

2. Equations (4) imply that $\underline{w}$ is the equilibrium distribution for or MC (see Definition 1.3). It is easy to show that every limiting distribution is an equilibrium distribution. The reverse is wrong: $\underline{\mu^{(n)}}$ may not converge to a limit even if the equilibrium distribution is unique.
The answer to the second question is given by the following

Theorem 4.5
Let $\mathbb{P}$ be a transition matrix of a finite regular MC. Let n_i be the number of visits to state i by time n. Then

$$\frac{n_i}{n} \to w_i \quad \textit{as} \quad n \to \infty$$

in the following sense: for any $\varepsilon > 0$

$$P\{\,|\frac{n_i}{n} - w_i| \ge \varepsilon\,\} \to 0 \ \textit{as} \ \ n \to \infty. \tag{6}$$

This theorem is a particular case of the following Law of Large Numbers. Let $f(j)$ be a function defined on S and taking real values. Set $E(f) \stackrel{\text{def}}{=} \sum_{j=1}^{m} w_j f(j)$.

Theorem 4.6
Let $\mathbb{P}$ be a transition matrix of a finite regular MC. Let n_i be the number of visits to state i by time n. Then

$$\frac{1}{n}[f(X_0) + f(X_1) + ... + f(X_{n-1})] \to E(f) \quad \textit{as} \quad n \to \infty$$

in the following sense: for any $\varepsilon > 0$

$$P\{\,|\frac{1}{n}[f(X_0) + f(X_1) + ... + f(X_{n-1})] - E(f)| \ge \varepsilon\,\} \to 0 \ \textit{as} \ \ n \to \infty. \tag{7}$$

Remarks.
1. Note that the results stated in Theorems 4.5 and 4.6 do not depend on the choice of the initial distribution of the MC!
2. To obtain Theorem 4.5 from Theorem 4.6, fix i and set $f(j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j = i \end{cases}$ Then $f(X_0) + f(X_1) + ... + f(X_{n-1}) = n_i$ and $E(f) = w_i$ which means that (7) implies (6).

Example. Explain why the following chains, given by their transition matrices, are regular. In each case find the limiting distribution. What proportion of time do you expect the chain to spend in each state in the long run?

a)

$$\begin{pmatrix} 1/10 & 1/2 & 0 & 2/5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

b)

$$\begin{pmatrix} 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Solution to problem (a) was explained in detail in lectures. Solutions to both problems can be found on the course's web page: http://www.maths.qmul.ac.uk/∼ig/MAS338/ and go to Coursework 2007/8, CW3.