Chapter 4

Markov Chains

4.1 Introduction and Definitions

Consider a sequence of consecutive *times* (or *trials* or *stages*): $n = 0, 1, 2, ...$ Suppose that at each time a probabilistic experiment is performed, the outcome of which determines the *state* of the system at that time. For convenience, denote (or label) the states of the system by $\{0, 1, 2, ...\}$, a discrete (finite or countably infinite) *state space*: at each time these states are mutually exclusive and exhaustive.

Denote the event 'the system is in state k at time n' by $X_n = k$. For a given value of n, X_n is a random variable and has a probability distribution

$$\{\mathrm{P}(X_n = k) : k = 0, 1, 2, ...\}$$

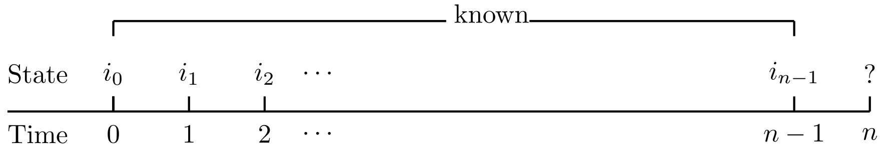
– termed the *absolute probability distribution* at time n. We refer to the sequence $\{X_n\}$ as a *discrete-time stochastic process.*

Example Consider a sequence of independent Bernoulli trials, each with the same probability of success. Let X_n be the number of successes obtained after n trials ($n = 1, 2, ...$). We might find

Trial n	1	2	3	4	5	6	7	8	...
Outcome	S	S	S	F	S	S	F	F	...
X_n	1	2	3	3	4	5	5	5	...

(Here we ignore $n = 0$ or define $X_0 = 0$). This is a particular *realization* of the stochastic process $\{X_n\}$. We know that $\mathrm{P}(X_n = k)$ is binomial.

The distinguishing features of a stochastic process $\{X_n\}$ are the relationships between $X_0, X_1, X_2, ...$ Suppose we have observed the system at times $0, 1, ..., n-1$ and we wish to make a probability statement about the state of the system at time n.



The most *general* model has the state of the system at time n being dependent on the entire past history of the system. We would then be interested in conditional probabilities

$$\mathrm{P}(X_n = i_n | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}).$$

The *simplest* model assumes that the states at different times are independent events, so that

$$P(X_n = i_n | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}) = P(X_n = i_n), \quad n = 0, 1, 2, ... : i_0 = 0, 1, 2, ..., \text{etc.}$$

The next simplest (Markov) model introduces the simplest form of dependence.

Definition A sequence of r.v.s $X_0, X_1, X_2, ...$ is said to be a (discrete) *Markov chain* if

$$\begin{aligned} P(X_n = j | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i) &= P(X_n = j | X_{n-1} = i) \\ &\quad \text{for all possible } n, i, j, i_0, ..., i_{n-2}. \end{aligned} \tag{4.1}$$

Hence in a Markov chain we do *not* require knowledge of what happened at times $0, 1, ..., (n-2)$ but only what happened at time $(n-1)$ in order to make a conditional probability statement about the state of the system at time n. We describe the system as making a *transition* from state i to state j at time n if $X_{n-1} = i$ and $X_n = j$ with (one-step) *transition probability* $P(X_n = j | X_{n-1} = i)$.

We shall only consider systems for which the transition probabilities are independent of time, so that

$$P(X_n = j | X_{n-1} = i) = p_{ij} \quad \text{for } n \geq 1 \text{ and all } i, j. \tag{4.2}$$

This is the *stationarity assumption* and $\{X_n : n = 0, 1, ...\}$ is then termed a *(time)-homogeneous* Markov chain.

The transition probabilities $\{p_{ij}\}$ form the *transition probability matrix* $\boldsymbol{P}$:

$$\boldsymbol{P} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots\cdots \\ p_{10} & p_{11} & p_{12} & \cdots\cdots \\ p_{20} & p_{21} & p_{22} & \cdots\cdots \\ \cdots & \cdots & \cdots & \cdots\cdots \\ \cdots & \cdots & \cdots & \cdots\cdots \end{pmatrix}.$$

The $\{p_{ij}\}$ have the properties

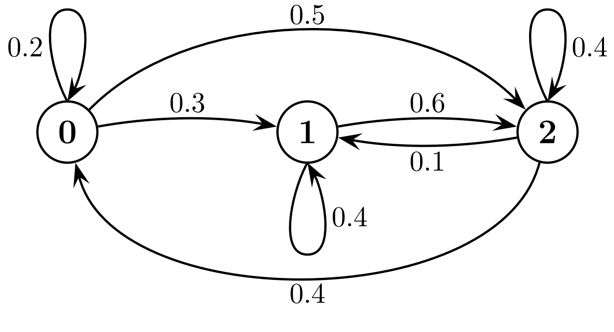
$$\begin{aligned} p_{ij} &\geq 0, \quad \text{all } i, j \\ \text{and} \quad \sum_{\text{all } j} p_{ij} &= 1, \quad \text{all } i \end{aligned} \tag{4.3}$$

(note that $i \to i$ transitions are possible). Such a matrix is termed a *stochastic matrix*.

It is often convenient to show $\boldsymbol{P}$ on a *state* (or *transition*) diagram: each vertex (or node) in the diagram corresponds to a state and each arrow to a non-zero p_{ij}. For example,

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0 & 0.4 & 0.6 \\ 0.5 & 0.1 & 0.4 \end{pmatrix}$$

is represented by



Knowledge of $\boldsymbol{P}$ and the *initial distribution* $\{\mathrm{P}(X_0 = k), k = 0, 1, 2, ...\}$ enables us, at least in theory, to calculate all probabilities of interest, e.g. *absolute* probabilities at time n,

$$\mathrm{P}(X_n = k), \qquad k = 0, 1, 2, ...$$

and *conditional* probabilities (or m-step transition probabilities)

$$\mathrm{P}(X_{m+n} = j | X_n = i), \qquad m \geq 2.$$

Notes:

(i) Some systems may be more appropriately defined for $n = 1, 2, ...$ and/or a state space which is a subset of $\{0, 1, 2, ...\}$: the above discussion is readily modified to take account of such variations.

(ii) If the system is *known* to be in state l at time 0, the initial distribution is

$$\begin{aligned} \mathrm{P}(X_0 = l) &= 1 \\ \mathrm{P}(X_0 = k) &= 0, \qquad \text{for } k \neq l. \end{aligned}$$

4.2 Some simple examples

Example 4.1 *An occupancy problem.*

Balls are distributed, one after the other, at random among 4 cells. Let X_n be the number of empty cells remaining after the nth ball has been distributed.

The *stages* are : $n = 1, 2, 3, ...$

The *state space* is: $\{0, 1, 2, 3\}$.

The possible *transitions* are:

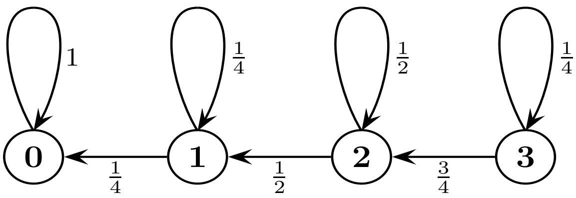
Transition X_{n-1}	$\rightarrow$	X_n	Conditional Probability
0		0	1
1		0	$\frac{1}{4}$
		1	$\frac{3}{4}$
2		1	$\frac{2}{4}$
		2	$\frac{2}{4}$
3		2	$\frac{3}{4}$
		3	$\frac{1}{4}$

Other transitions are impossible (i.e. have zero probabilities).

The probability distribution over states at stage n can be found from a knowledge of the state at stage $(n-1)$, the information from earlier stages not being required. Hence $\{X_n\}$ is a Markov chain. Furthermore, the transition probabilities are not functions of n, so the chain is homogeneous. The transition probability matrix is

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & \frac{2}{4} & \frac{2}{4} & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}$$

and the transition diagram is



♢

Example 4.2 *Random walk with absorbing barriers.*

A *random walk* is the path traced out by the motion of a particle which takes repeatedly a step of one unit in some direction, the direction being randomly chosen.

Consider a *one*-dimensional random walk on the x-axis, where there are absorbing barriers at $x = 0$ and $x = M$ (a positive integer), with

$$\begin{aligned} \text{P(particle moves one unit to the } \textit{right}) &= p \\ \text{P(particle moves one unit to the } \textit{left}) &= 1 - p = q, \end{aligned}$$

– except that if the particle is at $x = 0$ or $x = M$ it stays there.

The steps or times are: $n = 0, 1, 2, ...$

Let X_n = position of particle after step (or at time) n.

The state space is: $\{0, 1, 2, ..., M\}$.

The transition probabilities are homogeneous and given by:

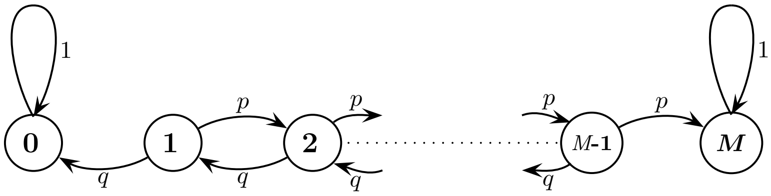
$$\begin{aligned} p_{i,i+1} &= p, \quad p_{i,i-1} = q; \quad i = 1, 2, ..., M-1 \\ p_{00} &= 1, \quad p_{M,M} = 1, \end{aligned}$$

all other p_{ij} being zero.

$\{X_n\}$ is a homogeneous Markov chain with

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & . & . & . & . & . & 0 \\ q & 0 & p & . & . & . & . & . & 0 \\ 0 & q & 0 & p & . & . & . & . & 0 \\ . & . & . & & . & . & . & . & . \\ . & . & . & & . & . & . & . & . \\ . & . & . & & . & . & . & . & . \\ 0 & . & . & & . & 0 & q & 0 & p \\ 0 & . & . & & . & 0 & 0 & 0 & 1 \end{pmatrix}$$

and the transition diagram is



◇

Example 4.3 *Discrete-time queue*

Customers arrive for service and take their place in a waiting line. During each time period a single customer is served, provided there is at least one customer waiting. During a service period new customers may arrive: the number of customers arriving in a time period is denoted by Y, with probability distribution

$$\mathrm{P}(Y = k) = a_k, \quad k = 0, 1, 2, ...$$

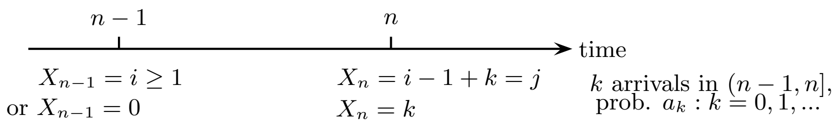
We assume that the numbers of arrivals in different periods are independent.

Here the times (each marking the end of its respective period) are: 0,1,2,...

Let X_n = number of customers waiting at time n.

The state space is: $\{0, 1, 2, ...\}$

The situation at time n can be pictured as follows:



X_n is a homogeneous Markov chain (why?) with

$$\boldsymbol{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & . & . & . & . \\ a_0 & a_1 & a_2 & a_3 & . & . & . & . \\ 0 & a_0 & a_1 & a_2 & . & . & . & . \\ 0 & 0 & a_0 & a_1 & . & . & . & . \\ . & . & . & . & & & & \\ . & . & . & . & & & & \\ . & . & . & . & & & & \end{pmatrix}.$$

◇

4.3 Calculation of probabilities

The *absolute* probabilities at time n can be calculated from those at time $(n-1)$ by invoking the law of total probability. Conditioning on the state at time $n-1$, we have:

$$\begin{aligned} \mathrm{P}(X_n = j) &= \sum_i \mathrm{P}(X_n = j | X_{n-1} = i)\mathrm{P}(X_{n-1} = i) \\ &= \sum_i p_{ij}\mathrm{P}(X_{n-1} = i) \end{aligned}$$

(since $\{X_{n-1} = i : i = 0, 1, 2, ...\}$ are mutually exclusive and exhaustive events).

In matrix notation:

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)}\boldsymbol{P}, \qquad n \geq 1 \tag{4.4}$$

where

$$\boldsymbol{p}^{(n)} = (\mathrm{P}(X_n = 0), \mathrm{P}(X_n = 1), ...)$$

(row vector). Repeated application of (4.4) gives

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(r)}\boldsymbol{P}^{n-r}, \qquad 0 \leq r \leq n-1 \tag{4.5}$$

and in particular

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^{n} \tag{4.6}$$

i.e.

$$\mathrm{P}(X_n = j) = \sum_i p_{ij}^{(n)}\mathrm{P}(X_0 = i) \tag{4.7}$$

where $p_{ij}^{(n)}$ denotes the (i, j) element of $\boldsymbol{P}^n$.

To obtain the *conditional* probabilities (or n-step transition probabilities), we condition on the initial state:

$$\mathrm{P}(X_n = j) = \sum_i \mathrm{P}(X_n = j | X_0 = i)\mathrm{P}(X_0 = i). \tag{4.8}$$

Then, comparing with (4.7), we deduce that

$$\mathrm{P}(X_n = j | X_0 = i) = p_{ij}^{(n)}. \tag{4.9}$$

Also, because the Markov chain is homogeneous,

$$\mathrm{P}(X_{r+n} = j | X_r = i) = p_{ij}^{(n)}.$$

So

$$\mathrm{P}(X_t = j | X_s = i) = p_{ij}^{(t-s)}, \quad t > s. \tag{4.10}$$

The main labour in actual calculations is the evaluation of $\boldsymbol{P}^n$. If the elements of $\boldsymbol{P}$ are numerical, $\boldsymbol{P}^n$ may be computed by a suitable numerical method when the number of states is finite: if they are algebraic, there are special methods for some forms of $\boldsymbol{P}$.

4.4 Classification of States

[*Note*: this is an extensive subject, and only a limited account is given here.]

Suppose the system is in state k at time $n = 0$. Let

$$f_{kk} = \mathrm{P}(X_n = k \text{ for some } n \geq 1 | X_0 = k), \qquad (4.11)$$

i.e. f_{kk} is the probability that the system returns *at some time* to the state k.

The state k is called *persistent* or *recurrent* if $f_{kk} = 1$, i.e. a return to k is certain. Otherwise ($f_{kk} < 1$) the state k is called *transient* (there is a positive probability that k is never re-entered).

If $p_{kk} = 1$, state k is termed an *absorbing* state.

The state k is *periodic with period* $t_k > 1$ if

$$\begin{array}{ll} p_{kk}^{(n)} > 0 & \text{when } n \text{ is a multiple of } t_k \\ \text{and} \quad p_{kk}^{(n)} = 0 & \text{otherwise.} \end{array}$$

Thus $t_k = \gcd\{n : p_{kk}^{(n)} > 0\}$: e.g. if $p_{kk}^{(n)} > 0$ only for $n = 4, 8, 12, ..$, then $t_k = 4$. State k is termed *aperiodic* if no such $t_k > 1$ exists.

State j is *accessible* (or *reachable*) from state i ($i \rightarrow j$) if $p_{ij}^{(n)} > 0$ for *some* $n > 0$. If $i \rightarrow j$ *and* $j \rightarrow i$, then states i and j are said to *communicate* ($i \leftrightarrow j$).

It can be shown that if $i \leftrightarrow j$ then states i and j

(i) are both transient or both recurrent;

(ii) have the same period.

A set C of states is called *irreducible* if $i \leftrightarrow j$ for all $i, j \in C$, so all the states in an irreducible set have the same period and are either all transient or all recurrent.

A set C of states is said to be *closed* if no state *outside* C is accessible from any state *in* C, i.e.

$$p_{ij} = 0 \quad \text{for all } i \in C, j \notin C.$$

(Thus, an absorbing state is a closed set with just one state.)

It can be shown that the entire state space can be uniquely partitioned as follows:

$$T \cup C_1 \cup C_2 \cup \cdots \qquad (4.12)$$

where T is the set of *transient* states and $C_1, C_2, ...$ are irreducible closed sets of *recurrent* states (some of the C_i may be absorbing states).

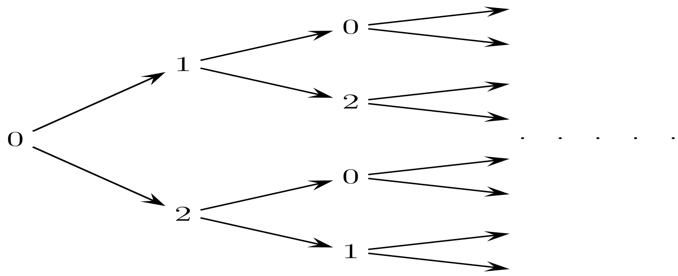
Quite often, the entire state space is irreducible, so the terms irreducible, aperiodic etc. can be applied to the Markov chain as a whole. An irreducible chain contains at most one closed set of states. In a *finite* chain, it is impossible for all states to be transient: if the chain is irreducible, the states are recurrent.

Let's consider some examples.

Example 4.4 (i) State space $S = \{0, 1, 2\}$, with

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

The possible direct transitions are: $0 \to 1, 2$; $1 \to 0, 2$; $2 \to 0, 1$. So $i \leftrightarrow j$ for all $i, j \in S$. S is the only closed set, and the Markov chain is irreducible, which in turn implies that all 3 states are recurrent.



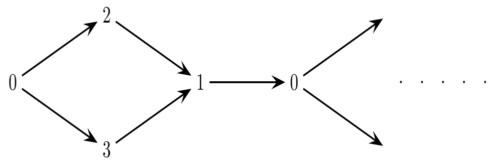
Also $p_{00} = 0, p_{00}^{(2)} > 0, p_{00}^{(3)} > 0, ..., p_{00}^{(n)} > 0$.

So state 0 is aperiodic, which implies that all states are aperiodic.

(ii) State space $S = \{0, 1, 2, 3\}$, with

$$P = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The possible direct transitions are: $0 \to 2, 3$: $1 \to 0$: $2 \to 1$: $3 \to 1$, so again $i \leftrightarrow j$ for all $i, j \in S$. S is the only closed set, the Markov chain is irreducible, and all 4 states are recurrent.



Also $p_{00} = 0, p_{00}^{(2)} = 0, p_{00}^{(3)} = 1 > 0....$

Thus state 0 is periodic with period $t_0 = 3$, and so all states have period 3.

(iii) State space $S = \{0, 1, 2, 3, 4\}$, with

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

The possible direct transitions are: $0 \to 0, 1$; $1 \to 0, 1$; $2 \to 2, 3$; $3 \to 2, 3$; $4 \to 0, 1, 4$. We conclude that $\{0, 1\}$ is a closed set, irreducible and aperiodic, and its states are recurrent; similarly for $\{2, 3\}$: state 4 is transient and aperiodic. Thus $S = T \cup C_1 \cup C_2$, where

$$T = \{4\}, \quad C_1 = \{0, 1\}, \quad C_2 = \{2, 3\}.$$

♢

4.5 The Limiting Distribution

Markov's Theorem

Consider a finite, aperiodic, irreducible Markov chain with states $0, 1, ..., M$. Then

$$p_{ij}^{(n)} \to \pi_j \quad \text{as } n \to \infty, \text{ for all } i, j; \tag{4.13a}$$

or

$$\boldsymbol{P}^n \to \begin{pmatrix} \pi_0 & \pi_1 & .. & .. & .. & \pi_M \\ \pi_0 & \pi_1 & .. & .. & .. & \pi_M \\ .. & .. & .. & .. & .. & .. \\ \pi_0 & \pi_1 & .. & .. & .. & \pi_M \end{pmatrix} \text{ as } n \to \infty. \tag{4.13b}$$

The limiting probabilities $\{\pi_j : j = 0, ..., M\}$ are the unique solution of the equations

$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}, \qquad j = 0, ..., M \tag{4.14a}$$

satisfying the normalisation condition

$$\sum_{i=0}^{M} \pi_i = 1. \tag{4.14b}$$

The equations (4.14a) may be written compactly as

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}, \tag{4.15}$$

where

$$\boldsymbol{\pi} = (\pi_0, \pi_1, ..., \pi_M).$$

Also

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^n \to \boldsymbol{\pi} \quad \text{as } n \to \infty. \tag{4.16}$$

Example 4.5

Consider the Markov chain (i) in Example 4.4 above. Since it is finite, aperiodic and irreducible, there is a unique limiting distribution $\boldsymbol{\pi}$ which satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$, i.e.

$$\begin{aligned} \pi_0 &= \pi_0 p_{00} + \pi_1 p_{10} + \pi_2 p_{20}, \quad &\text{i.e.} \quad \pi_0 &= \tfrac{1}{2}\pi_1 + \tfrac{1}{2}\pi_2 \\ \pi_1 &= \pi_0 p_{01} + \pi_1 p_{11} + \pi_2 p_{21}, \quad &\text{i.e.} \quad \pi_1 &= \tfrac{1}{2}\pi_0 + \tfrac{1}{2}\pi_2 \\ \pi_2 &= \pi_0 p_{02} + \pi_1 p_{12} + \pi_2 p_{22}, \quad &\text{i.e.} \quad \pi_2 &= \tfrac{1}{2}\pi_0 + \tfrac{1}{2}\pi_1 \end{aligned}$$

together with the normalisation condition

$$\pi_0 + \pi_1 + \pi_2 = 1$$

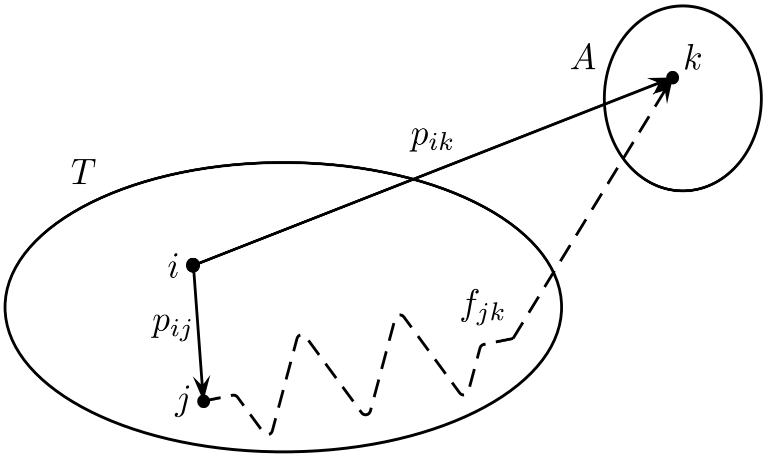
The best general approach to solving such equations is to set one $\pi_i = 1$, deduce the other values and then normalise at the end: note that one of the equations in (4.14a) is *redundant* and can be used as a check at the end. So here, set $\pi_0 = 1$: then the first two equations in (4.14a) yield $\pi_1 = \pi_2 = 1$. Since $\sum_i \pi_i = 3$, normalisation yields

$$\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}.$$

◇

4.6 Absorption in a finite Markov chain

Consider a finite Markov chain consisting of a set T of transient states and a set A of absorbing states (termed an *absorbing Markov chain*). Let A_k denote the event of absorption in state k and f_{ik} the probability of A_k when starting from the transient state i.



Conditioning on the first transition (first step analysis) as indicated on the diagram, and using the law of total probability, we have

$$\begin{aligned}
\mathrm{P}(A_k|X_0=i) &= \sum_{j\in A\cup T} \mathrm{P}(A_k|X_0=i,X_1=j)\mathrm{P}(X_1=j|X_0=i) \\
&= 1.\mathrm{P}(X_1=k|X_0=i)+\sum_{j\in T}\mathrm{P}(A_k|X_1=j)\mathrm{P}(X_1=j|X_0=i),
\end{aligned}$$

i.e.

$$f_{ik}=p_{ik}+\sum_{j\in T}p_{ij}f_{jk}. \qquad (4.17)$$

Also, let T_A denote the time to absorption (in *any* $k\in A$), and let μ_i be the mean time to absorption starting from state i. Then, again conditioning on the first transition, we obtain

$$\begin{aligned}
\mu_i=\mathrm{E}(T_A|X_0=i) &= \sum_{j\in A\cup T}\mathrm{E}(T_A|X_0=i,X_1=j)\mathrm{P}(X_1=j|X_0=i) \\
&= \sum_{j\in A}1.p_{ij}+\sum_{j\in T}\mathrm{E}(T_A|X_1=j)p_{ij} \\
&= \sum_{j\in A}p_{ij}+\sum_{j\in T}\{1+\mathrm{E}(T_A|X_0=j)\}p_{ij},
\end{aligned}$$

i.e.

$$\mu_i=1+\sum_{j\in T}p_{ij}\mu_j. \qquad (4.18)$$

(Note that in both (4.17) and (4.18) the summation $\sum_{j\in T}$ includes $j=i$.)

4.6.1 Absorbing random walk and Gambler's Ruin

Consider the random walk with absorbing barriers (at $0, M$) introduced in Example 4.2 above. The states $\{1, ..., M-1\}$ are transient. Then

$$f_{iM} = p_{iM} + \sum_{j \in T} p_{ij} f_{jM}, \qquad i = 1, ..., M-1$$

i.e.

$$\begin{aligned}
f_{1M} &= pf_{2M} \\
f_{iM} &= p_{i,i-1} f_{i-1,M} + p_{i,i+1} f_{i+1,M} = qf_{i-1,M} + pf_{i+1,M}, \qquad i = 2, ..., M-2 \\
f_{M-1,M} &= p + qf_{M-2,M}
\end{aligned}$$

For convenience, define $f_{0M} = 0, f_{MM} = 1$. Then

$$f_{iM} = qf_{i-1,M} + pf_{i+1,M}, \qquad i = 1, ..., M-1. \tag{4.19}$$

If we define

$$d_{iM} = f_{iM} - f_{i-1,M}, \quad i = 1, ..., M \tag{4.20}$$

these difference equations can be written as simple 1-step recursions:

$$pd_{i+1,M} = qd_{iM}, \quad i = 1, ..., M-1. \tag{4.21}$$

It follows that

$$d_{iM} = (q/p)^{i-1} d_{1M}, \quad i = 1, ..., M. \tag{4.22}$$

Now

$$\begin{aligned}
\sum_{j=1}^{i} d_{jM} &= (f_{1M} - f_{0M}) + (f_{2M} - f_{1M}) + \cdots + (f_{iM} - f_{i-1,M}) \\
&= f_{iM} - f_{0M} \\
&= f_{iM}.
\end{aligned}$$

so

$$\begin{aligned}
f_{iM} &= \sum_{j=1}^{i} (q/p)^{j-1} d_{1M} \\
&= \{1 + (q/p) + (q/p)^2 + \cdots + (q/p)^{i-1}\} f_{1M} \\
&= \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)} f_{1M}, & \text{if } p \neq \frac{1}{2} \\ i f_{1M}, & \text{if } p = \frac{1}{2} \end{cases}.
\end{aligned}$$

Using the fact that $f_{MM} = 1$, we deduce that

$$f_{1M} = \begin{cases} \dfrac{1-(q/p)}{1-(q/p)^M}, & \text{if } p \neq \frac{1}{2} \\ \dfrac{1}{M}, & \text{if } p = \frac{1}{2} \end{cases},$$

and so

$$f_{iM} = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^M}, & \text{if } p \neq \frac{1}{2} \\ \dfrac{i}{M}, & \text{if } p = \frac{1}{2} \end{cases}. \tag{4.23}$$

By a similar argument (or appealing to symmetry):

$$f_{i0} = \begin{cases} \dfrac{1-(p/q)^{M-i}}{1-(p/q)^M}, & \text{if } q \neq \frac{1}{2} \\ \dfrac{M-i}{M}, & \text{if } q = \frac{1}{2} \end{cases}. \tag{4.24}$$

We deduce that

$$f_{i0} + f_{iM} = 1$$

i.e. absorption at either 0 or M is certain to occur sooner or later.

Note that, as $M \to \infty$,

$$\begin{array}{lll} f_{iM} & \to & \begin{cases} 1-(q/p)^i, & \text{if } p > \frac{1}{2} \\ 0, & \text{if } p \le \frac{1}{2} \end{cases} : \\ f_{i0} & \to & \begin{cases} (q/p)^i, & \text{if } p > \frac{1}{2} \\ 1, & \text{if } p \le \frac{1}{2} \end{cases} . \end{array} \tag{4.25}$$

Similarly we have

$$\mu_i = 1 + p\mu_{i+1} + q\mu_{i-1}, \qquad 1 \le i \le M-1 \tag{4.26}$$

with $\mu_0 = \mu_M = 0$.

The solution is

$$\mu_i = \begin{cases} \dfrac{1}{p-q}\left(M\dfrac{1-(q/p)^i}{1-(q/p)^M} - i\right) & \text{if } p \ne \frac{1}{2} \\ i(M-i) & \text{if } p = \frac{1}{2} \end{cases} . \tag{4.27}$$

We have in fact solved the famous *Gambler's Ruin* problem. Two players, A and B, start with £a and £$(M-a)$ respectively (so that their total capital is £M. A coin is flipped repeatedly, giving heads with probability p and tails with probability $q = 1-p$. Each time 'heads' occurs, B gives £1 to A, otherwise A gives £1 to B. The game continues until one or other player runs out of money. After each flip the state of the system is A's current capital, and it is clear that this executes a random walk precisely as discussed above. Then

(i) $f_{aM} (f_{a0})$ is the probability that A (B) wins:

(ii) μ_a is the expected number of flips of the coin before one of the players becomes bankrupt and the game ends.

From the result (4.25) above, we see that if a gambler is playing against an infinitely rich adversary, then if $p > \frac{1}{2}$ there is a positive probability that the gambler's fortune will increase indefinitely, while if $p \le \frac{1}{2}$ the gambler is certain to go broke sooner or later.