MTH 7241 Fall 2022: Prof. C. King

Notes 4: Finite Markov chains

A. A. Markov

Andrei Andreyevich Markov (1856 - 1922) founded the modern theory of stochastic processes, and gave his name to the special class we will consider here. He was an early activist for human rights in Imperial Russia, and in 1912 he protested Leo Tolstoy's excommunication from the Russian Orthodox Church. Markov was undaunted by extensive calculations and was very good at them. In what has now become the famous first application of Markov chains, A. A. Markov studied the sequence of 20,000 letters in A. S. Pushkins poem "Eugeny Onegin", discovering that the stationary vowel probability is $p = 0.432$, that the probability of a vowel following a vowel is $p_1 = 0.128$, and that the probability of a vowel following a consonant is $p_2 = 0.663$.

Markov chains

Markov's great contribution was the systematic analysis of a class of sequences of random variables $X_1, X_2, \ldots$ which are dependent, but only in the simplest possible way. In a sense they are the nearest to independent chains. Thus for example the sequence of random steps $S_1, S_2, \ldots$ is independent, while the random walk $\{X_k = S_1 + \cdots + S_k\}$ is not independent. However they are both Markov chains.

The theory has an enormous range of applications, including:

- statistical physics
- queueing theory
- communication networks
- voice recognition
- bioinformatics
- Google's pagerank algorithm
- computer learning and inference
- economics
- gambling
- data compression

Definition of the chain

Let Ω be a finite or countably infinite sample space. A collection of Ω-valued random variables $\{X_0, X_1, X_2, \ldots\}$ is called a discrete-time Markov chain on Ω if it satisfies the *Markov condition*:

$$P(X_{n+1} = j \,|\, X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = j \,|\, X_n = i_n) \tag{1}$$

for all $n \geq 0$ and all states $j, i_0, \ldots, i_n \in \Omega$.

Regarding the index of X_n as a discrete time the Markov condition can be summarized by saying that the conditional distribution of the future state X_{n+1} conditioned on the present and past states $X_0, \ldots, X_n$ is equal to the conditional distribution of X_{n+1} conditioned on the present state X_n. In other words, the future (random) behavior of the chain only depends on where the chain sits right now, and not on how it got to its present position.

Ω = sample space

e.g.: X_n = stock price

Ω = range of possible values for stock price

$= \{0, 0.01, 0.02, 0.03, \ldots, 1.00, 1.01, \ldots\}$

e.g.: X_n = word in sentence (speech recognition).

Ω = set of possible words

$= \{\text{word}_1, \text{word}_2, \ldots, \text{word}_n\}$.

Markov chain:

$X_0 \to X_1 \to X_2 \to X_3 \to \cdots\cdots$

- sequence of Ω-valued random variables, discrete time indexed by integer subscript, related by conditional probabilities between successive variables.

We say that

$$P(X_{n+1} = j \,|\, X_n = i)$$

i and $j \in \Omega$
(possible states).

is the *transition probability* from state i to state j after n steps. We will mostly consider homogeneous chains, meaning that the transition probabilities do not depend on n. In this case for all n and $i, j \in \Omega$

$$P(X_{n+1} = j \,|\, X_n = i) = P(X_1 = j \,|\, X_0 = i) = p_{ij} \tag{2}$$

This defines the transition matrix P with entries p_{ij}. Any transition matrix P must satisfy these properties:

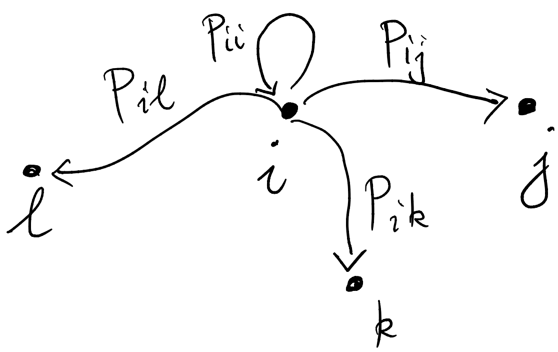
(P1) $p_{ij} \geq 0$ for all $i, j \in \Omega$

(P2) $\sum_{j \in \Omega} p_{ij} = 1$ for all $i \in \Omega$ → total probability for transitions starting at state i = 1.

Such a matrix is also called row-stochastic. So a square matrix is a transition matrix if and only if it is non-negative and row-stochastic.

An equivalent way to state the Markov condition is for any sequence of states $i, j, k, l, \ldots$,

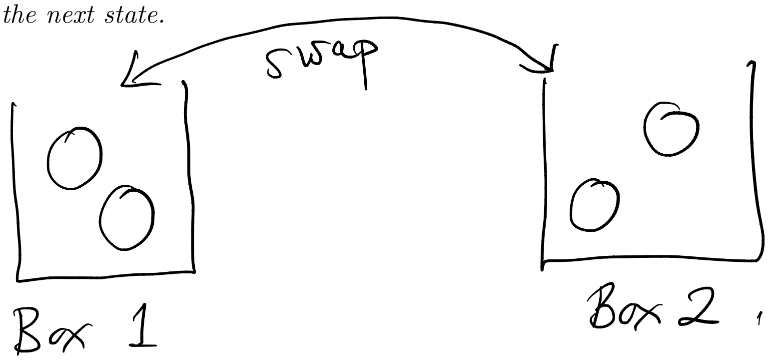
$$P(X_0 = i, \ldots, X_1 = j, X_2 = k, X_3 = l, \ldots) = P(X_0 = i)\, p_{ij}\, p_{jk}\, p_{kl}, \ldots$$



Example 1 *Consider the following model. There are four balls, two White and two Black, and two boxes. Two balls are placed in each box. The transition mechanism is that at each time unit one ball is randomly selected from each box, these balls are exchanged, and then placed back into the boxes. Let X_n be the number of White balls in the first box after n steps. The state space is $S = \{0, 1, 2\}$. The transition matrix is*

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1 & 0 \end{pmatrix} \tag{3}$$

Why is it a Markov chain? The transition mechanism only depends on the current state of the system. Once you know the current state (= number of balls in first box) you can calculate the probabilities of the next state.



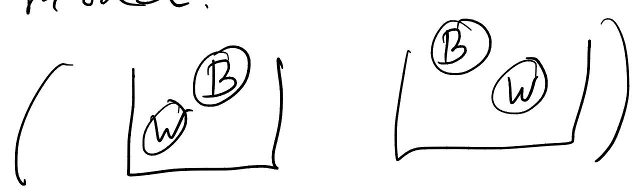
2 White
2 Black.

X_n = number of Black balls in Box 1 after n swaps.

Ran $X_n = \{0, 1, 2\} = \Omega$.

3 states for the model.

Suppose $X_0 = 1$.



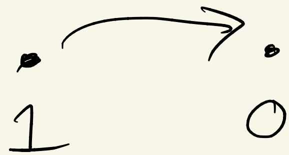
$$\mathbb{P}(X_1 = 0 \mid X_0 = 1) = \frac{1}{4}$$

independent $\begin{cases} \mathbb{P}(\text{choose B in Box 1}) = \frac{1}{2} \\ \mathbb{P}(\text{choose W in Box 2}) = \frac{1}{2} \end{cases}$

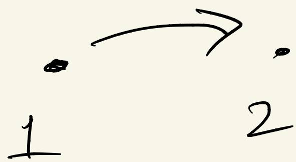
$$\mathbb{P}(X_1 = 2 \mid X_0 = 1) = \frac{1}{4}$$

$$\mathbb{P}(X_1 = 1 \mid X_0 = 1) = 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2}$$

$$\Rightarrow \quad P_{10} = \frac{1}{4}$$

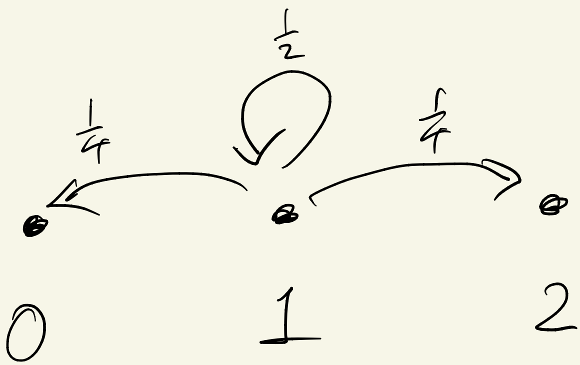


$$P_{12} = \frac{1}{4}$$



$$P_{11} = \frac{1}{2}$$

Better:



What is

$$\mathbb{P}(X_2 = 0 \mid X_1 = 1, X_0 = 1)$$

$$= \frac{1}{4}$$

$$= \mathbb{P}(X_2 = 0 \mid X_1 = 1, X_0 = 2)$$

This is the Markov condition:

a conditional probability for the future conditional on the present and the past is the same as the conditional prob. just conditional on the present.

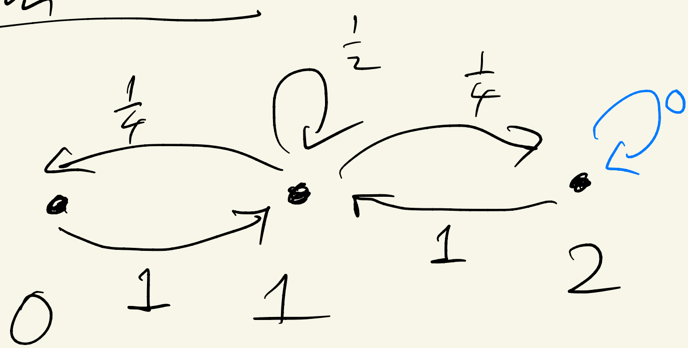
$$\mathbb{P}(X_{n+1}=j \mid X_n=i, \underbrace{X_{n-1}=k_{n-1}, X_{n-2}=k_{n-2},\dots, X_0=k_0}_{\text{past state}})$$

future state

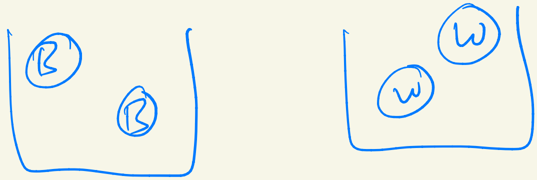
present state

$$= \mathbb{P}(X_{n+1}=j \mid X_n=i)$$

Balls in boxes,



$2 \to 2$?



NO

$$\mathbb{P}(X_0 = 1, X_1 = 0, X_2 = 1, X_3 = 2)$$

$$= \mathbb{P}(1 \to 0 \to 1 \to 2)$$

$\updownarrow$ Markov condition!

$$= \mathbb{P}(1 \to 0)\,\mathbb{P}(0 \to 1)\,\mathbb{P}(1 \to 2)$$

$$= \left(\frac{1}{4}\right)(1)\left(\frac{1}{4}\right) = \frac{1}{16}.$$

$$\mathbb{P}(X_3 = 2 \mid X_0 = 1)$$

$$\begin{tikzcd}
0 \arrow[r, bend right] & 1 \arrow[l, bend right] \arrow[loop above] \arrow[r, bend left] & 2 \arrow[l, bend left]
\end{tikzcd}$$

$$\mathbb{P}\left[\begin{array}{ccccccc} 1 & \longrightarrow & * & \longrightarrow & 1 & \longrightarrow & 2 \end{array}\right]$$

$$\begin{array}{cccc} X_0 & X_1 & X_2 & X_3 \\ & \left\{\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\right\} & & \end{array}$$

disjoint

$$= \mathbb{P}(1 \to 0 \to 1 \to 2)$$

sum

$$+ \mathbb{P}(1 \to 1 \to 1 \to 2)$$

$$+\mathbb{P}(1 \to 2 \to 1 \to 2)$$

$$= \left(\frac{1}{16}\right) + \left(\frac{1}{16}\right) + \left(\frac{1}{16}\right) = \frac{3}{16}.$$

$$\mathbb{P}(X_3 = 2 \mid X_0 = 1) = \frac{3}{16}.$$

<u>Systematic</u>

P_{ij} = transition probability $i \to j$

where $i, j \in \Omega$ are possible states.

Assume Ω is finite, $|\Omega| = m$.

p_{ij} $\quad i = 1, \ldots, m$

$\quad j = 1, \ldots, m.$

Transition matrix

$$P = (p_{ij})$$

$m \times m$ matrix

(i,j) entry of P is

$$(P)_{ij} = p_{ij}$$

Balls in boxes:

$$\Omega = \{0, 1, 2\}$$

$$P = \begin{array}{ccc|c} & & & \text{index } i \\ \left(\begin{array}{ccc} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{array}\right. & & \left.\right) & \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \\ 0 & 1 & 2 & \end{array}$$

$$\text{index } j \rightarrow \quad 0 \quad 1 \quad 2$$

$$= \begin{array}{c} \text{start state} \\ \left(\begin{array}{ccc} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \end{array}\right) \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \end{array}$$

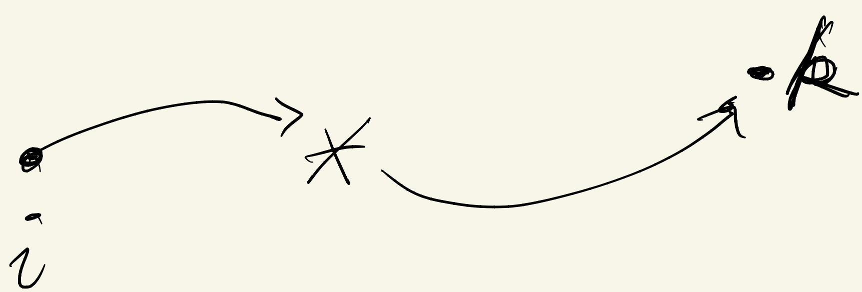
$$\text{destination state} \rightarrow \quad 0 \quad 1 \quad 2$$

Properties of Transition Matrix

- row sums must equal 1.
- $0 \le P_{ij} \le 1$ every entry.
- square matrix!
- size is $m \times m$, where m is number of states.

Ex.

$$\mathbb{P}(X_2 = k \mid X_0 = i)$$



Total probability (condition on X_1)

Markov!

$$= \sum_j \mathbb{P}(X_2 = k \mid X_1 = j, \cancel{X_0 = i})\, \mathbb{P}(X_1 = j \mid X_0 = i)$$

[Aside:

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\,\mathbb{P}(B^c)$$

$$\mathbb{P}(A \mid C) = \mathbb{P}(A \mid B, C)\,\mathbb{P}(B \mid C) + \mathbb{P}(A \mid B^c, C)\,\mathbb{P}(B^c \mid C)$$]

$$= \sum_j \mathbb{P}(X_2 = k \mid X_1 = j)\, \mathbb{P}(X_1 = j \mid X_0 = i)$$

$$= \sum_j P_{jk}\, P_{ij}$$

$$= \sum_j P_{ij}\, P_{jk}$$

$$= \sum_j (P)_{ij}\, (P)_{jk}$$

$$= (P \cdot P)_{ik}$$ matrix multiplication

$$= (P^2)_{ik}.$$

$$\boxed{\mathbb{P}(X_2 = k \mid X_0 = i) = (P^2)_{ik}}$$

Note: if P is a transition matrix, then also P^2 is a transition matrix.

So P^2 is the transition matrix for the 2-step chain

$$X_0 \to X_2 \to X_4 \to X_6 \to X_8 \to \cdots$$

or

$$X_1 \to X_3 \to X_5 \to X_7 \to \cdots$$

Balls in Boxes.

in

$$P = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cc} \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \\ \begin{array}{ccc} 0 & 1 & 2 \end{array} & \end{array}$$

out.

$\frac{1}{4} = P_{12}$

$= \mathbb{P}(X_1 = 2 | X_0 = 1)$

$$P^2 = P \cdot P$$

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

$$= \begin{array}{cc} \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} & \begin{array}{c} 0 \ \leftarrow \text{in} \\ 1 \\ 2 \end{array} \\ \text{out} \rightarrow \quad 0 \qquad 1 \qquad 2 & \end{array}$$

$$\frac{1}{2} = (P^2)_{21}$$

$$= \mathbb{P}(X_2 = 1 \mid X_0 = 2)$$

In a similar way, the matrix P^k is the transition for the k-step chain: (k is any integer).

$$\mathbb{P}(X_k = j \mid X_0 = i) = (P^k)_{ij}$$

<u>Ex:</u> balls in boxes:

$$\mathbb{P}(X_4 = 2 \mid X_0 = 0)$$

$$0 \to * \to * \to * \to 2.$$

$$= (P^4)_{02}$$

in

$$= \begin{array}{cc}
\begin{pmatrix} \frac{3}{16} & \frac{5}{8} & \frac{3}{16} \\[1ex] \frac{5}{32} & \frac{11}{16} & \frac{5}{32} \\[1ex] \frac{3}{16} & \frac{5}{8} & \frac{3}{16} \end{pmatrix} & \begin{array}{c} 0 \\[2ex] 1 \\[2ex] 2 \end{array} \\
\text{out} \to \quad 0 \qquad 1 \qquad 2 &
\end{array}$$

Notation:

n steps of chain

$$\mathbb{P}(X_n = j \mid X_0 = i)$$

$$= (P^n)_{ij}$$

$$= P_{ij}(n)$$ ← shorthand.

$$i \to * \to * \to * \cdots * \longrightarrow j$$

$\longleftarrow$ n steps $\longrightarrow$

Once the initial probability distribution of X_0 is specified, the joint distribution of the future X_i is determined by the transition matrix. So let $\alpha_i = P(X_0 = i)$ for all $i \in \Omega$, then for any sequence of states $i_0, i_1, \ldots, i_m$ we have

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_m = i_m) = \alpha_{i_0} p_{i_0,i_1} p_{i_1,i_2} \ldots p_{i_{m-1},i_m} \tag{4}$$

The transition matrix contains the information about how the chain evolves over successive transitions. For example,

$$\begin{aligned} P(X_2 = j|X_0 = i) &= \sum_k P(X_2 = j, X_1 = k|X_0 = i) \\ &= \sum_k P(X_2 = j|X_1 = k, X_0 = i)\, P(X_1 = k|X_0 = i) \\ &= \sum_k P(X_2 = j|X_1 = k)\, P(X_1 = k|X_0 = i) \\ &= \sum_k p_{kj}\, p_{ik} \\ &= \sum_k (P)_{ik}\, (P)_{kj} \\ &= (P^2)_{ij} \end{aligned} \tag{5}$$

So the matrix P^2 provides the transition rule for two consecutive steps of the chain. It is easy to check that P^2 is also row-stochastic, and hence is the transition matrix for a Markov chain, namely the two-step chain $X_0, X_2, X_4, \ldots,$ or $X_1, X_3, \ldots$. A similar calculation shows that for any $n \geq 1$

$$P(X_n = j|X_0 = i) = (P^n)_{ij} \tag{6}$$

and hence P^n is the n-step transition matrix. We write

$$p_{ij}(n) = (P^n)_{ij} = P(X_n = j|X_0 = i) \tag{7}$$

Note that $p_{ij} = 0$ means that the chain cannot move from state i to state j in one step. However it is possible in this situation that there is an integer n such that $p_{ij}(n) > 0$, meaning that it is possible to move from i to j in n steps. In this case we say that state j is *accessible* from state i.

$P_{ij} = 0 \Rightarrow \mathbb{P}(X_1 = j \mid X_0 = i) = 0$

If there is n such that $P_{ij}(n) > 0$

we say that state j is accessible from i.

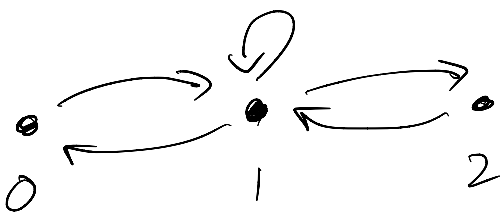
$i \to \times \to \times \cdots \times \to j$

Example 2 *For the balls in boxes example, all states are accessible from all other states.*

$P_{02} = 0$ but $P_{02}(2) > 0$.

$$\mathbb{P}(0 \to 1 \to 2) = (1)\left(\tfrac{1}{4}\right) = \tfrac{1}{4} > 0.$$

So state 2 is accessible from state 0.



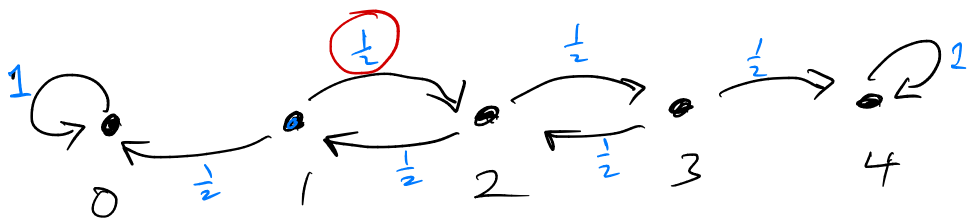
All states are accessible from all others.

Example 3 *The drunkard's walk. The state space is* $\Omega = \{0, 1, 2, 3, 4\}$*, and* X_n *is the drunkard's position after* n *steps. At each step he goes left or right with probability* $1/2$ *until he reaches an endpoint* 0 *or* 4*, where he stays forever. The transition matrix is*

$$P = \begin{array}{cc} & \text{in} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \\ \text{out} \to \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 & \end{array} \tag{8}$$

Again the transition mechanism depends only on the current state, which means that this is a Markov chain. Are all states accessible from all other states?

A finite state Markov chain can be usefully represented by a directed graph where the vertices are the states, and edges are the allowed one-step transitions.



States 0 and 4 are called <u>absorbing states</u>.

States $\{1, 2, 3\}$ are not accessible from state 0 or 4.

$$\mathbb{P}(X_3 = 1 \mid X_0 = 2)$$

$$= \mathbb{P}(2 \to * \to * \to 1)$$

$$= \mathbb{P}(2 \to 1 \to 2 \to 1) + \mathbb{P}(2 \to 3 \to 2 \to 1)$$

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

$$\mathbb{P}(X_3 = 1 \mid X_0 = 2) = \left(P^3\right)_{21}$$

Classification of states "f" = first.

Define

$$f_{ij}(n) = P(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j \,|\, X_0 = i) \tag{9}$$

to be the probability that starting in state i the chain first visits state j after n steps. Define

$$\begin{aligned} f_{ij} &= \sum_{n=1}^{\infty} f_{ij}(n) \\ &= P(X_n = j \text{ for some } n \geq 1 \,|\, X_0 = i) \end{aligned}$$

This is the probability that the chain eventually visits state j starting in state i.

Definition 1 *The state j is* persistent *if $f_{jj} = 1$. The state j is* transient *if $f_{jj} < 1$.*

So the state j is persistent if the chain must eventually return to state j, given that it started at state j. Since the clock starts over at each return to state j, we conclude that if the state j is persistent then the chain must return to j infinitely often.

Similarly if j is transient, then the chain will return to state j only finitely often. Eventually it will never return to the state.

One particular type of persistent state is the class of *absorbing states*: these are the states for which $p_{ii} = 1$. So once the chain enters an absorbing state it can never leave. The end states of the drunkard's walk are absorbing states.