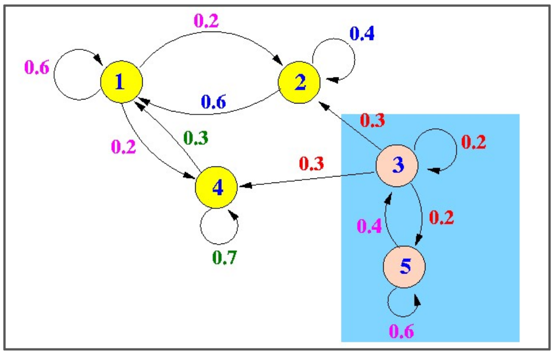
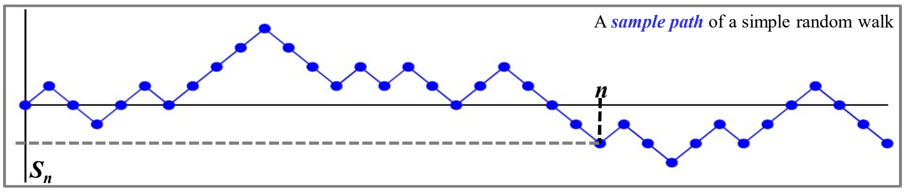
Chapter 6.
Transient and Recurrent States



$X_0, X_1, X_2, \ldots$ are ***i.i.d.*** random variables with $P(X_i = +1) = P(X_i = -1) = 1/2$.

$$S_n = X_0 + X_1 + \cdots + X_n$$

$\{S_n, n \geq 1\} = \{S_1, S_2, \ldots\}$ **: simple random walk** (Markov chain)



Definition. The discrete random variables $X_1, X_2, \ldots$ on $\mathbb{Z}^d$ are called steps of the random walk and have the following probability distribution: $\forall i \in \mathbb{N} : P(X_i = e) = \frac{1}{2d}$ if $e \in \mathbb{Z}^d$ and $\|e\| = 1$, and $P(X_i = e) = 0$ otherwise.

Definition. $S_0 = 0 \in \mathbb{Z}^d$ and $S_n = X_1 + \ldots + X_n$ for $n \in \mathbb{N}$ is called the position of the random walk at time n.

Polýa's Theorem. Simple random walks of dimension $d = 1, 2$ are recurrent, and of $d \geq 3$ are transient.

Transient and Recurrent States

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$ and transition function P. Set

Then ρ_{xy} denotes the probability that a Markov chain starting at x will be in state y at some positive time. In particular, ρ_{yy} denotes the probability that a Markov chain starting at y will ever return to y. A state y is called *recurrent* if $\rho_{yy} = 1$ and *transient* if $\rho_{yy} < 1$. If y is a recurrent state, a Markov chain starting at y returns to y with probability one. If y is a transient state, a Markov chain starting at y has positive probability $1 - \rho_{yy}$ of never returning to y. If y is an absorbing state, then $P_y(T_y = 1) = P(y, y) = 1$ and hence $\rho_{yy} = 1$; thus an absorbing state is necessarily recurrent.

Let $1_y(z)$, $z \in \mathscr{S}$, denote the indicator function of the set $\{y\}$ defined by

$$1_y(z) = \begin{cases} 1, & z = y, \\ 0, & z \neq y. \end{cases}$$

Let $N(y)$ denote the number of times $n \geq 1$ that the chain is in state y. Since $1_y(X_n) = 1$ if the chain is in state y at time n and $1_y(X_n) = 0$ otherwise, we see that

$$N(y) = \sum_{n=1}^{\infty} 1_y(X_n). \tag{31}$$

The event $\{N(y) \geq 1\}$ is the same as the event $\{T_y < \infty\}$. Thus

Let m and n be positive integers. By (27), the probability that a Markov chain starting at x first visits y at time m and next visits y n units of time later is $P_x(T_y = m)P_y(T_y = n)$. Thus

$$\begin{aligned} P_x(N(y) \geq 2) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P_x(T_y = m)P_y(T_y = n) \\ &= \left(\sum_{m=1}^{\infty} P_x(T_y = m)\right)\left(\sum_{n=1}^{\infty} P_y(T_y = n)\right) \\ &= \rho_{xy}\rho_{yy}. \end{aligned}$$

Similarly we conclude that

(32)

Since

$$P_x(N(y) = m) = P_x(N(y) \geq m) - P_x(N(y) \geq m + 1),$$

it follows from (32) that

(33)

Also

$$P_x(N(y) = 0) = 1 - P_x(N(y) \geq 1),$$

so that

(34) $$\boxed{P_x(N(y) = 0) = 1 - \rho_{xy}.}$$

These formulas are intuitively obvious. To see why (33) should be true, for example, observe that a chain starting at x visits state y exactly m times if and only if it visits y for a first time, returns to y $m - 1$ additional times, and then never again returns to y.

We use the notation $E_x(\ )$ to denote expectations of random variables defined in terms of a Markov chain starting at x. For example,

(35) $$E_x(1_y(X_n)) = P_x(X_n = y) = P^n(x, y).$$

It follows from (31) and (35) that

$$E_x(N(y)) = E_x\left(\sum_{n=1}^{\infty} 1_y(X_n)\right)$$
$$= \sum_{n=1}^{\infty} E_x(1_y(X_n))$$
$$= \sum_{n=1}^{\infty} P^n(x, y).$$

Set

$$\boxed{G(x, y) = E_x(N(y)) = \sum_{n=1}^{\infty} P^n(x, y).}$$

Then $G(x, y)$ denotes the expected number of visits to y for a Markov chain starting at x.

Theorem 1 (i) *Let y be a transient state. Then*

$$P_x(N(y) < \infty) = 1$$

and

(36) $$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}}, \qquad x \in \mathscr{S},$$

$$\boxed{G(x, y) = E_x(N(y)) = \sum_{n=1}^{\infty} P^n(x, y)}$$

which is finite for all $x \in \mathscr{S}$.

(ii) *Let y be a recurrent state. Then $P_y(N(y) = \infty) = 1$ and $G(y, y) = \infty$. Also*

(37) $$P_x(N(y) = \infty) = P_x(T_y < \infty) = \rho_{xy}, \qquad x \in \mathscr{S}.$$

If $\rho_{xy} = 0$, then $G(x, y) = 0$, while if $\rho_{xy} > 0$, then $G(x, y) = \infty$.

◆ This theorem describes the fundamental difference between a ***transient state*** and a ***recurrent state***.

◆ If y is a ***transient*** state, then no matter where the Markov chain starts,

- it makes only a finite number of visits to y and
- the expected number of visits to y is finite.

◆ Suppose instead that y is ***recurrent*** state.

- Then if the Markov chain starts at y, it returns to y infinitely often.
- If the chain starts at some other state x, it may be impossible for it to ever hit y.
- If it is possible, however, and the chain does visit y at least once, then it does so infinitely often.

Proof. Let y be a transient state. Since $0 \le \rho_{yy} < 1$, it follows from (32) that

$$P_x(N(y) = \infty) = \boxed{}$$

By (33)

$$\begin{aligned} G(x, y) &= E_x(N(y)) \\ &= \sum_{m=1}^{\infty} m P_x(N(y) = m) \\ &= \sum_{m=1}^{\infty} m \rho_{xy} \rho_{yy}^{m-1}(1 - \rho_{yy}). \end{aligned}$$

Substituting $t = \rho_{yy}$ in the power series

$$\boxed{} \qquad \sum_{m=0}^{\infty} t^m = \frac{1}{1-t}.$$

we conclude that

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty.$$

This completes the proof of (i).

Now let y be recurrent. Then $\rho_{yy} = 1$ and it follows from (32) that

$$P_x(N(y) = \infty) = \lim_{m\to\infty} P_x(N(y) \geq m) = \lim_{m\to\infty} \rho_{xy}\rho_{yy}^{m-1}$$

$$= \lim_{m\to\infty} \rho_{xy} = \rho_{xy}.$$

In particular, $P_y(N(y) = \infty) = 1$. If a nonnegative random variable has positive probability of being infinite, its expectation is infinite. Thus

$$G(y, y) = E_y(N(y)) = \infty.$$

If $\rho_{xy} = 0$, then $P_x(T_y = m) = 0$ for all finite positive integers m, so (28) implies that $P^n(x, y) = 0$, $n \geq 1$; thus $G(x, y) = 0$ in this case. If $\rho_{xy} > 0$, then $P_x(N(y) = \infty) = \rho_{xy} > 0$ and hence

$$G(x, y) = E_x(N(y)) = \infty.$$

This completes the proof of Theorem 1. ∎

Let y be a transient state. Since

$$\sum_{n=1}^{\infty} P^n(x, y) = G(x, y) < \infty, \qquad x \in \mathscr{S},$$

we see that

$$\lim_{n\to\infty} P^n(x, y) = 0, \qquad x \in \mathscr{S}. \tag{38}$$

A Markov chain is called a *transient chain* if all of its states are transient and a *recurrent chain* if all of its states are recurrent. It is easy to see that a Markov chain having a finite state space must have at least one recurrent state and hence cannot possibly be a transient chain. For if $\mathscr{S}$ is finite and all states are transient, then by (38)

Decomposition of the state space

$x \to y$

Let x and y be two not necessarily distinct states. We say that *x leads to y* if $\rho_{xy} > 0$. It is left as an exercise for the reader to show that x leads to y if and only if $P^n(x, y) > 0$ for some positive integer n. It is also left to the reader to show that if x leads to y and y leads to z, then x leads to z.

Theorem 2 *Let x be a recurrent state and suppose that x leads to y. Then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.*

Proof. We assume that $y \neq x$, for otherwise there is nothing to prove. Since

$$P_x(T_y < \infty) = \rho_{xy} > 0,$$

we see that $P_x(T_y = n) > 0$ for some positive integer n. Let n_0 be the least such positive integer, i.e., set

$$n_0 = \min\{\, n \geq 1 : P_x(T_y = n) > 0 \,\}. \tag{39}$$

It follows easily from (39) and (28) that $P^{n_0}(x, y) > 0$ and

$$P^m(x, y) = 0, \qquad 1 \leq m < n_0. \tag{40}$$

Since $P^{n_0}(x, y) > 0$, we can find states $y_1, \ldots, y_{n_0-1}$ such that

$$P_x(X_1 = y_1, \ldots, X_{n_0-1} = y_{n_0-1}, X_{n_0} = y) = P(x, y_1) \cdots P(y_{n_0-1}, y) > 0.$$

None of the states $y_1, \ldots, y_{n_0-1}$ equals x or y; for if one of them did equal x or y, it would be possible to go from x to y with positive probability in fewer than n_0 steps, in contradiction to (40).

We will now show that $\rho_{yx} = 1$. Suppose on the contrary that $\rho_{yx} < 1$. Then a Markov chain starting at y has positive probability $1 - \rho_{yx}$ of never hitting x. More to the point, a Markov chain starting at x has the positive probability

of visiting the states $y_1, \ldots, y_{n_0-1}, y$ successively in the first n_0 times and never returning to x after time n_0. But if this happens, the Markov chain never returns to x at any time $n \geq 1$, so we have contradicted the assumption that x is a recurrent state.

Since $\rho_{yx} = 1$, there is a positive integer n_1 such that $P^{n_1}(y, x) > 0$. Now

$$\begin{aligned} P^{n_1+n+n_0}(y, y) &= P_y(X_{n_1+n+n_0} = y) \\ &\geq P_y(X_{n_1} = x, X_{n_1+n} = x, X_{n_1+n+n_0} = y) \\ &= P^{n_1}(y, x)P^n(x, x)P^{n_0}(x, y). \end{aligned}$$

Hence

$$G(y, y) \geq \sum_{n=n_1+1+n_0}^{\infty} P^n(y, y) = \sum_{n=1}^{\infty} P^{n_1+n+n_0}(y, y)$$

$$\geq P^{n_1}(y, x)P^{n_0}(x, y) \sum_{n=1}^{\infty} P^n(x, x) = P^{n_1}(y, x)P^{n_0}(x, y)G(x, x) = +\infty,$$

from which it follows that y is also a recurrent state.

Since y is recurrent and y leads to x, we see from the part of the theorem that has already been verified that $\rho_{xy} = 1$. This completes the proof. ∎

A nonempty set C of states is said to be *closed* if no state inside of C leads to any state outside of C, i.e., if

(41)

Equivalently C is closed if and only if

(42) $$P^n(x, y) = 0, \quad x \in C, y \notin C, \text{ and } n \geq 1.$$

Actually, even from the weaker condition

(43) $$P(x, y) = 0, \quad x \in C \text{ and } y \notin C,$$

we can prove that C is closed. For if (43) holds, then for $x \in C$ and $y \notin C$

$$\begin{aligned} P^2(x, y) &= \sum_{z \in \mathscr{S}} P(x, z)P(z, y) \\ &= \sum_{z \in C} P(x, z)P(z, y) = 0, \end{aligned}$$

and (42) follows by induction. If C is closed, then a Markov chain starting in C will, with probability one, stay in C for all time. If a is an absorbing state, then $\{a\}$ is closed.

A closed set C is called *irreducible* if x leads to y for all choices of x and y in C. It follows from Theorem 2 that if C is an irreducible closed set, then either every state in C is recurrent or every state in C is transient. The next result is an immediate consequence of Theorems 1 and 2.

Corollary 1 *Let C be an irreducible closed set of recurrent states. Then $\rho_{xy} = 1$, $P_x(N(y) = \infty) = 1$, and $G(x, y) = \infty$ for all choices of x and y in C.*

An *irreducible Markov chain* is a chain whose state space is irreducible, that is, a chain in which every state leads back to itself and also to every other state. Such a Markov chain is necessarily either a transient chain or a recurrent chain. Corollary 1 implies, in particular, that an irreducible recurrent Markov chain visits every state infinitely often with probability one.

We saw at page 11 that if $\mathscr{S}$ is finite, it contains at least one recurrent state. The same argument shows that any finite closed set of states contains at least one recurrent state. Now let C be a finite irreducible closed set. We have seen that either every state in C is transient or every state in C is recurrent, and that C has at least one recurrent state. It follows that every state in C is recurrent. We summarize this result:

Theorem 3 *Let C be a finite irreducible closed set of states. Then every state in C is recurrent.*

Consider a Markov chain having a finite number of states. Theorem 3 implies that if the chain is irreducible it must be recurrent. If the chain is not irreducible, we can use Theorems 2 and 3 to determine which states are recurrent and which are transient.

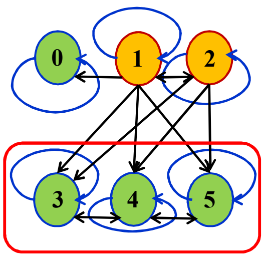
Example. Consider a Markov chain having the transition matrix

$$
\begin{array}{c}
\\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5
\end{array}
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\
0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\
0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4}
\end{bmatrix}
\end{array}.
$$

Determine which states are recurrent and which states are transient.

As a first step in studying this Markov chain, we determine by inspection which states lead to which other states. This can be indicated in matrix form as

$$
\begin{array}{c}
\\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5
\end{array}
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\
0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\
0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4}
\end{bmatrix}
\end{array}
$$



The x, y element of this matrix is $+$ or 0 according as ρ_{xy} is positive or zero, i.e., according as x does or does not lead to y. Of course, if $P(x, y) > 0$, then $\rho_{xy} > 0$. The converse is certainly not true in general. For example, $P(2, 0) = 0$; but

$$P^2(2, 0) = P(2, 1)P(1, 0) = \tfrac{1}{5} \cdot \tfrac{1}{4} = \tfrac{1}{20} > 0,$$

so that $\rho_{20} > 0$.

State 0 is an absorbing state, and hence also a recurrent state. We see clearly from the matrix of +'s and 0's that {3, 4, 5} is an irreducible closed set. Theorem 3 now implies that 3, 4, and 5 are recurrent states. States 1 and 2 both lead to 0, but neither can be reached from 0. We see from Theorem 2 that 1 and 2 must both be transient states. In summary, states 1 and 2 are transient, and states 0, 3, 4, and 5 are recurrent.

Let $\mathscr{S}_T$ denote the collection of transient states in $\mathscr{S}$, and let $\mathscr{S}_R$ denote the collection of recurrent states in $\mathscr{S}$. In Example $\mathscr{S}_T = \{1, 2\}$ and $\mathscr{S}_R = \{0, 3, 4, 5\}$. The set $\mathscr{S}_R$ can be decomposed into the disjoint irreducible closed sets $C_1 = \{0\}$ and $C_2 = \{3, 4, 5\}$. The next theorem shows that such a decomposition is always possible whenever $\mathscr{S}_R$ is nonempty.

Theorem 4 *Suppose that the set $\mathscr{S}_R$ of recurrent states is nonempty. Then $\mathscr{S}_R$ is the union of a finite or countably infinite number of disjoint irreducible closed sets $C_1, C_2, \ldots$.*

Proof. Choose $x \in \mathscr{S}_R$ and let C be the set of all states y in $\mathscr{S}_R$ such that x leads to y. Since x is recurrent, $\rho_{xx} = 1$ and hence $x \in C$. We will now verify that C is an irreducible closed set. Suppose that y is in C and y leads to z. Since y is recurrent, it follows from Theorem 2 that z is recurrent. Since x leads to y and y leads to z, we conclude that x leads to z. Thus z is in C. This shows that C is closed. Suppose that y and z are both in C. Since x is recurrent and x leads to y, it follows from Theorem 2 that y leads to x. Since y leads to x and x leads to z, we conclude that y leads to z. This shows that C is irreducible.

To complete the proof of the theorem, we need only show that if C and D are two irreducible closed subsets of $\mathscr{S}_R$, they are either disjoint or identical. Suppose they are not disjoint and let x be in both C and D. Choose y in C. Now x leads to y, since x is in C and C is irreducible. Since D is closed, x is in D, and x leads to y, we conclude that y is in D. Thus every state in C is also in D. Similarly every state in D is also in C, so that C and D are identical. ∎