Finite Markov Chains and Markov Decision Processes

Tomoyuki Shirai

Abstract Markov chains are important tools used for stochastic modeling in various areas of mathematical sciences. The first section of this article presents a survey of the basic notions of discrete-time Markov chains on finite state spaces together with several illustrative examples. Markov decision processes (MDPs), which are also known as stochastic dynamic programming or discrete-time stochastic control, are useful for decision making under uncertainty. The second section will provide a simple formulation of MDPs with finite state spaces and actions, and give two important algorithms for solving MDPs, value iteration and policy iteration, with an example on iPod shuffle.

Keywords Markov chain · Markov decision process · Mixing time · Coupling · Cutoff phenomenon

1 Markov Chains

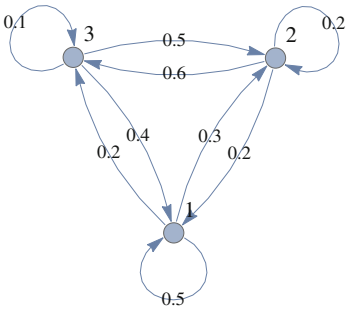
Throughout this article, we assume that S is a finite set of states and we denote the set $\{0, 1, 2, \dots\}$ by $\mathbf{T}$.

A discrete-time stochastic process on S is a sequence of S-valued random variables $\{X_t\}_{t\in\mathbf{T}}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. A *Markov chain* on S is a stochastic process having the following Markov property: for $0 \le t_0 < t_1 < \cdots < t_n < t$ and $x_0, x_1, \dots, x_n, x \in S$,

$$\mathbb{P}(X_t = x | X_{t_0} = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n) = \mathbb{P}(X_t = x | X_{t_n} = x_n).$$

T. Shirai (✉)
Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
e-mail: shirai@imi.kyushu-u.ac.jp

R. Nishii et al. (eds.), *A Mathematical Approach to Research Problems of Science and Technology*, Mathematics for Industry 5,
DOI: 10.1007/978-4-431-55060-0_15, © Springer Japan 2014



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

Fig. 1 State transition diagram and corresponding transition matrix

In particular, a Markov chain $X = \{X_t\}_{t\in\mathbf{T}}$ is said to be *time homogeneous* if $\mathbb{P}(X_{t+1} = y|X_t = x), x, y \in S$ does not depend on t. When a Markov chain X is time homogeneous, the $|S| \times |S|$ matrix $P = (p(x, y))_{x,y\in S}$ given by the one-step transition probability $p(x, y) := \mathbb{P}(X_{t+1} = y|X_t = x)$ is called a *transition matrix*. A time homogeneous Markov chain is completely determined by a transition matrix.

Lemma 1 *Let $X = \{X_t\}_{t\in\mathbf{T}}$ be a time homogeneous Markov chain. Then, for every $s, t \in \mathbf{T}$ and $x, y \in S$, $\mathbb{P}(X_{t+s} = y|X_s = x) = P^t(x, y)$.*

Throughout this section, we treat only time homogeneous Markov chains.

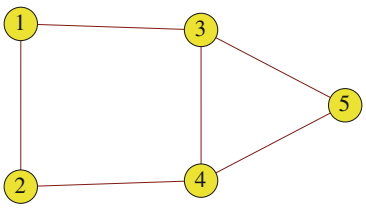
1.1 Examples of Markov Chains

Example 1 Let $S = \{1, 2, \ldots, n\}$. An n by n matrix $P = (p_{ij})_{i,j=1}^n$ is said to be a *stochastic matrix* if $p_{ij} \geq 0$ for all $i, j = 1, 2, \ldots, n$ and $\sum_{j=1}^n p_{ij} = 1$ for all $i = 1, 2, \ldots, n$. Every stochastic matrix P defines a Markov chain. If n is small, it is well described by using a diagram (Fig. 1).

Example 2 (Simple random walk (SRW) on a finite graph) Let $G = (V, E)$ be a finite connected graph and set $S = V$ with $|V| \geq 2$. An SRW on a finite graph G is a Markov chain on the vertex set S with the transition probability being $\deg(x)^{-1}$ at each vertex $x \in S$, where $\deg(x)$ is the degree of a vertex x in G. For example, in Fig. 2, $\deg(1) = \deg(2) = \deg(5) = 2$, and $\deg(3) = \deg(4) = 3$.

Example 3 (Ehrenfest's urn) In two urns, say U_1 and U_2, there are n balls in total. A ball is taken out uniformly at random and put into the other urn. Looking at the number of balls in U_1, we can regard it as a Markov chain on $S = \{0, 1, 2, \ldots, n\}$ with transition probability

$$p(k, k-1) = \frac{k}{n}, \quad p(k, k+1) = \frac{n-k}{n} \quad (k = 0, 1, 2, \ldots, n).$$



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

Fig. 2 Finite graph and transition matrix of SRW on it

Example 4 (SRW on a hypercube) Let $S = \{0, 1\}^n$. We can identify S with the vertices of a square when $n = 2$ and those of a cube when $n = 3$. Since the size of the transition matrix is 2^n, it is not practical to write it down. In this case, it is more convenient to give a transition rule algorithmically. The transition rule from a point $x = (x_1, \ldots, x_n) \in S$ is defined as follows:

1. Choose a coordinate i from $\{1, 2, \ldots, n\}$ uniformly at random.
2. Update x_i to be 1 if $x_i = 0$ and 0 if $x_i = 1$. That is, $x_i \mapsto 1 - x_i$.

This rule defines the SRW $X = \{X_t\}_{t \in \mathbf{T}}$ on S. For example, when $n = 5$, transition proceeds like

$$(1, 0, 1, 1, 0) \xrightarrow{3} (1, 0, 0, 1, 0) \xrightarrow{5} (1, 0, 0, 1, 1) \xrightarrow{2} (1, 1, 0, 1, 1) \xrightarrow{3} (1, 1, 1, 1, 1) \xrightarrow{1} \cdots$$

The number above each arrow indicates the coordinate chosen in step 1. If we use up-spin and down-spin instead of 1 and 0, we see that

$$(\uparrow, \downarrow, \uparrow, \uparrow, \downarrow) \xrightarrow{3} (\uparrow, \downarrow, \downarrow, \uparrow, \downarrow) \xrightarrow{5} (\uparrow, \downarrow, \downarrow, \uparrow, \uparrow) \xrightarrow{2} (\uparrow, \uparrow, \downarrow, \uparrow, \uparrow) \xrightarrow{3} (\uparrow, \uparrow, \uparrow, \uparrow, \uparrow) \xrightarrow{1} \cdots$$

It seems like a transition for the stochastic Ising model (a model for magnetism). One can easily see that $N_t = \sum_{i=1}^{n} (X_t)_i$, the number of 1's in X_t, is the same Markov chain as was given in Example 3.

Example 5 (Markov chain on the set of q-colorings) Let $G = (V, E)$ be a finite connected graph. For a fixed integer $q > \max_{x \in V} \deg(x)$, we consider a map $c : V \to \{1, 2, \ldots, q\}$. It can be regarded as a coloring of V by q-colors. We call a map c a q-coloring and denote the set of all q-colorings by S. If c satisfies $c(v) \neq c(w)$ whenever $vw \in E$, i.e., v and w are adjacent in G, we call it a proper q-coloring and denote the totality of proper q-colorings by S_{proper}. Even when it is difficult to identify the structure of S for a general graph G, we can define a natural Markov chain $\{c_t\}_{t \in \mathbf{T}}$ on S algorithmically:

1. A vertex in V is chosen uniformly at random.
2. If $v \in V$ is chosen at step 1, we set $A_v(c_t) = \{1, 2, \ldots, q\} \setminus \{c_t(w) : vw \in E\}$, which is the set of colors admissible for the vertex v. A color is chosen from $A_v(c_t)$ uniformly at random and $c_{t+1}(v)$ is updated to that color, leaving all the other vertices unchanged (Fig. 3).

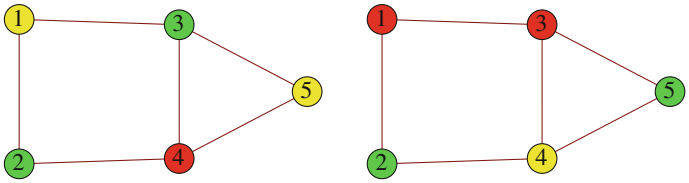


Fig. 3 *Left diagram* shows a proper 3-coloring, and *right diagram* shows an improper one

1.2 Irreducibility and Periodicity

It is important to know whether or not the Markov chain under consideration can traverse its state space.

Definition 1 We say that a Markov chain $X = \{X_t\}_{t \in \mathbf{T}}$ on S is *irreducible* if for any $x, y \in S$ there exists $t = t_{x,y} \in \mathbb{N}$ such that $\mathbb{P}(X_t = y | X_0 = x) > 0$.

Definition 2 Let $\mathrm{Per}(x) := \{t \in \mathbb{N} : \mathbb{P}(X_t = x \,|X_0 = x) > 0\}$. We call the greatest common divisor of $\mathrm{Per}(x)$ the period of a state $x \in S$. It is known that the period is constant on S when X is irreducible. In this case, the period can be considered as that of Markov chain X. If the period is 1, X is said to be *aperiodic*.

Example 6 (Random bishop/knight moves) The possible moves for a bishop and a knight from a particular square on a chessboard are shown in Fig. 4. The state space S comprises the 64 squares. The square to move to is chosen uniformly at random from the possible moves. In the example shown, the bishop chooses one of the squares with probability $1/13$ and moves to it, and the knight does the same with probability $1/8$. These transition rules define Markov chains on S. We call these chains "random bishop move" and "random knight move," respectively.

- (Irreducibility). The random bishop move is not irreducible. Indeed, by the transition rule, the bishop can only move on the squares of the same color as that of the initial place. Then, it is impossible for the bishop to jump to any square of the other color. By induction on the size of the chessboard, it can be shown that the random knight move is irreducible.
- (Periodicity). The period of the random knight move is two. Indeed, the random knight can move only to a square of the opposite color so that an even number of moves is required to return to the initial square. On the other hand, the random bishop move is aperiodic since it is clear that $\{2, 3\} \subset \mathrm{Per}(x)$ for every $x \in S$.

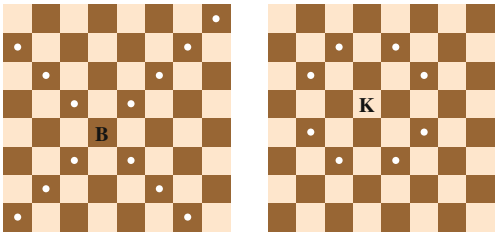


Fig. 4 Bishop (B) and knight (K) can move to a *square* with a *white dot*

1.3 Stationarity and Reversibility

It is important to study the behavior of a Markov chain $X = \{X_t\}_{t\in\mathbf{T}}$ as $t \to \infty$. By the Markov property, the distribution of X_t converges to a stationary distribution (under mild conditions) as $t \to \infty$ regardless of its initial distribution.

Definition 3 We say that π is a *stationary distribution* of a Markov chain X on S if it is a probability distribution and satisfies

$$\sum_{x\in S} \pi(x)p(x, y) = \pi(y), \ \forall y \in S.$$

We say that a Markov chain X or its transition matrix P is *reversible* with respect to π if the detailed balance condition

$$\pi(x)p(x, y) = \pi(y)p(y, x), \ \forall x, y \in S$$

holds. We call π a *reversible distribution* or a *reversible probability measure*.

It is easy to see the following.

Proposition 1 *If π is a reversible distribution, then it is also a stationary distribution.*

Remark 1 Suppose that P is irreducible. There exists a reversible distribution if and only if for any closed path $(x_1, x_2, \ldots, x_n, x_1)$, it holds that

$$p(x_1, x_2)p(x_2, x_3)\cdots p(x_n, x_1) = p(x_1, x_n)p(x_n, x_{n-1})\cdots p(x_2, x_1). \quad (1)$$

For fixed $a \in S$, we define $\tilde{\pi}(x) = \frac{p(a,x_1)p(x_1,x_2)\cdots p(x_n,x)}{p(x_1,a)p(x_2,x_1)\cdots p(x,x_n)}$ by taking a path $(a, x_1, \ldots, x_n, x)$. It does not depend on the choice of a path joining a and x under the condition (1), and it is a constant multiple of the reversible distribution.

Example 7 It is easy to show that the Markov chain defined in Example 3 is reversible with respect to $\pi(k) = \binom{n}{k}2^{-n}$. Indeed, the detailed balance condition $\tilde{\pi}(k)\frac{n-k}{n} =$

$\tilde{\pi}(k+1)\frac{k+1}{n}$, $k = 0, 1, \ldots, n-1$ with $\tilde{\pi}(0) = 1$ yields $\tilde{\pi}(k) = \binom{n}{k}$. Therefore, we obtain the reversible distribution $\pi(k) = \tilde{\pi}(k)/\sum_{j=0}^{n} \tilde{\pi}(j)$.

Example 8 Let C_n be the cycle graph with n vertices. The SRW on C_n is irreducible and reversible with respect to the uniform distribution. If n is odd, the SRW is aperiodic; if n is even, the SRW has period 2. A Markov chain on C_n moving to the right with probability $p(\neq 1/2)$ and to the left with probability $1-p(\neq 1/2)$ has the uniform distribution as the stationary distribution; however, it is not reversible since the condition (1) in Remark 1 fails.

The following two propositions are useful for identifying reversible distributions:

Proposition 2 *The SRW on a finite graph $G = (V, E)$ in Example 2 has the reversible distribution $\pi(x) = \frac{\deg(x)}{2|E|}$, where $2|E| = \sum_{x\in V} \deg(x)$ by the handshaking lemma.*

Proposition 3 *Suppose that the transition probability of an irreducible Markov chain on S is symmetric in the sense that $p(x, y) = p(y, x)$ for every $x, y \in S$. Then, the uniform distribution $\pi(x) = \frac{1}{|S|}$, $\forall x \in S$, is the reversible distribution.*

The next theorem is one of the most important facts in Markov chain theory.

Theorem 1 *Let $X = \{X_t\}_{t\in\mathbf{T}}$ be an irreducible Markov chain on a finite state space S.*

(1) There exists a unique stationary distribution π.
(2) If X is aperiodic, then the distribution $\mathbb{P}(X_t = \cdot | X_0 = x) = P^t(x, \cdot)$ of X_t starting at x converges to the stationary distribution π as $t \to \infty$ for any $x \in S$. In other words, P^t converges to the matrix Π whose row vectors are all $\Pi(x, \cdot) = \pi$ $(x \in S)$.
(3) For each $x \in S$, $\pi(x) = \frac{1}{\mathbb{E}_x[\tau_x^+]}$, where $\tau_x^+ = \inf\{t \geq 1 : X_t = x\}$.

Example 9 The Markov chain given in Example 2 has the stationary distribution $\pi = (\frac{1}{6}, \frac{1}{6}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6})$ from Proposition 2. Since P is irreducible and aperiodic, (2) of Theorem 1 implies

$$P^t = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}^t \to \begin{pmatrix} 1/6 & 1/6 & 1/4 & 1/4 & 1/6 \\ 1/6 & 1/6 & 1/4 & 1/4 & 1/6 \\ 1/6 & 1/6 & 1/4 & 1/4 & 1/6 \\ 1/6 & 1/6 & 1/4 & 1/4 & 1/6 \\ 1/6 & 1/6 & 1/4 & 1/4 & 1/6 \end{pmatrix} = \Pi \quad (t \to \infty)$$

By (3) of Theorem 1, we have $\mathbb{E}_x[\tau_x^+] = 6$ for $x = 1, 2, 5$ and $\mathbb{E}_x[\tau_x^+] = 4$ for $x = 3, 4$.

Example 10 For a random knight move starting from one of the corners on the chessboard, say c, it is easy to show that $\mathbb{E}_c[\tau_c^+] = 168$ by Proposition 2 and (3) of Theorem 1. Indeed, it is easy to check that $\deg(c) = 2$ and that

$$2|E| = \sum_{x \in S} \deg(x) = 2 \times 4 + 3 \times 8 + 4 \times 20 + 6 \times 16 + 8 \times 16 = 336.$$

Remark 2 We note that an irreducible Markov chain on S is aperiodic if there exists a state $x \in S$ such that $p(x, x) > 0$. To apply (2) of Theorem 1, we define the lazy version of a Markov chain with $P = (p(x, y))$ as the Markov chain with transition matrix $Q = (q(x, y))$ with

$$q(x, y) = \begin{cases} \frac{1}{2} p(x, y) & \text{if } y \neq x, \\ \frac{1}{2} + \frac{1}{2} p(x, x) & \text{if } y = x. \end{cases}$$

It is clear that $Q = \frac{1}{2}(I + P)$. If a fair coin is flipped and it comes up heads, then the Markov chain moves according to the original probability law P; if it comes up tails, then it stays at the present position. The stationary distribution of Q is the same as that of P. Even if P is periodic, Q becomes aperiodic.

Example 11 Let $G = (V, E)$ be a finite connected graph and suppose that $q > \max_{x \in S} \deg(x)$. In Example 5, a Markov chain was defined on the set of all q-colorings. By the transition rule, a vertex chosen in step 1 is colored differently from the vertices in its neighborhood. Through repeated transitions, at least after all the vertices are chosen in step 1, the state becomes a q-proper coloring even if it was originally a non-q-proper coloring. Moreover, once the state becomes q-proper, it will remain q-proper. This means that S_{proper} is closed with respect to this Markov chain. Although the Markov chain on S is not irreducible, that on S_{proper} is irreducible. Such a subset of a state space as S_{proper} is sometimes called an irreducible component. By Proposition 3, the stationary distribution is the uniform distribution on S_{proper}.

1.4 Coupon Collector's Problem

Coupon collector's problem is a classic problem in probability theory and has been extended in several ways. Here we consider the most basic one.

Problem 1 Suppose that there are n different kinds of coupons. One coupon is obtained with equal probability $\frac{1}{n}$ in each trial. How many trials does it take to collect a complete set of coupons?

The number of different coupons is considered to be a Markov chain $X = \{X_t\}_{t \in \mathbf{T}}$ on $S = \{0, 1, 2, \ldots, n\}$ with $X_0 = 0$. Since the probability of getting a new kind of coupon is $\frac{n-k}{n}$ if one has k different kinds already, the transition probability is given by

$$p(k, k) = \frac{k}{n}, \quad p(k, k+1) = \frac{n-k}{n} \quad (k \in S).$$

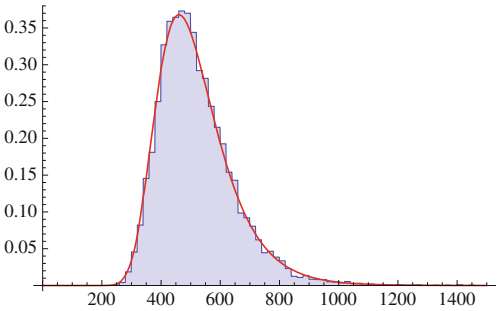


Fig. 5 Histogram of τ_{100} (simulation) and the limiting distribution $e^{-e^{-c}}$

By definition, this Markov chain only goes upwards. Let τ_n be a random variable taking values in $\mathbb{N}$ defined by

$$\tau_n = \inf\{t \in \mathbb{N} : X_t = n\},$$

which is the first time that a complete set of coupons has been collected; if the set $\{t \in \mathbb{N} : X_t = n\}$ is empty, τ_n is understood to be ∞. Problem 1 can thus be rephrased as the problem of studying the random variable τ_n.

Proposition 4 *(1)* $E[\tau_n] = n\sum_{k=1}^{n} \frac{1}{k} \sim n \log n$.[1] *(2)* $\lim_{n\to\infty} P(\tau_n \leq n \log n + cn) = e^{-e^{-c}}$ $(c \in \mathbb{R})$.

This proposition implies that the expected time to collect a complete set of coupons is about $n \log n$ and the probability that all kinds are not yet collected after $n \log n$ is exponentially small. For example, if $n = 100$, then $E[\tau_{100}] = 518.738\ldots$ (Fig. 5).

1.5 Mixing Time

The distribution at time t of an irreducible and aperiodic Markov chain on a finite state space S converges to the stationary distribution as $t \to \infty$ by Theorem 1. Here we consider the speed of convergence. For that we introduce a distance on $\mathscr{P}(S)$, the set of all probability measures on S.

Definition 4 For $\mu, \nu \in \mathscr{P}(S)$, we define the *total variation distance* by

$$\|\mu - \nu\|_{TV} = \max_{A \subset S} |\mu(A) - \nu(A)|.$$

This distance has several different expressions.

[1] $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

Proposition 5 *For* $\mu, \nu \in \mathscr{P}(S)$, $0 \le \|\mu - \nu\|_{TV} \le 1$ *and*

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{x\in S}|\mu(x) - \nu(x)| = \sum_{\substack{x\in S \\ \mu(x)\ge\nu(x)}} |\mu(x) - \nu(x)|$$
$$= \inf\{P(X \ne Y) : \ (X, Y) \textit{ is a coupling of}(\mu, \nu)\},$$

where a two-dimensional random variable (X, Y) *is said to be a coupling of* (μ, ν) *if the marginal distributions of* X *and* Y *are equal to* μ *and* ν*, respectively. Here we simply write* $\mu(x)$ *for* $\mu(\{x\})$.

Remark 3 When a Markov chain is irreducible and aperiodic, since S is finite, Theorem 1 implies that $d(t) := \max_{x\in S} \|P^t(x, \cdot) - \pi\|_{TV} \to 0$. Moreover, it is known that $d(t)$ is monotone decreasing.

Definition 5 From the remark above, we can define the *mixing time* by

$$t_{\text{mix}}(\varepsilon) := \inf\{t \in \mathbb{N} : \ d(t) \le \varepsilon\}$$

for given $\varepsilon \in (0, 1/2)$. In particular, we write $t_{\text{mix}} := t_{\text{mix}}(1/4)$. Here $1/4$ can be replaced with any $\varepsilon \in (0, 1/2)$.

Mixing time is the time when a Markov chain approaches the stationarity "sufficiently." Several results have been obtained for the following problem.

Problem 2 Given an increasing sequence of state spaces $\{S_n : n \in \mathbb{N}\}$ and Markov chains $X^{(n)} = \{X_t^{(n)}\}_{t\in\mathbf{T}}$ on S_n, one can define $t_{\text{mix}}^{(n)}$ for each $X^{(n)}$ on S_n. Analyze the asymptotic behavior of the mixing time $t_{\text{mix}}^{(n)}$ as $n \to \infty$.

1.6 Coupling of Markov Chains

The coupling method is often used for comparisons with probability distributions. In the example below, we use a coupling of Markov chains to derive an inequality.

Definition 6 (1) Let $X = \{X_t\}_{t\in\mathbf{T}}$ and $Y = \{Y_t\}_{t\in\mathbf{T}}$ be Markov chains on S starting at different initial states x and y, respectively. A Markov chain $\{(\tilde{X}_t, \tilde{Y}_t)\}_{t\in\mathbf{T}}$ on $S \times S$ is said to be a Markov *coupling* of X and Y if the probability law of $\{\tilde{X}_t\}_{t\in\mathbf{T}}$ (resp. $\{\tilde{Y}_t\}_{t\in\mathbf{T}}$) is equal to that of the given Markov chain X (resp. Y). We denote the probability law of this coupling $\{(\tilde{X}_t, \tilde{Y}_t)\}_{t\in\mathbf{T}}$ by $\mathbb{P}_{x,y}$.

(2) We define a coupling time by $\tau_{couple} = \inf\{t \ge 0 : \ \tilde{X}_t = \tilde{Y}_t\}$.

Example 12 Consider a Markov chain on $S = \{0, 1, 2, \ldots, n\}$. This chain jumps to one of its two neighbors with equal probability $1/2$ at $\{1, 2, \ldots, n-1\}$, to 0 or 1 with equal probability at 0, and to $n-1$ or n with equal probability at n. We construct

a coupling as follows: Toss a fair coin. Both $\tilde{X}_t$ and $\tilde{Y}_t$ move upwards if it comes up heads and both move downwards if it comes up tails. The important feature of this coupling is the fact that if $x \leq y$, then $\tilde{X}_t \leq \tilde{Y}_t$ for any $t \geq 0$. Therefore, since $\{\tilde{X}_t = n\} \subset \{\tilde{Y}_t = n\}$, we can see that if $x \leq y$ then

$$P^t(x, n) = \mathbb{P}_{x,y}(\tilde{X}_t = n) \leq \mathbb{P}_{x,y}(\tilde{Y}_t = n) = P^t(y, n).$$

In other words, $P^t(x, n)$ is an increasing function of x for each t. This fact is not so easy to prove by simply using matrix computations.

1.7 Upper Estimate of Mixing Time via Coupling of Markov Chains

The expected coupling time is used as an upper bound of t_{mix}.

Proposition 6 *Let $\mathbb{P}_{x,y}$ be a coupling of two Markov chains starting at x and y. Then, $t_{\mathrm{mix}} \leq 4 \max_{x,y \in S} \mathbb{E}_{x,y}[\tau_{couple}]$.*

From Proposition 6, it is important to construct a "nice" coupling with small coupling time. Here we give two examples.

1.7.1 Mixing Time of LSRW on Cycle Graph C_n

First we estimate the mixing time for the lazy version of the SRW on C_n given in Example 8. We construct a coupling $\{(\tilde{X}_t, \tilde{Y}_t)\}_{t \in \mathbf{T}}$ of two LSRWs starting at x and y respectively as follows:

1. Toss a fair coin. If it comes up heads, $\tilde{X}_t$ moves according to the transition rule; if it comes up tails, $\tilde{Y}_t$ does.
2. After the two chains meet, they move together as a single LSRW, keeping $\tilde{X}_t = \tilde{Y}_t$ for $t \geq \tau_{\mathrm{couple}}$.

Looking at either $\tilde{X}_t$ or $\tilde{Y}_t$ reveals that each chain is obviously an LSRW on C_n. Let us consider the coupling time of this chain $\{(\tilde{X}_t, \tilde{Y}_t)\}_{t \in \mathbf{T}}$. Let Z_t be the shortest path distance between $\tilde{X}_t$ and $\tilde{Y}_t$. It is thus a Markov chain on $\{0, 1, \ldots, \lfloor n/2 \rfloor\}$. (The transition rule at $\lfloor n/2 \rfloor$ is a little different depending on whether n is even or odd.) Then, the coupling time of $\tilde{X}_t$ and $\tilde{Y}_t$ is equal to the first hitting time of Z_t at 0. It is known to be of $O(n^2)$. Therefore, by Proposition 6, we can conclude that $t_{\mathrm{mix}}^{(n)} = O(n^2)$.

1.7.2 Mixing Time of LSRW on Hypercube

Consider the lazy version of the SRW on hypercube $S = \{0, 1\}^n$ given in Example 4. A coupling $\{(\tilde{X}_t, \tilde{Y}_t)\}_{t\in\mathbf{T}}$ of two LSRWs starting at different initial states is constructed as follows:

1. A coordinate i is chosen from $\{1, 2, \ldots, n\}$ uniformly at random.
2. In accordance with the heads or tails of a fair coin flip, set $\tilde{X}_t(i) = \tilde{Y}_t(i) = 1$ or $\tilde{X}_t(i) = \tilde{Y}_t(i) = 0$.

For example, a transition when $n = 5$ proceeds like

$$\begin{pmatrix}1&0&1&1&0\\0&0&0&0&0\end{pmatrix} \overset{3,\text{heads}}{\to} \begin{pmatrix}1&0&1&1&0\\0&0&1&0&0\end{pmatrix} \overset{5,\text{heads}}{\to} \begin{pmatrix}1&0&1&1&1\\0&0&1&0&1\end{pmatrix} \overset{3,\text{tails}}{\to} \begin{pmatrix}1&0&0&1&1\\0&0&0&0&1\end{pmatrix} \overset{1,\text{tails}}{\to} \cdots$$

Suppose i is chosen at step 1. No matter what the value of the i-th coordinate is, at step 2, it keeps that value with probability $1/2$ and is updated to the other value with probability $1/2$. Therefore, if we look at either $\tilde{X}_t$ or $\tilde{Y}_t$ only, we see nothing but an LSRW. Under this coupling, once the i-th coordinate is chosen at step 1, the values of the i-th coordinate of $\tilde{X}_t$ and $\tilde{Y}_t$ will remain the same. Therefore, the coupling time of the coupled chain is the first time when all the coordinates at which the values are different at $t = 0$ (e.g., $\{1, 3, 4\}$ in the example above) are chosen. If we regard $\{1, 3, 4\}$ as the coupons yet to be collected, the coupling time is smaller than τ_n defined in coupon collector's problem in Sect. 1.4. Therefore, $\mathbb{E}_{x,y}[\tau_{\text{couple}}] \le \mathbb{E}[\tau_n] \le n \log n + n$. By Proposition 6, we see that $t_{\text{mix}}^{(n)} \le 4(n \log n + n)$. It is known that $t_{\text{mix}}^{(n)} \sim \frac{1}{2} n \log n$.

1.8 Cutoff Phenomenon

The cutoff phenomenon is said to occur when the total variation distance $d(t)$ keeps nearly 1 before the mixing time t_{mix} and abruptly drops to near 0 around the mixing time t_{mix}. This implies that the distribution of X_t is far from the stationarity before time t_{mix} and close to stationarity after time t_{mix}. This phenomenon is formulated as follows.

Definition 7 A sequence of Markov chains has a *cutoff* if

$$t_{\text{mix}}^{(n)}(\varepsilon) \sim t_{\text{mix}}^{(n)}(1-\varepsilon) \quad \text{for every } \varepsilon \in (0, 1/2)$$

as $n \to \infty$, which is equivalent to

$$\lim_{n\to\infty} d_n(c t_{\text{mix}}^{(n)}) = \begin{cases} 1 & \text{if } c < 1, \\ 0 & \text{if } c > 1. \end{cases}$$

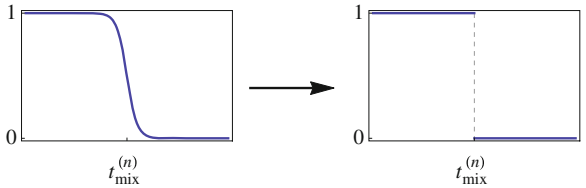


Fig. 6 Cutoff phenomenon. Graph of $d_n(t)$ rescaled by $t_{\mathrm{mix}}^{(n)}$ as $n \to \infty$

The total variation distance $d_n(t)$ converges to a step function as $n \to \infty$ by rescaling time t by $t_{\mathrm{mix}}^{(n)}$ (Fig. 6).

The following is a more precise version of the above.

Definition 8 A sequence of Markov chains has a *cutoff with a window of size* w_n if $w_n = o(t_{\mathrm{mix}}^{(n)})$ and for every $\varepsilon \in (0, 1/2)$ there exists $c_\varepsilon > 0$ such that

$$t_{\mathrm{mix}}^{(n)}(\varepsilon) - t_{\mathrm{mix}}^{(n)}(1-\varepsilon) \le c_\varepsilon w_n \quad (\forall n \in \mathbb{N}),$$

which is equivalent to

$$\lim_{c \to -\infty} \liminf_{n \to \infty} d_n(t_{\mathrm{mix}}^{(n)} + c w_n) = 1, \quad \lim_{c \to +\infty} \limsup_{n \to \infty} d_n(t_{\mathrm{mix}}^{(n)} + c w_n) = 0.$$

Example 13 (1) The LSRW on the hypercube $\{0, 1\}^n$ has a cutoff at $t_{\mathrm{mix}}^{(n)} \sim \frac{1}{2} n \log n$ with a window of size n.

(2) The SRW on cycle graph C_n does not have a cutoff.

(3) A biased random walk on $\{0, 1, \ldots, n\}$ moves upwards with probability $p > 1/2$ and downwards with probability $1 - p$. Then its lazy version has a cutoff around $t_{\mathrm{mix}}^{(n)} \sim (p - 1/2)^{-1} n$ with a window of size $\sqrt{n}$.

2 Markov Decision Processes

On many occasions one has to make a decision to minimize a cost or maximize a reward. Markov decision processes (MDPs) provide a model for use in such situations.

Here we give a formulation of MDPs. Let S be a finite state space and A a finite set of actions. For each $a \in A$ a transition matrix $P(a) = (p_{xy}(a))_{x,y \in S}$ is given. A function $c : S \times A \to [0, \infty)$ is called a *cost function*. A *policy* is a sequence $u = \{u_t\}_{t \in \mathbf{T}}$ of functions $u_t : S^{t+1} \to A$. For given $\{P(a)\}_{a \in A}$, we define a stochastic process $\{X_t\}_{t \in \mathbf{T}}$ on S associated with policy u and initial state $\mu = (\mu_x)_{x \in S}$ by the following properties: for $x_0, x_1, \ldots, x_{t+1} \in S$,

1. $\mathbb{P}^u(X_0 = x_0) = \mu_{x_0}$.
2. $\mathbb{P}^u(X_{t+1} = x_{t+1} | X_0 = x_0, \ldots, X_t = x_t) = p_{x_t x_{t+1}}(u_t(x_0, \ldots, x_t))$.

When the initial state μ is the delta measure at x, we denote the probability law of $\{X_t\}_{t\in\mathbf{T}}$ by $\mathbb{P}^u_x$. This process is not, in general, a Markov chain since the conditional probability depends on the past not just on the present state. A policy u is said to be a *stationary policy* if there exists a map $u : S \to A$ such that $u_t(x_0, \ldots, x_t) = u(x_t)$ for any $t = 0, 1, \ldots$. Here we abuse the notation of u. If a policy u is stationary, then the corresponding stochastic process is a Markov chain.

In what follows, for simplicity, we assume the following:

(A1) There exists an absorbing state $z \in S$ in the sense that $p_{zy}(a) = \delta_{zy}$ and $c(z, a) = 0$ for any $a \in A$. We denote the set of all absorbing states by S_{abs}.
(A2) For $x \in S \setminus S_{abs}$, $c(x, a) > 0$ for every $a \in A$.
(A3) There exists a stationary policy u such that for every $x \in S \setminus S_{abs}$ there exists $t = t_x \in \mathbb{N}$ so that $\mathbb{P}^u_x(X_t \in S_{abs}) > 0$.

Let τ be the first hitting time to S_{abs}, i.e., $\tau = \inf\{t \geq 0 : X_t \in S_{abs}\}$. We define the expected total cost associated with a policy u by

$$V^u(x) = \mathbb{E}^u_x\left[\sum_{t=0}^{\tau-1} c(X_t, u_t(X_0, X_1, \ldots, X_t))\right] \quad (x \in S)$$

and the optimal total cost by

$$V^*(x) = \inf_u V^u(x) \quad (x \in S).$$

It is clear that

$$c_{\min}\mathbb{E}^u_x[\tau] \leq V^u(x) \leq c_{\max}\mathbb{E}^u_x[\tau], \tag{2}$$

where $c_{\min} = \min_{x\in S\setminus S_{abs}, a\in A} c(x, a)$ and $c_{\max} = \max_{x\in S, a\in A} c(x, a)$. This implies, under (A2), that $\max_{x\in S} V^u(x) < \infty$ is equivalent to $\max_{x\in S} \mathbb{E}^u_x[\tau] < \infty$.

Lemma 2 *Let u be a stationary policy as in (A3). Then,* $\max_{x\in S} \mathbb{E}^u_x[\tau] < \infty$. *In particular,* $\max_{x\in S} V^*(x) < \infty$.

We note that if a policy u is a stationary policy associated with $u : S \to A$, then $V^u(x)$ satisfies

$$V^u(x) = c(x, u(x)) + \sum_{y\in S} p_{xy}(u(x))V^u(y). \tag{3}$$

Example 14 For $n \geq 2$, let $S = \{0, 1, 2, \ldots, n\}$ be a state space with 0 being the absorbing state. There are two actions $A = \{a_1, a_2\}$. If one chooses a_1, then one goes downward by 1 in every state; if one chooses a_2, then one jumps to 0 or $n - 1$ with equal probability $1/2$ at n and goes downward by 1 otherwise. Suppose that the cost of action a_1 (resp. a_2) is 1 (resp. C), i.e., $c(x, a_1) = 1$ (resp. $c(x, a_2) = C$)

for $x = 1, 2, \ldots, n$. Suppose $C > 1$ for simplicity. It is clear that $V^*(x) = x$ for $x \in \{0, 1, \ldots, n-1\}$ and

$$V^*(n) = \begin{cases} C + \frac{n-1}{2} & \text{if } 1 < C \leq \frac{n+1}{2}, \\ n & \text{if } C \geq \frac{n+1}{2} \end{cases}$$

for $x = n$. One should choose action a_2 at n for the former and action a_1 for the latter.

In Example 14, we can compute the optimal cost V^* explicitly. However, it is not easy to determine the optimal cost in general. So the question is how to estimate the optimal cost V^*. Here we give upper and lower estimates for V^*.

2.1 Lower Bound: Value Iteration

For a lower bound, we define the minimum expected cost incurred before time t inductively by

$$V_t(x) = \min_{a \in A} \left\{ c(x, a) + \sum_{y \in S} p_{xy}(a) V_{t-1}(y) \right\}, \quad V_0(x) = 0 \ (\forall x \in S), \tag{4}$$

which is often called the *Bellman equation with finite horizon*. By induction, it is easy to see that $V_t(x)$ is increasing in t. Hence, there exists an increasing limit $\lim_{t \to \infty} V_t(x) \in [0, \infty]$. We can show that the limit is equal to the optimal value $V^*(x)$.

Proposition 7 *For each $x \in S$, $V_t(x)$ is increasing in t and converges to $V^*(x)$ as $t \to \infty$. In particular, $V_t(x) \leq V^*(x)$ for any t.*

We apply Proposition 7 to Example 14. Since $V_t(0) \equiv 0$ and $C > 1$, we see that

$$V_t(x) = \begin{cases} 1 + V_{t-1}(x-1) & \text{for } x = 1, 2, \ldots, n-1, \\ \min\{1 + V_{t-1}(n-1),\ C + \frac{1}{2} V_{t-1}(n-1)\} & \text{for } x = n. \end{cases}$$

This implies that $V_t(x) = \min\{x, t\}$ and hence $V^*(x) = x$ for $x = 1, 2, \ldots, n-1$. When $t \geq n$, as $V_{t-1}(n-1) = n-1$, we have

$$V^*(n) = V_t(n) = \begin{cases} C + \frac{n-1}{2} & \text{if } 1 < C \leq \frac{n+1}{2}, \\ n & \text{if } C \geq \frac{n+1}{2}. \end{cases}$$

2.2 Upper Bound: Policy Iteration

Next we consider an upper bound for V^*. For a given stationary policy u_0 such that $\max_{x\in S} V^{u_0}(x) < \infty$, one can choose a stationary policy u_1 such that for each $x \in S$ action $a = u_1(x)$ minimizes the function $a \mapsto c(x,a) + \sum_{y\in S} p_{xy}(a)V^{u_0}(y)$. For such a stationary policy u_0, we inductively define a sequence of stationary policies $\{u_t\}_{t\in\mathbf{T}}$ by

$$u_t(x) \in \arg\min_{a\in A} \left\{ c(x,a) + \sum_{y\in S} p_{xy}(a)V^{u_{t-1}}(y) \right\} \quad (x \in S), \tag{5}$$

where $\arg\min_{a\in A} f(a)$ is the set of arguments for which $f(a)$ attains its minimum and $u_t(x)$ is arbitrarily chosen from the right-hand side.

Proposition 8 *For a stationary policy u_0 such that $\max_{x\in S} V^{u_0}(x) < \infty$, we define $\{u_t\}_{t\in\mathbf{T}}$ as described above. Then, $V^{u_t}(x)$ is decreasing in t and converges to $V^*(x)$ as $t \to \infty$ for each $x \in S$. In particular, $V^*(x) \le V^{u_t}(x)$ for any t.*

We apply Proposition 8 to Example 14. For simplicity, we assume that $n \ge 3$. The $n = 2$ case is left to the reader as an exercise. First, we suppose a policy $u_0(x) = a_2$ for every x. Then,

$$V^{u_0}(x) = \begin{cases} Cx & \text{for } x = 0, 1, \ldots, n-1, \\ \frac{C}{2}(1+n) & \text{for } x = n. \end{cases}$$

It is clear that $u_1(x) = a_1$ for $x = 0, 1, \ldots, n-1$ since $C > 1$. For $x = n$,

$$c(n, a_i) + \sum_{y\in S} p_{ny}(a_i)V^{u_0}(y) = \begin{cases} 1 + C(n-1) & \text{for } i = 1, \\ C + \frac{1}{2}C(n-1) & \text{for } i = 2. \end{cases}$$

Then, it is easy to see that $u_1(n) = a_2$ when $n \ge 3$ since $C > 1$ and that

$$V^{u_1}(x) = \begin{cases} x & \text{for } x = 0, 1, \ldots, n-1, \\ c + \frac{1}{2}(n-1) & \text{for } x = n. \end{cases}$$

Similarly, it is clear that $u_2(x) = a_1$ for $x = 1, \ldots, n-1$ and that

$$c(n, a_i) + \sum_{y\in S} p_{ny}(a_i)V^{u_1}(y) = \begin{cases} 1 + (n-1) = n & \text{for } i = 1, \\ C + \frac{1}{2}(n-1) & \text{for } i = 2 \end{cases}$$

for $x = n$. Therefore, we have

$$u_2(x) = a_1 \ (x = 1, \ldots, n-1), \quad u_2(n) = \begin{cases} a_2 & \text{if } 1 < C \le \frac{n+1}{2}, \\ a_1 & \text{if } C \ge \frac{n+1}{2} \end{cases} \tag{6}$$

and

$$V^{u_2}(x) = \begin{cases} x & \text{for } x = 0, 1, \ldots, n-1, \text{ or for } x = n \text{ and } C \ge \frac{n+1}{2}, \\ C + \frac{n-1}{2} & \text{for } x = n \text{ and } 1 < C \le \frac{n+1}{2}. \end{cases}$$

It can be easily seen that $u_t(x) = u_2(x)$ for $t \ge 2$. Therefore, $u_2(x)$ given in (6) is an optimal policy.

2.3 An Example: iPod Shuffle

An iPod shuffle is an MP3 music player with a clickable control pad and an external button for switching between two different modes; one is sequential play mode and the other is random shuffle mode. Suppose that the playlist for your iPod is sorted by song title, say, $S = \{1, 2, \ldots, n\}$, and n is assumed to be the song you want to listen to. If you click the control pad in the sequential play mode, x goes to $x+1$, and if you click the control pad in the random shuffle mode, the next song is chosen uniformly at random from $\{1, 2, \ldots, n\}$. Intuitively, if you start at a song close enough to n, it might be better to stay in the sequential play mode, and if you start at a song far from n, it might be better to switch to the random shuffle mode until a song close to n is reached. The question is, what is the threshold for switching between two modes? This problem is well-modeled by a Markov decision process.

Example 15 (iPod shuffle) Let $S = \{1, 2, \ldots, n\}$ be a state space with n being the absorbing state. There are two actions $A = \{a_1, a_2\}$. If action a_1 is chosen, one moves upward by 1; if action a_2 is chosen, one jumps to a state uniformly at random. The costs of action a_1 and a_2 are 1 and T, respectively. We assume that $1 < T \ll n$. We apply Proposition 7 to this example. It follows from (4) that

$$V_t(x) = \min \left\{ 1 + V_{t-1}(x+1), \ T + \frac{1}{n} \sum_{y=1}^{n} V_{t-1}(y) \right\}, \quad x = 1, \ldots, n-1, \tag{7}$$

and $V_t(n) = 0 (\forall t = 0, 1, \ldots)$. From this expression, by induction, it is easy to see that $V_t(x)$ is decreasing in x for each t and that there exist $v_t > 0$ and $K_t \in \{1, \ldots, n\}$ such that

$$V_t(x) = \begin{cases} v_t & \text{for } x = 1, 2, \ldots, n - K_t, \\ n - x & \text{for } x = n - K_t + 1, n - K_t + 2, \ldots, n. \end{cases}$$

By Proposition 7, $V_t(x) \nearrow V^*(x)$ as $t \to \infty$, and hence we obtain $v > 0$ and $K \in \{1, \ldots, n\}$ such that

$$V^*(x) = \begin{cases} v & \text{for } x = 1, 2, \ldots, n-K, \\ n-x & \text{for } x = n-K+1, n-K+2, \ldots, n. \end{cases}$$

On the other hand, from (7),

$$V^*(x) = \min\left\{1 + V^*(x+1),\ T + \frac{1}{n}\sum_{y=1}^{n} V^*(y)\right\}, \quad \text{for } x = 1, 2, \ldots, n-1. \tag{8}$$

The second argument on the right-hand side does not depend on x and is equal to

$$C_n(v, K, T) = T + \frac{1}{n}\left\{(n-K)v + \frac{1}{2}K(K-1)\right\}.$$

Setting $x = 1$ in (8) yields

$$\begin{cases} v = \min\{1 + v,\ C_n(v, K, T)\} & K = 0, 1, \ldots, n-2, \\ v = \min\{n-1,\ C_n(v, n-1, T)\} & K = n-1, \\ n-1 = \min\{n-1,\ C_n(v, n, T)\} & K = n. \end{cases}$$

Since we assumed that $T \ll n$, we have that $C_n(v, n, T) < n-1$, and so $K \neq n$. It is also easy to see that $v = C_n(v, K, T)$ for $K \leq n-1$, which implies that

$$v = \frac{1}{2}(K-1) + \frac{nT}{K}. \tag{9}$$

Setting $x = n-K$ and $x = n-K+1$ in (8) yields $v = \min\{K, C_n(v, K, T)\}$ and $K - 1 = \min\{K-1, C_n(v, K, T)\}$. Hence, we have

$$K - 1 \leq v = C_n(v, K, T) \leq K.$$

By solving these inequalities together with (9), we have

$$\frac{\sqrt{1+8nT} - 1}{2} \leq K \leq \frac{\sqrt{1+8nT} + 1}{2}.$$

Therefore, we can see that $v \sim K \sim \sqrt{2nT}$ as $n \to \infty$. □

Remark 4 We refer the reader to Levin et al. [1] for a comprehensive account of the topics covered in Sect. 1, especially mixing time and cutoff phenomenon. Norris [2] provides additional details for the explanations in Sect. 2. The iPod example in Sect. 2.3 is taken from Norvig [3].

References

1. D.A. Levin, Y. Peres, E.L. Wilmer, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, 2009)
2. J.R. Norris, *Markov Chains* (Cambridge University Press, Cambridge, 1997)
3. P. Norvig, Doing the Martin Shuffle (with your iPod). Available at http://norvig.com/ipod.html