

Stochastic Processes II: Course Outline

Chris Hayduk

May 23, 2021

Note:

- Section 1 comes from *Essential of Stochastic Processes*, Third Edition by Durrett
- Sections 2-11 come from *Markov Chains and Mixing Times*, Second Edition by Levin and Peres
- Section 12 comes from the lecture notes *Mixing Times of Markov Chains: Techniques and Examples* by Berestycki

See each corresponding source for more details and examples.

1 Martingales

Thinking of M_n as the amount of money at time n for a gambler betting on a fair game and X_n as the outcomes of the gambling game, we say that M_0, M_1, \dots is a **martingale** with respect to X_0, X_1, \dots if

1. for any $n \geq 0$ we have $E|M_n| < \infty$
2. the value of M_n can be determined from the values of X_n, \dots, X_0 and M_0
3. for any possible values x_n, \dots, x_0 ,

$$E(M_{n+1} - M_n | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0, M_0 = m_0) = 0$$

Why is the above definition useful?

- The first condition, $E|M_n| < \infty$, is needed to guarantee that the conditional expectation makes sense.
- To motivate the second, we note that in many of our example X_n will be a Markov chain and $M_n = f(X_n, n)$. The conditioning event is formulated in terms of X_n because in passing from the random variables X_n that are driving the process the martingale to M_n , there may be a loss of information.

- The third, defining property says that conditional on the past up to time n , the average profit from the bet on the n th game is 0

In most cases, casino games are not fair but biased against the player. We say that M_n is a **supermartingale** with respect to X_n if a gambler's expected winnings on one play are negative

$$E(M_{n+1} - M_n | A_v) \leq 0$$

where $A_v = \{X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0, M_0 = m_0\}$

If we reverse the sign and suppose

$$E(M_{n+1} - M_n | A_v) \geq 0$$

then M_n is called a **submartingale** with respect to X_n .

Theorem 5.1. Let X_n be a Markov chain with transition probability p and let $f(x, n)$ be a function of the state x and the time n so that

$$f(x, n) \geq \sum_y p(x, y) f(y, n+1)$$

Then $M_n = f(X_n, n)$ is a supermartingale with respect to X_n .

Theorem 5.6. If M_m is a supermartingale and $m \leq n$, then $EM_m \geq EM_n$.

Theorem 5.7. If M_m is a submartingale and $0 \leq m < n$, then $EM_m \leq EM_n$,

Theorem 5.8. If M_m is a martingale and $0 \leq m < n$, then $EM_m = EM_n$.

The most famous result of martingale theory is that “you can't beat an unfavorable game.” To formulate and prove this, we will introduce a family of betting strategies that generalize the doubling strategy. The amount of money we bet on the n th game, H_n , clearly, cannot depend on the outcome of that game, nor is it sensible to allow it to depend on the outcome of games that will be played later. We say that H_n is an admissible gambling strategy or **predictable process** if for each n the value of H_n can be determined from $X_{n-1}, X_{n-2}, \dots, X_0, M_0$

To motivate the next definition, which we give for a general M_m , think of H_m as the amount of stock we hold between time $m-1$ and m , and M_m the price of a stock at time m . Then our wealth at time n is

$$W_n = W_0 + \sum_{m=1}^n H_m (M_m - M_{m-1})$$

since the change in our wealth from time $m-1$ to m is the amount we hold times the change in the price of the stock: $H_m (M_m - M_{m-1})$. To formulate the doubling strategy in this setting, let $X_m = 1$ if the result of the m th game is a win and -1 if the m th result is

a loss, and let $M_n = X_1 + \dots + X_n$ be the net profit of a gambler who bets on unit every time.

Theorem 5.9. Suppose that M_n is a supermartingale with respect to X_n , H_n is predictable, and $0 \leq H_n \leq c_n$, where c_n is a constant that may depend on n . Then

$$W_n = W_0 + \sum_{m=1}^n H_m(M_m - M_{m-1})$$

is a supermartingale.

We say that T is a **stopping time with respect to X_n** if the occurrence (or nonoccurrence) of the event $\{T = n\}$ can be determined from the information known at the time n , $X_n, X_{n-1}, \dots, X_0, M_0$.

Theorem 5.10. If M_n is a supermartingale with respect to X_n and T is a stopping time, then the stopped process $M_{T \wedge n}$ is a supermartingale with respect to X_n . In particular, $EM_{T \wedge n} \leq EM_0$.

Theorem 5.11. Suppose M_n is a martingale and T a stopping time with $P(T < \infty) = 1$ and $|M_{T \wedge n}| \leq K$ for some constant K . Then $EM_T = EM_0$.

2 Refresher on Markov Chain Basic Definitions

A chain P is called **irreducible** if for any two states $x, y \in \mathcal{X}$ there exists an integer t (possibly depending on x and y) such that $P^t(x, y) > 0$.

Let $\mathcal{T}(x) := \{t \geq 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for the chain to return to starting position x . The **period** of state x is defined to be the greatest common divisor of $\mathcal{T}(x)$.

Lemma 1.6. If P is irreducible, then $\gcd \mathcal{T}(x) = \gcd \mathcal{T}(y)$ for all $x, y \in \mathcal{X}$

For an irreducible chain, the period of the chain is defined to be the period which is common to all the states. The chain will be called **aperiodic** if all states have period 1. If a chain is not aperiodic, we call it **periodic**.

Proposition 1.7. If P is a aperiodic and irreducible, then there is an integer r_0 such that $P^r(x, y) > 0$ for all $x, y \in \mathcal{X}$ and $r \geq r_0$.

Given an arbitrary transition matrix P , let $Q = \frac{I+P}{2}$ where I is the $|\mathcal{X}| \times |\mathcal{X}|$ identity matrix. Since $Q(x, x) > 0$ for all $x \in \mathcal{X}$, the transition matrix Q is aperiodic. We call Q a **lazy version of P** .

The **time reversal** of an irreducible Markov chain with transition matrix P and stationary distribution π is the chain with matrix

$$\hat{P}(x, y) := \frac{\pi(y)P(y, x)}{\pi(x)}$$

The stationary equation $\pi = \pi P$ implies that \hat{P} is a stochastic matrix.

Suppose a probability distribution π on \mathcal{X} satisfies

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

for all $x, y \in \mathcal{X}$. The equation above is called the **detailed balance equation**. A chain satisfying detailed balance is called **reversible**.

Proposition 1.23. Let (X_t) be an irreducible Markov chain with transition matrix P and stationary distribution π . Write (\hat{X}_t) for the time-reversed chain with transition matrix \hat{P} . Then π is stationary for \hat{P} , and for any $x_0, \dots, x_t \in \mathcal{X}$ we have

$$P_\pi\{X_0 = x_0, \dots, X_t = x_t\} = P_\pi\{\hat{X}_0 = x_t, \dots, \hat{X}_t = x_0\}$$

A Markov chain is called **transitive** if for each pair $(x, y) \in \mathcal{X} \times \mathcal{X}$ there is a bijection $\varphi = \varphi_{(x, y)} : \mathcal{X} \rightarrow \mathcal{X}$ such that

$$\varphi(x) = y$$

and

$$P(z, w) = P(\varphi(z), \varphi(w))$$

for all $z, w \in \mathcal{X}$

Proposition 2.16. Let P be the transition matrix of a transitive Markov chain on a finite state space \mathcal{X} . Then the uniform probability distribution on \mathcal{X} is stationary for P .

3 Coupon Collecting

A company issues n different types of coupons. A collector desires a complete set. We suppose each coupon he acquires is equally likely to be each of the n types. How many coupons must he obtain so that his collection contains all n types?

It may not be obvious why this is a Markov chain. Let X_t denote the number of different types represented among the collector's first t coupons. Clearly, $X_0 = 0$. When the collector has coupons of k different types, there are $n - k$ types missing. Of the n possibilities for his next coupon, only $n - k$ will expand his collection. Hence,

$$P\{X_{t+1} = k + 1 \mid X_t = k\} = \frac{n - k}{n}$$

and

$$P\{X_{t+1} = k \mid X_t = k\} = \frac{k}{n}$$

Every trajectory of this chain is non-decreasing. Once the chain arrives at state n (corresponding to a complete collection), it is absorbed there. We are interested in the number of steps required to reach the absorbing state.

Proposition 2.3. Consider a collector attempting to collect a complete set of coupons. Assume that each new coupon is chosen uniformly and independently from the set of n possible types, and let τ be the (random) number of coupons collected when the set first contains every type. Then

$$E(\tau) = n \sum_{k=1}^n \frac{1}{k}$$

Proposition 2.4. Let τ be a coupon collector random variable, as in Proposition 2.3. For any $c > 0$,

$$P\{\tau > \lceil n \log n + cn \rceil\} \leq e^{-c}$$

4 Random Walks on Groups

Recall that a **group** is a set G endowed with an associative operation $\cdot : G \times G \rightarrow G$ and an **identity** $\text{id} \in G$ such that for all $g \in G$,

1. $\text{id} \cdot g = g$ and $g \cdot \text{id} = g$,
2. there exists an **inverse** $g^{-1} \in G$ for which $g \cdot g^{-1} = g^{-1} \cdot g = \text{id}$

Given a probability distribution μ on a group (G, \cdot) , we define the **left random walk on G with increment distribution μ** as follows: it is a Markov chain with state space G and which moves by multiplying the current state *on the left* by a random element of G selected according to μ . Equivalently, the transition matrix P of this chain has entries

$$P(g, hg) = \mu(h)$$

for all $g, h \in G$.

Proposition 2.12. Let P be the transition matrix of a random walk on a finite group G and let U be the uniform probability distribution on G . Then U is a stationary distribution for P .

For a set $H \subset G$, let $\langle H \rangle$ be the smallest group containing all the elements of H ; recall that every element of $\langle H \rangle$ can be written as a product of elements in H and their inverses. A set H is said to **generate** G if $\langle H \rangle = G$.

Proposition 2.13. Let μ be a probability distribution on a finite group G . The random walk on G with increment distribution μ is irreducible if and only if $S = \{g \in G : \mu(g) > 0\}$

generates G .

We call a set S of generators of G **symmetric** if $s \in S$ implies $s^{-1} \in S$. In parallel fashion, we call a probability distribution μ on a group G **symmetric** if $\mu(g) = \mu(g^{-1})$ for every $g \in G$.

Proposition 2.14. The random walk on a finite group G with increment distribution μ is reversible if μ is symmetric.

Every random walk on a group is transitive, so Proposition 2.16 from Section 2 holds (random walk on a group has a uniform probability distribution on G as the stationary distribution for P).

5 Intro to Markov Chain Mixing

The **total variation distance** between two probability distributions μ and ν on \mathcal{X} is defined by

$$\|\mu - \nu\|_{\text{TV}} = \max_{A \subset \mathcal{X}} |\mu(A) - \nu(A)|$$

This definition is explicitly probabilistic: the distance between μ and ν is the maximum difference between the probabilities assigned to a single event by the two distributions.

Proposition 4.2. Let μ and ν be two probability distributions on \mathcal{X} . Then

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|$$

A **coupling** of two probability distributions μ and ν is a pair of random variables (X, Y) defined on a single probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν . That is, a coupling (X, Y) satisfies $P\{X = x\} = \mu(x)$ and $P\{Y = y\} = \nu(y)$.

Coupling is a general and powerful technique; it can be applied in many different ways. Here, we offer a gentle introduction by showing the close connection between couplings of random variables and the total variation distance between those variables.

Proposition 4.7. Let μ and ν be two probability distributions on \mathcal{X} . Then

$$\|\mu - \nu\|_{\text{TV}} = \inf\{P\{X \neq Y\} : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

We will in fact show that there is a coupling (X, Y) which attains this infimum. We call such a coupling **optimal**.

Theorem 4.9 (Convergence Theorem). Suppose that P is irreducible and aperiodic with stationary distribution π . Then there exist constants $\alpha \in (0, 1)$ and $C > 0$ such that

$$\max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq C\alpha^t$$

Bounding the maximal distance (over $x_0 \in \mathcal{X}$) between $P^t(x_0, \cdot)$ and π is among our primary objectives. It is therefore convenient to define

$$d(t) := \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$$

It is often possible to bound $\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}}$ uniformly over all pairs of states (x, y) . We therefore make the definition

$$\bar{d}(t) := \max_{x, y \in \mathcal{X}} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}}$$

The relationship between d and \bar{d} is given below:

Lemma 4.10. If $d(t)$ and $\bar{d}(t)$ are as defined above, then

$$d(t) \leq \bar{d}(t) \leq 2d(t)$$

Lemma 4.11. The function \bar{d} is submultiplicative: $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$

It is useful to introduce a parameter which measures the time required by a Markov chain for the distance to stationarity to be small. The **mixing time** is defined by

$$t_{\text{mix}}(\epsilon) := \min\{t : d(t) \leq \epsilon\}$$

and

$$t_{\text{mix}} := t_{\text{mix}}(1/4)$$

We have that when ℓ is a positive integer,

$$d(\ell t_{\text{mix}}(\epsilon)) \leq \bar{d}(t_{\text{mix}}(\epsilon))^\ell \leq (2\epsilon)^\ell$$

In particular, taking $\epsilon = 1/4$ above yields

$$d(\ell t_{\text{mix}}) \leq 2^{-\ell}$$

and

$$t_{\text{mix}}(\epsilon) \leq \lceil \log_2 \epsilon^{-1} \rceil t_{\text{mix}}$$

For a distribution μ on a group G , the **reversed distribution** $\hat{\mu}$ is defined by $\hat{\mu}(g) := \mu(g^{-1})$ for all $g \in G$. Let P be the transition matrix of the random walk with increment distribution μ . Then the random walk with increment distribution $\hat{\mu}$ is exactly the time reversal \hat{P} of P .

Lemma 4.13. Let P be the transition matrix of a random walk on a group G with increment distribution μ and let \hat{P} be that of the walk on G with increment distribution $\hat{\mu}$. Let π be the uniform distribution on G . Then for any $t \geq 0$,

$$\|P^t(\text{id}, \cdot) - \pi\|_{\text{TV}} = \|\hat{P}^t(\text{id}, \cdot) - \pi\|_{\text{TV}}$$

Corollary 4.14. If t_{mix} is the mixing time of a random walk on a group and $\widehat{t_{\text{mix}}}$ is the mixing time of the reversed walk, the $t_{\text{mix}} = \widehat{t_{\text{mix}}}$

6 Coupling

6.1 Definition

As we defined previously, a coupling of two probability distributions μ and ν is a pair of random variables (X, Y) defined on the same probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν .

Couplings are useful because a comparison between distributions is reduced to a comparison between random variables. Proposition 4.7 characterized $\|\mu - \nu\|_{\text{TV}}$ as the minimum, over all couplings (X, Y) of μ and ν , of the probability that X and Y are different. This provides an effective method of obtaining upper bounds on the total variation distance.

In this chapter, we will extract more information by coupling not only pairs of distributions, but entire Markov chain trajectories.

We define a **coupling of Markov chains** with transition matrix P to be a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that both (X_t) and (Y_t) are Markov chains with transition matrix P , although the two chains may have different starting distributions.

Given a Markov chain \mathcal{X} with transition matrix P , a **Markovian coupling** of two P -chains is a Markov chain $\{(X_t, Y_t)\}_{t \geq 0}$ with state space $\mathcal{X} \times \mathcal{X}$ which satisfies, for all x, y, x', y' ,

$$\begin{aligned} P\{X_{t+1} = x' \mid X_t = x, Y_t = y\} &= P(x, x') \\ P\{Y_{t+1} = y' \mid X_t = x, Y_t = y\} &= P(y, y') \end{aligned}$$

Any Markovian coupling of Markov chains with transition matrix P can be modified so that the two chains stay together at all times after their first simultaneous visit to a single state - more precisely, so that

$$\text{if } X_s = Y_s, \text{ then } X_t = Y_t \text{ for } t \geq s \quad (1)$$

6.2 Bounding Total Variation Distance

We can use coupling to bound total variation distance.

Theorem 5.4. Let $\{(X_t, Y_t)\}$ be a coupling satisfying (1) for which $X_0 = x$ and $Y_0 = y$. Let τ_{couple} be the coalescence time of the chains:

$$\tau_{\text{couple}} := \min\{t : X_s = Y_s \text{ for all } s \geq t\}$$

Then

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq P_{x,y}\{\tau_{\text{couple}} > t\}$$

Corollary 5.5. Suppose that for each pair of states $x, y \in \mathcal{X}$ there is a coupling (X_t, Y_t) with $X_0 = x$ and $Y_0 = y$. For each such coupling, let τ_{couple} be the coalescence time of the chains as defined above. Then

$$d(t) \leq \max_{x,y \in \mathcal{X}} P_{x,y}\{\tau_{\text{couple}} > t\}$$

and therefore $t_{\text{mix}} \leq 4 \max_{x,y} E_{x,y}(\tau_{\text{couple}})$

Proposition 5.7. Let Q be an irreducible transition matrix and consider the lazy chain with transition matrix $P = (Q + I)/2$. The distributions at time t and $t + 1$ satisfy

$$\|P^t(x, \cdot) - P^{t+1}(x, \cdot)\|_{\text{TV}} \leq \frac{1}{\sqrt{t}}$$

7 Strong Stationary Times

7.1 Top-to-Random Shuffle

Consider the top-to-random shuffle: done by taking the top card and inserting it uniformly at random in the deck. Let \top be the time *one move after the first occasion when the original bottom card has moved to the top of the deck*. We show now that the arrangement of cards at τ_{top} is distributed normally on the set \mathcal{S}_n of all permutations of $\{1, \dots, n\}$ and moreover, this random element of \mathcal{S}_n is independent of the time τ_{top} .

Proposition 6.1. Let (X_t) be the random walk on \mathcal{S}_n corresponding to the top-to-random shuffle on n cards. Given at time t that there are k cards under the original bottom card, each of the $k!$ possible orderings of these cards are equally likely. Therefore, if τ_{top} is one shuffle after the first time that the original bottom card moves to the top of the deck, then the distribution $X_{\tau_{\text{top}}}$ is uniform over \mathcal{S}_n , and the time τ_{top} is independent of $X_{\tau_{\text{top}}}$.

The above proposition implies that, for any t , given that $\tau_{\text{top}} = t$, the distribution of X_t is uniform. In this chapter, we show that we can use the distribution of the random time τ_{top} to bound t_{mix} , the fixed number of steps needed for the distribution of the chain to be approximately stationary.

7.2 Markov Chains with Filtrations

Let $\{\mathcal{F}_t\}$ be a *filtration*, a sequence of σ -algebras such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all t . We say that $\{X_t\}$ is *adapted* to $\{\mathcal{F}_t\}$ if X_t is \mathcal{F}_t -measurable for all t . If $\mathcal{H}_t = \sigma(X_0, X_1, \dots, X_t)$, then $\{\mathcal{H}_t\}$ is called the *natural filtration*. Clearly, $\{X_t\}$ is adapted to the natural filtration.

Suppose that $\{X_t\}$ is adapted to $\{\mathcal{F}_t\}$. We say that $\{X_t\}$ is a **Markov chain with respect to $\{\mathcal{F}_t\}$** if

$$P_x\{X_{t+1} = y \mid \mathcal{F}_t\} = P(X_t, y)$$

where P is a transition matrix.

A **stopping time** for the filtration $\{\mathcal{F}_t\}$ is a random variable τ with values in $\{0, 1, \dots\}$ such that $\{\tau = t\} \in \mathcal{F}_t$. If the filtration of a stopping time is not specific, it will be assumed to be the natural filtration.

7.3 Stationary Times

Let (X_t) be an irreducible Markov chain with stationary distribution π . Suppose that $\{\mathcal{F}_t\}$ is a filtration and $\{X_t\}$ is adapted to $\{\mathcal{F}_t\}$. A **stationary time** τ for (X_t) is an $\{\mathcal{F}_t\}$ -stopping time, possibly depending on the starting position x , such that the distribution of X_τ is π :

$$P_x\{X_\tau = y\} = \pi(y) \text{ for all } y$$

We want to use τ_{top} to bound t_{mix} . To carry out this program, we need a property of τ_{top} stronger than the above. We will need that τ_{top} is independent of $X_{\tau_{\text{top}}}$.

7.4 Strong Stationary Times and Bounding Distance

Let (X_t) be a Markov chain with respect to the filtration $\{\mathcal{F}_t\}$ with stationary distribution π . A **strong stationary time** for (X_t) and starting position x is an $\{\mathcal{F}_t\}$ -stopping time τ , such that for all times t and y ,

$$P_x\{\tau = t, X_\tau = y\} = P_x\{\tau = t\}\pi(y)$$

In other words, X_τ has distribution π and is independent of τ .

Proposition 6.11. If τ is a strong stationary time for starting state x , then

$$\|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq P_x\{\tau > t\}$$

We introduce a new distance parameter $s_x(t)$, called the **separation distance** and defined by

$$s_x(t) := \max_{y \in \mathcal{X}} \left[1 - \frac{P^t(x, y)}{\pi(y)} \right]$$

The distance $s_x(t)$ is weakly decreasing in t . We also define

$$s(t) := \max_{x \in \mathcal{X}} s_x(t)$$

The Convergence Theorem implies that $s(t) \rightarrow 0$ as $t \rightarrow \infty$ for aperiodic irreducible chains.

Note that separation distance is a stronger notion of distance than $d(t)$. That is, generally $d(t) \leq s(t)$.

Lemma 6.12. If τ is a strong stationary time for starting state x , then

$$s_x(t) \leq P_x\{\tau > t\}$$

Definition 6.13. Given starting state x , a state y is a **halting state** for a stopping time τ if $X_t = y$ implies $\tau \leq t$. For example, when starting the lazy random walk on the hypercube at $(0, \dots, 0)$, the state $(1, \dots, 1)$ is a halting state for the stopping time τ_{refresh} .

Proposition 6.14. If there exists a halting state for starting state x , then τ is an optimal strong stationary time for x , i.e.

$$s_x(t) = P_x\{\tau > t\}$$

and it is stochastically dominated under P_x by every other strong stationary time.

8 Lower Bounds on Mixing Times

So far, we have focused on finding upper bounds on t_{mix} . It is natural to ask if a given upper bound is the best possible, and so in this chapter we turn to methods of obtaining lower bounds on t_{mix} .

8.1 Counting Bound

If the possible location of a chain after t steps do not form a significant fraction of the state space, then the distribution of the chain at time t cannot be close to uniform. This idea can be used to obtain lower bounds on the mixing time.

Let (X_t) be a Markov chain with irreducible and aperiodic transition matrix P on the state space \mathcal{X} . Define $d_{\text{out}}(x) := |\{y : P(x, y) > 0\}|$ to be the number of states accessible in one step from x , and let

$$\Delta := \max_{x \in \mathcal{X}} d_{\text{out}}(x)$$

Denote by \mathcal{X}_t^x the set of states accessible from x in exactly t steps and observe that $|\mathcal{X}_t^x| \leq \Delta^t$. If $\Delta^t < (1 - \epsilon)|\mathcal{X}|$, then from the definition of total variation distance we have that,

$$\|P^t(x, \cdot) - \pi\|_{\text{TV}} \geq P^t(x, \mathcal{X}_t^x) - \pi(\mathcal{X}_t^x) \geq 1 - \frac{\Delta^t}{|\mathcal{X}|} > \epsilon$$

This implies that

$$t_{\text{mix}}(\epsilon) \geq \frac{\log(|\mathcal{X}|(1 - \epsilon))}{\log \Delta}$$

In the reversible case when $\Delta \geq 3$, we have

$$|\mathcal{X}_t^x| \leq 1 + \Delta \sum_{j=0}^{t-1} (\Delta - 1)^j \leq 3(\Delta - 1)^t$$

so

$$t_{\text{mix}}(\epsilon) \geq \frac{\log(|\mathcal{X}|(1 - \epsilon)/3)}{\log(\Delta - 1)}$$

8.2 Diameter Bound

Given a transition matrix P on \mathcal{X} , construct a graph with vertex set \mathcal{X} and which includes the edge $\{x, y\}$ for all x and y with $P(x, y) + P(y, x) > 0$. Define the **diameter** of a Markov chain to be the diameter of this graph, that is, the maximal graph distance between distinct vertices.

Let P be an irreducible and aperiodic transition matrix on \mathcal{X} with diameter L and suppose that x_0 and y_0 are states at maximum graph distance L . Then $P^{\lfloor (L-1)/2 \rfloor}(x_0, \cdot)$ and $P^{\lfloor (L-1)/2 \rfloor}(y_0, \cdot)$ are positive on disjoint vertex sets. Consequently, $\bar{d}(\lfloor (L-1)/2 \rfloor) = 1$ and for any $\epsilon < 1/2$,

$$t_{\text{mix}}(\epsilon) \geq \frac{L}{2}$$

8.3 Bottleneck Ratio

Bottlenecks in the state space \mathcal{X} of a Markov chain are geometric features that control mixing time. A bottleneck makes portions of \mathcal{X} difficult to reach from some starting locations, limiting the speed of convergence.

As usual, P is the irreducible and aperiodic transition matrix for a Markov chain on \mathcal{X} with stationary distribution π .

The **edge measure** Q is defined by,

$$Q(x, y) := \pi(x)P(x, y), \quad Q(A, B) = \sum_{x \in A, y \in B} Q(x, y)$$

Here $Q(A, B)$ is the probability of moving from A to B in one step when starting from the stationary distribution.

The **bottleneck ratio** of the set S is defined to be,

$$\Phi(S) := \frac{Q(S, S^c)}{\pi(S)}$$

while the bottleneck ratio of the whole chain (also known as the **expansion**) is

$$\Phi_* := \min_{S: \pi(S) \leq \frac{1}{2}} \Phi(S)$$

For simple random walk on a graph with vertices \mathcal{X} and edge set E ,

$$Q(x, y) = \begin{cases} \frac{\deg(x)}{2|E|} \frac{1}{\deg(x)} = \frac{1}{2|E|} & \text{if } \{x, y\} \text{ is an edge,} \\ 0 & \text{otherwise} \end{cases}$$

In this case, $2|E|Q(S, S^c)$ is the size of the **boundary** ∂S of S , the collection of edges having one vertex in S and one vertex in S^c . The bottleneck ratio, in this case, becomes,

$$\Phi(S) = \frac{|\partial S|}{\sum_{x \in S} \deg(x)}$$

Theorem 7.4. If Φ_* is the bottleneck ratio defined previously, then

$$t_{\text{mix}} = t_{\text{mix}}(1/4) \geq \frac{1}{4\Phi_*}$$

8.4 Distinguishing Statistics

One way to produce a lower bound on the mixing time t_{mix} is to find a statistic f (a real-valued function) on \mathcal{X} such that the distance between the distribution of $f(X_t)$ and the distribution of f under the stationary distribution π can be bounded from below.

Let μ and ν be two probability distributions on \mathcal{X} , and let f be a real-valued function defined on \mathcal{X} . We write E_μ to indicate expectations of random variables (on a sample space \mathcal{X}) with respect to the probability distribution μ :

$$E_\mu(f) := \sum_{x \in \mathcal{X}} f(x) \mu(x)$$

Proposition 7.9. For $f : \mathcal{X} \rightarrow \mathbb{R}$, define $\sigma_*^2 := \max\{\text{Var}_\mu(f), \text{Var}_\nu(f)\}$. If

$$|E_\nu(f) - E_\mu(f)| \geq r\sigma_*$$

then

$$\|\mu - \nu\|_{\text{TV}} \geq 1 - \frac{8}{r^2}$$

In particular, if for a Markovian chain (X_t) with transition matrix P the function f satisfies

$$|E_x[f(X_t)] - E_\pi(f)| \geq r\sigma_*$$

then

$$\|P^t(x, \cdot) - \pi\|_{\text{TV}} \geq 1 - \frac{8}{r^2}$$

Lemma 7.10. Let μ and ν be probability distributions on \mathcal{X} , and let $f : \mathcal{X} \rightarrow \Lambda$ be a function on \mathcal{X} , where Λ is a finite set. Then

$$\|\mu - \nu\|_{\text{TV}} \geq \|\mu f^{-1} - \nu f^{-1}\|_{\text{TV}}$$

Lemma 7.13. Consider the coupon collecting problem with n distinct coupon types, and let $I_j(t)$ be the indicator event that the j -th coupon has not been collected by time t . Let $R_t = \sum_{j=1}^n I_j(t)$ be the number of coupon types not collected by time t . The random variables $I_j(t)$ are negatively correlated, and letting $p = (1 - \frac{1}{n})^t$, we have for $t \geq 0$

$$\begin{aligned} E(R_t) &= np \\ \text{Var}(R_t) &\leq np(1-p) \leq \frac{n}{4} \end{aligned}$$

Proposition 7.14. For the lazy random walk on the n -dimensional hypercube,

$$d\left(\frac{1}{2}n \log n - \alpha n\right) \geq 1 - 8e^{2-2\alpha}$$

9 The Symmetric Group and Shuffling Cards

9.1 The Symmetric Group

A permutation of $\{1, 2, \dots, n\}$ is a bijection from $\{1, 2, \dots, n\}$ to itself. The set of all permutations forms a group, the symmetric group \mathcal{S}_n under the composition operation. The identity element of \mathcal{S}_n is the identity function $\text{id}(k) = k$. Every $\sigma \in \mathcal{S}_n$ has a well-defined inverse function, which is its inverse in the group.

A probability distribution μ on the symmetric group describes a mechanism for shuffling cards: apply permutation σ to the deck with probability $\mu(\sigma)$. Repeatedly shuffling the deck using this mechanism is equivalent to running the random walk on the group with increment distribution μ . As discussed earlier, as long as the support of μ generates all of \mathcal{S}_n , the resulting chain is irreducible. If $\mu(\text{id}) > 0$, then it is aperiodic. Every shuffle chain has uniform stationary distribution.

We will consider a permutation as a map from positions to labels. For example, the permutation σ defined by,

$$\begin{aligned}\sigma(1) &= 3 \\ \sigma(2) &= 1 \\ \sigma(3) &= 2 \\ \sigma(4) &= 4\end{aligned}$$

corresponds to placing card 3 into position 1, card 1 into position 2, card 2 into position 3, and card 4 into position 4.

We will often find it convenient to use **cycle notation** for permutations. In this notation, (abc) refers to the permutation σ for which $b = \sigma(a)$, $c = \sigma(b)$, and $a = \sigma(c)$. When several cycles are written consecutively, they are performed one at a time *from right to left*.

A cycle of length n is called an n -cycle. A **transposition** is a 2-cycle.

We describe a simple algorithm for generating an *exactly* uniform random permutation. Let σ_0 be the identity permutation. For $k = 1, 2, \dots, n - 1$, inductively construct σ_k from σ_{k-1} by swapping the cards at location k and J_k , where J_k is an integer picked uniformly in $\{k, \dots, n\}$, independently of $\{J_1, \dots, J_{k-1}\}$. More precisely,

$$\sigma_k(i) = \begin{cases} \sigma_{k-1}(i) & \text{if } i \neq J_k, i \neq k \\ \sigma_{k-1}(J_k) & \text{if } i = k \\ \sigma_{k-1}(k) & \text{if } i = J_k \end{cases}$$

That is, $\sigma_k = \sigma_{k-1} \circ (kJ_k)$. This method requires $n - 1$ steps, which is optimal.

Given a permutation $\sigma \in \mathcal{S}_n$, consider the sign of the product

$$M(\sigma) = \prod_{1 \leq i < j \leq n} (\sigma(j) - \sigma(i))$$

Clearly, $M(\text{id}) > 0$ since every term is positive. For every $\sigma \in \mathcal{S}_n$ and every transposition (ab) , we have,

$$M(\sigma \circ (ab)) = -M(\sigma)$$

Why? We assume that $a < b$. Then for every c such that $a < c < b$, two factors change sign (the one that pairs c with a and also the one that pairs c with b), while the single factor

containing both a and b also changes sign.

Call a permutation σ **even** if $M(\sigma) > 0$ and otherwise call σ **odd**. Note that a permutation is even (odd) if and only if every way of writing it as a product of transpositions contains an even (odd) number of factors. The set of all even permutations in \mathcal{S}_n forms a subgroup, known as the **alternating group** A_n .

Note that an m -cycle can be written as a product of $m - 1$ transpositions. Hence, an m cycle is odd (even) when m is even (odd), and the sign of any permutation is determined by its disjoint cycle decomposition.

9.2 Random Transposition

To avoid periodicity, the random transposition shuffle is defined as follows: at time t , choose two cards, labeled L_t and R_t , independently and uniformly at random. If L_t and R_t are different, transpose them. Otherwise, do nothing. The resulting distribution μ satisfies,

$$\mu(\sigma) = \begin{cases} 1/n & \text{if } \sigma = \text{id} \\ 2/n^2 & \text{if } \sigma = (ij) \\ 0 & \text{otherwise} \end{cases}$$

This chain is irreducible and aperiodic.

9.2.1 Lower bound

Lower bound is given by,

Proposition 8.4. Let $0 < \epsilon < 1$. For the random transposition chain on an n -card deck,

$$t_{\text{mix}}(\epsilon) \geq \frac{n-1}{2} \log \left(\frac{1-\epsilon}{6} n \right)$$

9.2.2 Upper bound via coupling

Now for an upper bound. For the coupling, we take a slightly different view of generating the transpositions. At each time t , the shuffler chooses a card with label $X_t \in [n]$ and, independently, a position $Y_t \in [n]$; she then transposes the card labeled X_t with the card in position Y_t . Of course, if the card in position Y_t already has the label X_t , the deck is left unchanged.

To couple two decks, use the same choices (X_t) and (Y_t) to shuffle both. Let (σ_t) and (σ'_t) be the two trajectories. What can happen in one step? Let a_t be the number of cards that occupy the same position in both σ_t and σ'_t .

- If the card labeled X_t is in the same position in both decks, then $a_{t+1} = a_t$

- If X_t is in different positions in the two decks but position Y_t is occupied by the same card, then performing the specified transposition breaks one alignment but also forms a new one. We have $a_{t+1} = a_t$
- If X_t is in different positions in the two decks and if the cards at position Y_t in the two decks do not match, then at least one new alignment is made - and possibly as many as three.

Proposition 8.5. Let τ be the time required for the two decks to coincide. Then, no matter the initial configurations of the two decks, $E(\tau) < \frac{\pi^2}{6}n^2$

9.2.3 Upper bound via strong stationary time

Proposition 8.6. In the random transposition shuffle, let R_t and L_t be the cards chosen by the right and left hands, respectively, at time t . Assume that when $t = 0$, no cards have been marked. At time t , mark card R_t if both of the following are true:

- R_t is unmarked
- Either L_t is a marked card or $L_t = R_t$

Let τ be the time when every card has been marked. Then τ is a strong stationary time for the chain.

9.3 Riffle Shuffles

A method often used to shuffle real decks of 52 cards is the following: first, the shuffler cuts the deck into two piles. Then, the piles are “riffled” together: she successively drops cards from the bottom of each pile to form a new pile. There are two undetermined aspects of this procedure. First, the numbers of cards in each pile after the initial cut can vary. Second, real shufflers drop varying numbers of cards from each stack as the deck is reassembled.

For mathematicians, there is a tractable mathematical model for riffle shuffling. Here are three ways to shuffle a deck of n cards:

1. Let M be a Binomial($n, 1/2$) random variable, and split the deck into its top M cards and its bottom $n - M$ cards. There are $\binom{n}{M}$ ways to riffle these two piles together, preserving the relative order within each pile (first select the positions for the top M cards; then fill in both piles). Choose one of these arrangements uniformly at random.
2. Let M be a Binomial($n, 1/2$) random variable, and split the deck into its top M cards and its bottom $n - M$ cards. The two piles are then held over the table and cards are dropped one by one, forming a single pile once more, according to the following recipe: if, at a particular moment, the left pile contains a cards and the right pile contains b cards, then drop the card on the bottom of the left pile with probability $a/(a + b)$ and the card on the bottom of the right pile with probability $b/(a + b)$. Repeat this procedure until all cards have been dropped.

3. Label the n cards with n independent fairly chosen bits. Pull all the cards labeled 0 to the top of the deck, preserving their relative order.

A **rising sequence** of a permutation σ is a maximal set of consecutive values that occur in the correct relative order in σ .

We claim that *methods (1) and (2) generate the same distribution Q on permutations, where*

$$Q(\sigma) = \begin{cases} (n+1)/2^n & \text{if } \sigma = \text{id} \\ 1/2^n & \text{if } \sigma \text{ has exactly two rising sequences,} \\ 0 & \text{otherwise} \end{cases}$$

Recall that for a distribution R on \mathcal{S}_n , the **reverse distribution** \hat{R} satisfies $\hat{R}(\rho) = R(\rho^{-1})$. We claim that *method (3) generates \hat{Q}* . Why? The cards labeled 0 form one increasing sequence ρ^{-1} , and the cards labeled 1 form the other. (Again, there are $n+1$ ways to get the identity permutation, namely, all strings of the form $00\dots 011\dots 1$). Alternatively, the number M of cards labeled 0 has a $\text{Binomial}(n, 1/2)$ distribution, and given M , the locations of these cards are uniform among the $\binom{n}{M}$ possibilities. Thus, method (3) is indeed the reversal of method (1).

Thanks to Lemma 4.13 (which says that a random walk on a group and its inverse, both started from the identity, have the same distance from uniformity after the same number of steps), it will suffice to analyze method (3).

Consider repeated inverse riffle shuffles using method (3). For the first shuffle, each card is assigned a random bit, and all the 0's are pulled ahead of all the 1's. For the second shuffle, each card is again assigned a random bit, and all the 0's are pulled ahead of all the 1's. Considering both bits (and writing the second bit on the left), we see that cards labeled 00 precede those labeled 01, which precede those labeled 10, which precede those labeled 11. After k shuffles, each card will be labeled with a string of k bits, and cards with different labels will be in lexicographic order (cards with the same label will be in their original relative order).

Proposition 8.11. Let τ be the number of inverse riffle shuffles required for all cards to have different bitstring labels. Then τ is a strong stationary time.

Proposition 8.12. For the riffle shuffle on an n -card deck, $t_{\text{mix}} \leq 2 \log_2(4n/3)$ for sufficiently large n .

Proposition 8.13. Fix $0 < \epsilon, \delta < 1$. Consider riffle shuffling an n -card deck. For sufficiently large n ,

$$t_{\text{mix}}(\epsilon) \geq (1 - \delta) \log_2 n$$

10 Eigenvalues

10.1 The Spectral Representation of a Reversible Transition Matrix

For a transition matrix P , a function f on \mathcal{X} is an eigenfunction with corresponding eigenvalue λ if $Pf = \lambda f$. If P is not reversible, then the eigenfunctions and eigenvalues may not be real.

We begin by collecting some elementary facts about the eigenvalues of transition matrices:

Lemma 12.1. Let P be the transition matrix of a finite Markov chain.

1. If λ is an eigenvalue of P , the $|\lambda| \leq 1$
2. If P is irreducible, the vector space of eigenfunctions corresponding to the eigenvalue 1 is the one-dimensional space generated by the column vector $\vec{1} := (1, 1, \dots, 1)^T$
3. If P is irreducible and aperiodic, then -1 is not an eigenvalue of P

Denote by $\langle \cdot, \cdot \rangle$ the usual inner product on $\mathbb{R}^{\mathcal{X}}$, given by $\langle f, g \rangle = \sum_{x \in \mathcal{X}} f(x)g(x)$. We will also need another inner product, denoted by $\langle \cdot, \cdot \rangle_{\pi}$ and defined by

$$\langle f, g \rangle_{\pi} := \sum_{x \in \mathcal{X}} f(x)g(x)\pi(x)$$

We write $\ell^2(\pi)$ for the vector space $\mathbb{R}^{\mathcal{X}}$ equipped with the above inner product.

Recall that the transition matrix P is reversible with respect to the stationary distribution π if $\pi(x)P(x, y) = \pi(y)P(y, x)$ for all $x, y \in \mathcal{X}$. The reason for introducing the above inner product is,

Lemma 12.2. Let P be reversible with respect to π .

1. The inner product space $(\mathbb{R}^{\mathcal{X}}, \langle \cdot, \cdot \rangle_{\pi})$ has an orthonormal basis of real-valued eigenfunctions $\{f_j\}_{j=1}^{|\mathcal{X}|}$ corresponding to eigenvalues $\{\lambda_j\}$.
2. The matrix P can be decomposed as

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{j=1}^{|\mathcal{X}|} f_j(x)f_j(y)\lambda_j^t$$

3. The eigenfunction f_1 corresponding to the eigenvalue 1 can be taken to be the constant $\vec{1}$, in which case,

$$\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{j=2}^{|\mathcal{X}|} f_j(x)f_j(y)\lambda_j^t$$

Lemma 12.3. If φ is an eigenfunction of the transition matrix P with eigenvalue $\lambda \neq 1$, then $E_{\pi}(\varphi) = 0$

10.2 The Relaxation Time

Define

$$\lambda_* := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P, \lambda \neq 1\}$$

The difference $\gamma_* := 1 - \lambda_*$ is called the **absolute spectral gap**. Lemma 12.1 implies that if P is aperiodic and irreducible, then $\gamma_* > 0$.

For a reversible matrix P , we label the eigenvalues P in decreasing order

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\mathcal{X}|} \geq -1$$

The **spectral gap** of a reversible chain is defined by $\gamma := 1 - \lambda_2$. If the chain is lazy, then $\gamma_* = \gamma$

The **relaxation time** t_{rel} of a reversible Markov chain with absolute spectral gap γ_* is defined to be

$$t_{\text{rel}} := \frac{1}{\gamma_*}$$

One operational meaning of the relaxation time comes from the inequality

$$\text{Var}_\pi(P^t f) \leq (1 - \gamma_*)^{2t} \text{Var}_\pi(f)$$

Theorem 12.4. Let P be the transition matrix of a reversible, irreducible Markov chain with state space \mathcal{X} , and let $\pi_{\min} := \min_{x \in \mathcal{X}} \pi(x)$. Then,

$$\begin{aligned} t_{\text{mix}}(\epsilon) &\leq t_{\text{rel}} \log \left(\frac{1}{\epsilon \pi_{\min}} \right) \\ t_{\text{mix}}^{(\infty)}(\epsilon) &\leq \lceil t_{\text{rel}} \log \left(\frac{1}{\epsilon \pi_{\min}} \right) \rceil \end{aligned}$$

Theorem 12.5. Suppose that $\lambda \neq 1$ is an eigenvalue for the transition matrix P of an irreducible and aperiodic Markov chain. Then,

$$t_{\text{mix}}(\epsilon) \geq \left(\frac{1}{1 - |\lambda|} - 1 \right) \log \left(\frac{1}{2\epsilon} \right)$$

In particular, for reversible chains,

$$t_{\text{mix}}(\epsilon) \geq (t_{\text{rel}} - 1) \log \left(\frac{1}{2\epsilon} \right)$$

Remark 12.6. If the absolute spectral gap γ_* is small because the smallest eigenvalue $\lambda_{|\mathcal{X}|}$ is near -1 , but the spectral gap γ is not small, the slow mixing suggested by this lower bound can be rectified by passing to a lazy chain to make the eigenvalues positive.

Corollary 12.7. For a reversible, irreducible, and aperiodic Markov chain,

$$\lim_{t \rightarrow \infty} d(t)^{1/t} = \lambda_*$$

10.3 Eigenvalues and Eigenfunctions of Some Simple Random Walks

10.3.1 The cycle

Let $\omega = e^{2\pi i/n}$. In the complex plane, the set $W_n := \{\omega, \omega^2, \dots, \omega^{n-1}, 1\}$ of the n -th roots of unity forms a regular n -gon inscribed in the unit circle. Since $\omega^n = 1$, we have

$$\omega^j \omega^k = \omega^{k+j} = \omega^{k+j \bmod n}$$

Hence (W_n, \cdot) is a cyclic group of order n , generated by ω . In this section, we view simple random walk on the n -cycle as the random walk on the (multiplicative) group W_n with increment distribution uniform on $\{\omega, \omega^{-1}\}$. Let P be the transition matrix of this walk. Every (possibly complex-valued) eigenfunction f of P satisfies

$$\lambda f(\omega^k) = Pf(\omega^k) = \frac{f(\omega^{k-1}) + f(\omega^{k+1})}{2}$$

for $0 \leq k \leq n-1$.

For $0 \leq j \leq n-1$, define $\varphi_j(\omega^k) := \omega^{kj}$. Then,

$$P\varphi_j(\omega^k) = \frac{\varphi_j(\omega^{k-1}) + \varphi_j(\omega^{k+1})}{2} = \frac{\omega^{j(k-1)} + \omega^{j(k+1)}}{2} = \omega^{jk} \left(\frac{\omega^{-j} + \omega^j}{2} \right)$$

Hence φ_j is an eigenfunction of P with eigenvalue

$$\lambda_j = \frac{\omega^j + \omega^{-j}}{2} = \cos(2\pi j/n)$$

What is the underlying geometry? For any ℓ and j , the average of the vectors $\omega^{\ell-j}$ and $\omega^{\ell+j}$ is a scalar multiple of ω^ℓ . Since the chord connecting $\omega^{\ell+j}$ with $\omega^{\ell-j}$ is perpendicular to ω^ℓ , the projection of $\omega^{\ell+j}$ onto ω^ℓ has length $\cos(2\pi j/n)$.

Because φ_j is an eigenfunction of the real matrix P with a real eigenvalue, both its real part and its imaginary part are eigenfunctions. In particular, the function $f_j : W_n \rightarrow \mathbb{R}$ defined by,

$$f_j(\omega^k) = \operatorname{Re}(\varphi_j(\omega^k)) = \operatorname{Re}(e^{2\pi ijk/n}) = \cos\left(\frac{2\pi jk}{n}\right)$$

is an eigenfunction. We have $\lambda_2 = \cos(2\pi/n) = 1 - \frac{4\pi^2}{2n^2} + O(n^{-4})$, so the spectral gap γ is of order n^{-2} .

When $n = 2m$ is even, $\cos(2\pi m/n) = -1$ is an eigenvalue, so $\gamma_* = 0$. The walk in this case is periodic.

11 The Cutoff Phenomenon

11.1 Definition

Suppose, for a sequence of Markov chains indexed by $n = 1, 2, \dots$, the mixing time for the n -th chain is denoted by $t_{\text{mix}}^{(n)}(\epsilon)$. This sequence of chains has a **cutoff** if, for all $\epsilon \in (0, 1)$,

$$\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\epsilon)}{t_{\text{mix}}^{(n)}(1 - \epsilon)} = 1$$

The upper and lower bound for the top-to-random chain (given by $d_n(n \log n + \alpha n) \leq e^{-\alpha}$ and $\liminf_{n \rightarrow \infty} d_n(n \log n - \alpha n) \geq 1 - 2e^{2-\alpha}$) show that the total variation distance d_n for the n -card chain “falls off a cliff” at $t_{\text{mix}}^{(n)}$. More precisely, when time is rescaled by $n \log n$, as $n \rightarrow \infty$, the function d_n approaches a step function:

$$\lim_{n \rightarrow \infty} d_n(cn \log n) = \begin{cases} 1 & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases}$$

In fact, this property characterizes when a sequence of chains has a cutoff.

Lemma 18.1. Let $t_{\text{mix}}^{(n)}$ and d_n be the mixing time and distance to stationarity, respectively, for the n -th chain in a sequence of Markov chains. The sequence has a cutoff if and only if

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

A sequence of Markov chains has a cutoff **window** of size $O(w_n)$ if $w_n = o(t_{\text{mix}}^{(n)})$ and,

$$\begin{aligned} \lim_{\alpha \rightarrow -\infty} \liminf_{n \rightarrow \infty} d_n(t_{\text{mix}}^{(n)} + \alpha w_n) &= 1 \\ \lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n(t_{\text{mix}}^{(n)} + \alpha w_n) &= 0 \end{aligned}$$

We say a family of chains has a **pre-cutoff** if it satisfies the weaker condition

$$\sup_{0 < \epsilon < 1/2} \limsup_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\epsilon)}{t_{\text{mix}}^{(n)}(1 - \epsilon)} < \infty$$

11.2 Examples of Cutoff

11.2.1 Biased random walk on a line segment

Theorem 18.2. The lazy random walk (X_t) with bias $\beta = p - 1/2$ on $\{0, 1, 2, \dots, n\}$ has a cutoff at $\beta^{-1}n$ with a window of size $O(\sqrt{n})$. More precisely, there is constant $c(\beta) > 0$ such that for all $\alpha \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} d_n\left(\frac{n}{\beta} + \alpha\sqrt{n}\right) = \Phi(-c(\beta)\alpha)$$

where Φ is the standard normal distribution function.

11.2.2 Random walk on the hypercube

Theorem 18.3. The lazy random walk on the n -dimensional hypercube has a cutoff at $(1/2)n \log n$ with a window size of $O(n)$.

12 More on Riffle Shuffle

Theorem 11.1. There is a cutoff phenomenon at time $t_{\text{mix}} = \frac{3}{2} \log_2 n$

Theorem 11.2 After m shuffles,

$$P(X_m = \sigma) = \frac{1}{2^{mn}} \binom{2^m + n - R(\sigma)}{n}$$

where $R(\sigma)$ is the number of rising sequences of σ .

Theorem 11.3. Let $m = \log_2(n^{3/2}c)$. Then

$$d(m) = 1 - 2\Phi\left(-\frac{1}{4\sqrt{3}c}\right) + O(n^{-1/4})$$

where $\Phi(x)$ is the cumulative distribution function of a standard normal random variable:

$$\Phi(x) = \int_{-\infty}^x e^{-u^2/2} \frac{du}{\sqrt{2\pi}}$$