
Lecture 17: Approximate Counting and Sampling

1 Introduction

We continue the discussion of polynomial time approximate sampling schemes started in the previous lecture. To begin with, we review concepts of finite state Markov chains. Then we discuss techniques to bound the mixing time of Markov chains. In particular we discuss two techniques: the method of *canonical paths*, and *coupling*. Finally, we discuss two applications: random walk on hypercubes, and random walk on matchings.

2 Review of Finite state Markov chains

Let \mathcal{M} denote a finite state Markov chain with state space Ω (assumed to be finite), and (P_{xy}) the transition probabilities, where

$$\begin{aligned} P_{xy} &= Pr[X_1 = y | X_0 = x] \\ P_{xy}^t &= Pr[X_t = y | X_0 = x] \end{aligned}$$

A probability distribution π ($\pi(x) \geq 0$, $\forall x \in \Omega$, and $\sum_{x \in \Omega} \pi(x) = 1$) on the state space Ω is called a *stationary distribution*, if $\pi(y) = \sum_{x \in \Omega} \pi(x) P_{xy}$, $\forall y \in \Omega$. From the theory of Markov chains we have the following result:

THEOREM 1

An irreducible, aperiodic Markov chain has a unique stationary distribution π such that, for all $x, y \in \Omega$, we have $P_{xy}^t \rightarrow \pi(y)$ as $t \rightarrow \infty$.

We use the *total variation distance* as a measure of distance between two probability distributions. It is defined as

$$D(\pi, \pi') = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)| = \max_{A \subset \Omega} (|\pi(A) - \pi'(A)|)$$

We will be interested in computing bounds on the mixing time of Markov chains with huge state spaces, in which the states are typically combinatorial objects. We define the mixing time, $\tau_x(\epsilon)$ to be

$$\tau_x(\epsilon) = \min\{t \mid \forall t' \geq t \ D(P_x^{t'}, \pi) \leq \epsilon\}$$

where $P_x^{t'}$ is the probability distribution of the states at time t' , given that the chain started from state x . The mixing time is the minimum time taken, such that the probability

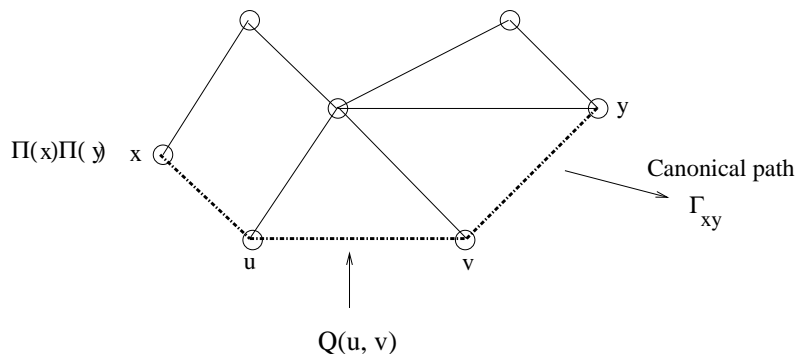


Figure 1: An example showing a canonical path between the states x and y . The dashed-dotted line in the figure is the canonical path associated with the pair of states (x, y) .

distribution of the states of the Markov chain is within a tolerance ϵ of the unique stationary distribution with respect to the total variation distance.

In this handout we assume that all the Markov chains considered are *Lazy*, i.e., $P_{xx} \geq \frac{1}{2}$ for all $x \in \Omega$.

Also, a probability measure π is called *reversible* if it satisfies

$$\pi(x)P_{xy} = \pi(y)P_{yx}, \quad \forall x, y \in \Omega$$

A Markov chain is called *reversible* if there exists a probability distribution π satisfying the above equality. It is easy to check that such a π is a stationary distribution. For notational convenience, define $Q(x, y) := \pi(x)P_{xy}$.

From now on we restrict our attention to finite, irreducible, aperiodic, lazy and reversible Markov chains. Also, π is always used to denote the unique stationary probability distribution of this chain. We now describe the method of Canonical Paths, which is a very useful tool to obtain bounds on the mixing time.

3 The Method of Canonical Paths

Associate with each pair of states (x, y) a *canonical path* $\Gamma_{xy} = x_0, x_1, x_2, \dots, x_{|\Gamma_{xy}|}$, such that $x_0 = x$, $x_{|\Gamma_{xy}|} = y$ and $P_{x_i x_{i+1}} > 0$ for all $i = 0, 1, \dots, |\Gamma_{xy}| - 1$. An instance of a canonical path in a Markov chain is given in figure 1.

Define,

$$\bar{\rho} = \max_{(u,v):Q(u,v)>0} \frac{1}{Q(u,v)} \sum_{\Gamma_{xy} \ni (u,v)} \pi(x)\pi(y)|\Gamma_{xy}| \quad (1)$$

where the sum is over all pairs of states x, y such that (u, v) is in the corresponding canonical path Γ_{xy} . Here, $|\Gamma_{xy}|$ denotes the length of the canonical path associated with the pair of states (x, y) . We can relate this to a multi-commodity flow problem by assuming that the capacity of each edge (u, v) to be $Q(u, v)$ and $\pi(x)\pi(y)$ to be the flow between the pair of states x, y . The usefulness of $\bar{\rho}$ stems from the following theorem:

THEOREM 2

For all x and $\epsilon > 0$ we have,

$$\tau_x(\epsilon) \leq \bar{\rho} \ln \left(\frac{1}{\epsilon \pi(x)} \right)$$

The proof of this result can be found on the home page of Mark Jerrum, in the file entitled “Chapter 5”. The above theorem gives us a bound on the mixing time of Markov chains. This result is true for all choices of canonical paths Γ_{xy} . However, one needs to make a proper choice of the canonical path in order to get tight bounds on the mixing time.

3.1 Application: Random walk on a hypercube

Consider a Markov chain whose state space is the hypercube, $\{0, 1\}^n =: \Omega$. Define the transition probabilities by

$$P_{xy} = \begin{cases} \frac{1}{2} & \text{if } x = y \\ \frac{1}{2n} & \text{if the Hamming distance between } x \text{ and } y \text{ is } 1 \\ 0 & \text{otherwise} \end{cases}$$

It is easy to see that the chain is irreducible, aperiodic and lazy. Also, by symmetry $\pi(x) = 2^{-n}$, $\forall x \in \Omega$ is a stationary probability distribution. It is also easy to check that the chain is reversible.

Now, with every pair of states (x, y) we associate a canonical path Γ_{xy} according to the following rule: The intermediate states in the path are obtained by flipping bits where x and y differ, in left to right order. Note that our canonical path never uses the self loop edges. For example, consider the case when $n = 5$, $x = (1, 1, 0, 1, 1)$ and $y = (1, 0, 0, 0, 0)$. The canonical path associated with x and y is

$$(1, 1, 0, 1, 1) \rightarrow (1, 0, 0, 1, 1) \rightarrow (1, 0, 0, 0, 1) \rightarrow (1, 0, 0, 0, 0)$$

We now try to compute $\bar{\rho}$ for this example from (1). Consider an edge (u, v) where $u = (u_1, u_2, \dots, u_i, u_{i+1}, \dots, u_n)$ and $v = (u_1, u_2, \dots, \bar{u}_i, u_{i+1}, \dots, u_n)$. The transition from state u to state v is performed by flipping the i th bit. In order to compute a bound on $\bar{\rho}$, we need to find out all the pairs (x, y) for which $\Gamma_{xy} \ni (u, v)$. Given the pair (u, v) it is clear that the following bits are determined:

$$\begin{aligned} (y_1, y_2, \dots, y_{i-1}) &= (u_1, u_2, \dots, u_{i-1}) \\ x_i &= u_i \\ y_i &= \bar{u}_i \\ (x_{i+1}, x_{i+2}, \dots, x_n) &= (u_{i+1}, u_{i+2}, \dots, u_n) \end{aligned}$$

Hence the undetermined bits are x_1, x_2, \dots, x_{i-1} , and $y_{i+1}, y_{i+2}, \dots, y_n$. Therefore the number of possibilities for (x, y) such that $\Gamma_{xy} \ni (u, v)$ is $2^{i-1} \cdot 2^{n-i} = 2^{n-1}$. Also, $Q(u, v) := \Pi(u)P_{uv} = 2^{-n} \cdot \frac{1}{2n}$. So,

$$\begin{aligned} \sum_{\Gamma_{xy} \ni (u, v)} \pi(x)\pi(y)|\Gamma_{xy}| &\leq \sum_{\Gamma_{xy} \ni (u, v)} 2^{-n} \cdot 2^{-n} \cdot n \\ &= 2^{n-1} \cdot 2^{-n} \cdot 2^{-n} \cdot n \end{aligned}$$

The inequality in the first step is obtained by upper bounding the length of the canonical path, $|\Gamma_{xy}|$ by n . Using this bound in (1), we get

$$\begin{aligned}\bar{\rho} &\leq \frac{2^{n-1} \cdot 2^{-n} \cdot 2^{-n} \cdot n}{Q(u, v)} \\ &= \frac{2^{n-1} \cdot 2^{-n} \cdot 2^{-n} \cdot n}{2^{-n} \cdot \frac{1}{2^n}} \\ &= n^2\end{aligned}\tag{2}$$

Hence, we get the following polynomial time bound on the mixing time

$$\begin{aligned}\tau_x(\epsilon) &\leq \bar{\rho} \ln \left(\frac{1}{\epsilon \pi(x)} \right) = n^2 \ln \left(\frac{2^n}{\epsilon} \right) \\ &= n^2 \left(n + \ln \frac{1}{\epsilon} \right)\end{aligned}$$

This bound on the mixing time is very loose and we can improve it to get a $(n \log n)$ bound. For this we use the method of *coupling* of Markov chains.

The method of Markov chain Coupling

Under the technique of *coupling*, two correlated copies of the same Markov chain are run together and we try to compute a bound on how quickly they collide. Let the two copies of the chain be denoted by $\{X_t\}$ and $\{Y_t\}$, where X_0 and Y_0 are arbitrary states of Ω . The Markov chain (X_t, Y_t) on $\Omega \times \Omega$ is coupled such that

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

for all $x, y, x', y' \in \Omega$. In other words, if we simply observe the individual coordinates, they behave like P_{xy} . We demonstrate the advantage of coupling for the example of a random walk on the hypercube.

Our idea is to force the two copies of the chain starting at different states to come together. Let Y_t be another copy of the chain with Y_0 is uniformly distributed on $\{0, 1\}^n$. We construct a coupling of X_t and Y_t by letting U_1, U_2, \dots be uniform on $\{1, 2, \dots, 2n\}$. At time t , the j th coordinates of X and Y are each set equal to 1 if $U_t = 2j - 1$ and are each set equal to 0 if $U_t = 2j$. The other coordinates are left unchanged. It is each to check that our construction constitutes a coupling, i.e., the marginals look like individual random walks on the hypercube. Also, note that once the two chains collide, i.e, $X_t = Y_t$, they always remain glued to each other, i.e, $X_n = Y_n$ for all $n \geq t$.

Define the coupling time of this two dimensional Markov chain to be

$$\gamma(\epsilon) = \min\{t \mid \text{After } t \text{ steps, } Pr(X_t = Y_t \mid X_0 = x, Y_0 = y) \geq 1 - \epsilon\} \quad \forall x, y \in \Omega$$

By arguments similar to the coupon collector problem we can show that the coupling time, $\gamma(\epsilon) = O(n \log n + \log(\frac{1}{\epsilon}))$. The advantage of computing the coupling time stems from the following result, which is true for any coupling of Markov chains.

EXERCISE 1

$$\max_x \tau_x(\epsilon) \leq \gamma(\epsilon)$$

Hence we have shown that the mixing time for a random walk on a hypercube is bounded by $\gamma(\epsilon) = O\left(n \log n + \log\left(\frac{1}{\epsilon}\right)\right)$.

Method of Encoding function

While bounding the mixing time for the random walk on a hypercube using the method of canonical paths we assumed the knowledge of the number of states. In many applications, the states in the Markov chain are combinatorial objects, like the matchings in a graph, and hence the number of states is unknown or hard to evaluate. We now give an alternative approach to cases where the number of states is unknown.

For a fixed pair of states (u, v) , we define a one-one function

$$\eta_{u,v} : \{(x, y) \mid (u, v) \in \Gamma_{xy}\} \longrightarrow \Omega$$

where Ω is the state space of the chain. The existence of this one-one function implies that the size of the domain of this function is at most $|\Omega|$. This function $\eta_{u,v}$ is typically called an encoding function.

For the example of a random walk on the hypercube, define

$$\eta_{u,v}(x, y) = (x_1, x_2, \dots, x_i, y_{i+1}, y_{i+2}, \dots, y_n)$$

It is easy to check the following fact.

EXERCISE 2 $\eta_{u,v}(x, y)$ as defined above uniquely identifies x and y and hence is a one-one function.

We will also use the facts that: for all x , $\pi(x) = \frac{1}{|\Omega|}$, for $x \neq y$, $\Gamma_{xy} \leq n$ and $P_{xy} = \frac{1}{2n}$ if a transition from x to y is possible. Thus $Q(u, v) = \pi(u)P_{uv} = \frac{1}{|\Omega|} \cdot \frac{1}{2n}$, and

$$\sum_{\Gamma_{xy} \ni (u,v)} \pi(x)\pi(y)|\Gamma_{xy}| \leq |\Omega| \left(\frac{1}{|\Omega|^2} \right) \cdot n$$

Hence we have $\bar{\rho} \leq 2n^2$, which is a slightly looser bound than the one in (2).

3.2 Application: Monomer-dimer systems

Next, we apply the method of canonical paths to a problem of sampling from monomer-dimer systems. A monomer-dimer system is a cover of the vertices of a graph $G = (V, E)$ by monomers (molecules covering one vertex) and dimers (molecules covering a pair of adjacent vertices). An example of a monomer-dimer system for a 4×4 grid graph is given in figure 2. Note that a monomer-dimer configuration with k dimers corresponds to a matching of size k . The *weight* or *energy* of a monomer-dimer system is λ^k , where λ is a given non-negative constant and k is the number of dimers. For a given graph $G = (V, E)$, let m_k be the

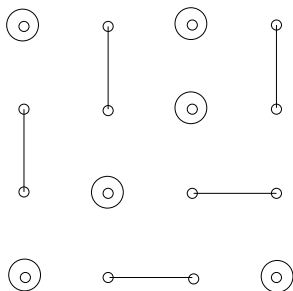


Figure 2: Example of a monomer-dimer system

number of monomer-dimer systems with k dimers. Let $Z(\lambda) = \sum m_k \lambda^k$ (known as the *partition function*) be the sum of the weights of the monomer-dimer systems. The partition function can also be viewed as $Z(\lambda) = \sum_{M: M \text{ is a Matching}} \lambda^{|M|}$.

It is known that computing $Z(\lambda)$ is $\#\mathbf{P}$ hard. Our ultimate goal is to approximately compute $Z(\lambda)$. However, our immediate goal is to sample from the monomer-dimer systems, such that the probability of a monomer-dimer system of weight k is $\frac{\lambda^k}{Z(\lambda)}$. How to go from approximate sampling to approximate counting will be discussed in the next lecture. Since the dimers in a monomer-dimer system form a matching, the object is to sample from the matchings, where a matching of size k has probability $\frac{\lambda^k}{Z(\lambda)}$.

We define a Markov chain whose states are the matchings in $G = (V, E)$. The transitions from a state M to a new state M' are made as follows:

1. With probability $\frac{1}{2}$, $M' = M$.
2. Select an edge $e = \{u, v\}$ uniformly at random from E

$$M' = \begin{cases} M - e & \text{if } e \in M \\ M + e & \text{if both } u \text{ and } v \text{ are unmatched in } M \\ M + e - e' & \text{if exactly one of } u \text{ and } v \text{ is matched in } M \text{ and } e' \text{ is the matching edge} \\ M & \text{otherwise} \end{cases}$$

3. Go to M' with probability $\min \left\{ 1, \frac{\lambda^{|M'|}}{\lambda^{|M|}} \right\}$.

For this Markov chain it is easy to check that:

EXERCISE 3 : The chain is reversible with stationary distribution $\pi(M) = \frac{\lambda^{|M|}}{Z(\lambda)}$.

We can also show that

THEOREM 3

The mixing time of this chain satisfies:

$$\tau_x(\epsilon) \leq 4|E| \left(n\lambda(\ln n\lambda) + \ln \frac{1}{\epsilon} \right)$$

The above theorem can be proved by the use of the method of Canonical Paths. The detailed proof of this theorem is given in the next lecture.