CS294 Markov Chain Monte Carlo: Foundations & Applications	**Fall 2009**

Lecture 5: September 17

Lecturer: Prof. Alistair Sinclair *Scribes: Suman Ravuri*

5.1 Coupling for bounding the mixing time

Consider as usual an ergodic (i.e., irreducible, aperiodic) Markov chain on some state space Ω. Consider two particles started at positions x and y in Ω, each individually moving through the state space according to the Markov transition matrix P, but whose evolutions may be coupled in some way. We will show below that the time until the two particles meet gives a bound on the mixing time; more precisely, we will show that

$$\Delta(t) \leq \max_{x,y} \Pr[\text{two particles started at positions } x,\, y \text{ have not met by time } t],$$

where we recall that

$$\Delta(t) = \max_{x} \|p_x^{(t)} - \pi\|.$$

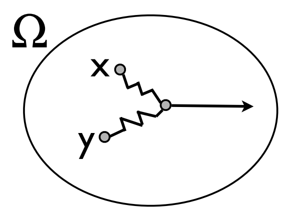

Figure 5.1: Coupling

More formally:

Definition 5.1 *A* coupling *of a Markov chain P is a pair process (X_t, Y_t) such that:*

1. *each of $(X_t, \cdot)$ and $(\cdot, Y_t)$, viewed in isolation, is a faithful copy of the Markov chain; that is,*

$$\Pr[X_{t+1} = b \,|\, X_t = a] = P(a, b) = \Pr[Y_{t+1} = b \,|\, Y_t = a];$$

and

2. *if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.*

Figure 5.1 gives a pictorial illustration of this definition.

Now define the random variable $T_{xy} = \min\{t : X_t = Y_t \,|\, X_0 = x, Y_0 = y\}$ to be the (stopping) time until the two processes meet. The following claim gives the desired upper bound on the mixing time:

Claim 5.2

$$\Delta(t) \le \max_{x,y} Pr[T_{xy} > t]$$

.

Proof: Recall that we defined in Lecture 3 that:

$$D(t) = \max_{x,y} \|p_x^{(t)} - p_y^{(t)}\|$$

and that $\Delta(t) \le D(t)$. Now,

$$\begin{aligned}
\Delta(t) &\le D(t) \\
&= \max_{x,y} \|p_x^{(t)} - p_y^{(t)}\| \\
&\le \max_{x,y} \Pr[X_t \ne Y_t \,|\, X_0 = x, Y_0 = y] \\
&= \max_{x,y} \Pr[T_{xy} > t \,|\, X_0 = x, Y_0 = y].
\end{aligned}$$

The only real content in this proof is the third line, where we use the coupling lemma from Lecture 3. This completes our proof. ■

Coupling ideas for analyzing the time-dependent behavior of Markov chains can be traced back to Doeblin in the 1930s. However, the modern development of the topic was initiated by David Aldous [Al83].

5.2 Examples

5.2.1 Simple random walk on the hypercube $\{0,1\}^n$

The n-dimensional cube is a graph with 2^n vertices, each of which can be encoded as an n-bit binary string $b_1 b_2 \cdots b_n$, whose neighbours are the strings which differ from it by Hamming distance exactly 1. We define a random walk on the cube by the following:

1. With probability 1/2, do nothing.
2. Else, pick a coordinate $i \in \{1, \ldots, n\}$ uniformly at random and flip coordinate x_i (i.e. $x_i \to 1 - x_i$).

This setup is clearly equivalent to the following:

1. Pick a coordinate $i \in \{1, \ldots, n\}$ uniformly at random *and* a bit $b \in \{0,1\}$ uniformly at random.
2. Set $x_i = b$.

This second description of the random walk dynamics suggests the following coupling: make X_t and Y_t choose the *same* i and b at every step. Clearly this is a valid coupling: obviously each of X_t and Y_t is performing exactly the above random walk.

To analyze the time T_{xy}, notice that once every $i \in \{1, \ldots, n\}$ has been chosen at least once, X_t must equal Y_t. (This is because, once a coordinate i has been chosen, X_t and Y_t agree on that coordinate at all future times.) Thus for any x and y, T_{xy} is stochastically dominated by the time for a coupon collector to collect all n coupons. Thus $\Pr[T_{xy} > n \ln n + cn] < e^{-c}$, and hence by Claim 5.2 we have $\Delta(n \ln n + cn) \le e^{-c}$; therefore in particular

$$\tau_{\text{mix}} \le n \ln n + O(n),$$

and more generally

$$\tau(\varepsilon) \le n \ln n + \lceil n \ln(\varepsilon^{-1}) \rceil.$$

An exact analysis of this very simple random walk reveals that in fact $\tau_{\text{mix}} \sim (1/2) n \ln n$, so our analysis is tight up to a factor of 2.

5.2.2 Another random walk on the hypercube

The above coupling was a bit trite because of the self-loop probability of 1/2. In this example we consider the same random walk but with a different self-loop probability. This means we will have to be a bit more careful in defining the coupling.

1. With probability $\frac{1}{n+1}$, do nothing.
2. Otherwise, with probability $\frac{1}{n+1}$ for each neighbor, go to one of the neighbors.

This can be reformulated as:

1. Pick $i \in \{0, 1, \ldots, n\}$.
2. If $i = 0$ do nothing; otherwise, flip x_i.

Note that the self-loop probability here is $\frac{1}{n+1}$.

To define a coupling here, we'll write $d(X_t, Y_t)$ for the number of coordinates in which X_t and Y_t differ.

If $d(X_t, Y_t) > 1$, then:

1. if X_t picks $i = 0$ then Y_t picks $i = 0$.
2. if X_t picks coordinate i where they agree, Y_t also picks i.
3. if X_t picks coordinate i where they *disagree*, then Y_t picks coordinate $f(i)$, where f is a cyclic permutation on the disagreeing coordinates.

   (For example, if $X_t = 110100$ and $Y_t = 000010$, then the permutation could be the one that sends $1, 2, 4, 5$ to $2, 4, 5, 1$ respectively.)

If $d(X_t, Y_t) = 1$, then, letting i_0 be the disagreeing coordinate:

1. if X_t picks $i = 0$, then Y_t picks i_0.
2. if X_t picks $i = i_0$, then Y_t picks 0.
3. else, both X_t and Y_t pick the same i.

Note that the distance between X_t and Y_t never increases under this coupling.

In this scenario, the time for the distance between X_t and Y_t to decrease to at most 1 is dominated by the coupon collector for $n/2$ coupons, and so with high probability this is $\sim (n/2)\ln n + O(n)$. The time for the last single disagreement to disappear (that is, the regime in which $d(X_t, Y_t) = 1$, if this happens) has a geometric distribution with mean n, so contributes $O(n)$ to the time for the chains to meet. Summing these two contributions and using Claim 5.2 gives

$$\tau_{\text{mix}} \leq \frac{1}{2} n \ln n + O(n).$$

For this example, Diaconis and Shahshahani [DS81] have shown that the true value of the mixing time is $\tau_{\text{mix}} \sim (1/4) n \ln n + O(n)$, so again our analysis is tight up to a factor of 2.

Exercise: For every self-loop probability δ with $\delta > \text{const}/n$ and $\delta < 1 - \text{const}/n$, show that $\tau_{\text{mix}} \leq c_\delta n \ln n + O(n)$, where c_δ is a constant. (The exact value of c_δ is not important.)

Exercise: Give a strong stationary time argument for the random walk on $\{0,1\}^n$ with self-loop probability 1/2 (that is, the walk discussed in section 5.2.1).

5.2.3 Top-in-at-Random Shuffle

As previously discussed, the top-in-at-random shuffle involves repeatedly taking the top card from a deck of n cards and inserting it at a position chosen uniformly at random in the deck. Analyzing this shuffle by coupling is best done through its inverse, defined as follows:

- Pick a card c from the deck uniformly at random.
- Move card c to the top of the deck.

Recall from Lecture 4 that the mixing times of the original shuffle and the inverse shuffle are identical.

To construct a coupling, we will envision the two decks of n cards X_t and Y_t at time t. We define the coupling by making both X_t and Y_t choose the *same* card c (which of course is not necessarily in the same position in both decks) and move it to the top. Now the key observation is the following: once a card has been chosen in the coupling, this card will be in the same position in both decks for the rest of time. [**Exercise:** Check this.] T_{xy} is therefore once again dominated by the coupon collector random variable for n coupons. This leads to

$$\tau_{\text{mix}} \leq n \ln n + O(n)$$

and

$$\tau(\varepsilon) \leq n \ln n + \lceil n \ln \varepsilon^{-1} \rceil.$$

As we saw in the previous lecture, this value is tight.

5.2.4 Random Transposition Shuffle

This shuffle is defined as follows:

- Pick cards c and c' uniformly at random.
- Switch c and c'.

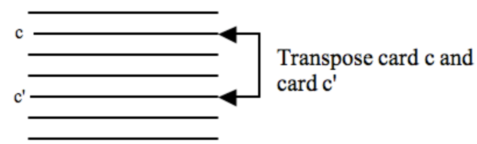

Figure 5.2: Random Transposition Shuffle with cards c and c'

An equivalent, more convenient description is the following:

- Pick card c and position p uniformly at random.
- Exchange card c with the card at position p in the deck.

It is easy to define a coupling using this second definition: namely, make X_t and Y_t choose the same c and p at each step. This coupling ensures that the distance between X and Y is non-increasing. More explicitly, writing $d_t = d(X_t, Y_t)$ for the number of positions at which the two decks differ, we have the following case analysis:

1. If card c is in the same position in both decks, then $d_{t+1} = d_t$.
2. If card c is in different positions in the two decks, there are two possible subcases:
   (a) If the card at position p in both decks is the same, then $d_{t+1} = d_t$.
   (b) Otherwise, $d_{t+1} \leq d_t - 1$.

Thus we get a decrease in distance only in case 2(b), and this occurs with probability

$$\Pr[d_{t+1} < d_t] = \left(\frac{d_t}{n}\right)^2.$$

Therefore, the time for d_t to decrease from value d is stochastically dominated by a geometric random variable with mean $\left(\frac{n}{d}\right)^2$. This implies that $E[T_{xy}] \leq \sum_{d=1}^{n} \left(\frac{n}{d}\right)^2$, which is $O(n^2)$.

Invoking Markov's inequality, we get that $\Pr[T_{xy} > cn^2] < c' = \frac{1}{2\varepsilon}$ for a suitable constant c, which leads to the bound

$$\tau_{\text{mix}} \leq cn^2.$$

Actually, for this shuffle it is known that

$$\tau_{\text{mix}} \sim \frac{1}{2} n \ln n,$$

so our analysis in this case is off by quite a bit.

Exercise: Design a better coupling that gives $\tau_{\text{mix}} \leq O(n \ln n)$.

Remark: It turns out that for any ergodic Markov chain there is *always* a coupling that is optimal, in the sense that the coupling time satisfies

$$Pr[T_{xy} > t] = D_{xy}(t).$$

This is a very general theorem of Griffeath [Gr78]. However, the couplings that achieve the mixing time may involve looking arbitrarily far into the future in the two Markov chains, and thus are not useful in practice. The couplings we have used so far—and almost all couplings used in algorithmic applications—are *Markovian* couplings: the evolution of X_t and Y_t depends only on the current values of X_t and Y_t. For some Markov chains, there is (provably) no Markovian coupling that achieves the mixing time [KR99]. Recently some exciting progress has been made in the use of non-Markovian couplings; see, for example [HV03].

5.2.5 Graph Colorings

This section introduces the problem of sampling colorings of a graph using a Markov chain. This is of interest, for example, when estimating the number of legal colorings for a given graph and color set, a fundamental combinatorial problem. It also has applications to the antiferromagnetic Potts model in statistical physics. Throughout, our input will be an undirected graph $G = (V, E)$, where V is a set of n vertices, and E is a set of edges connecting V. The maximum degree of any vertex in V is denoted Δ. We will also be given a set of colors $Q = \{1, \ldots, q\}$.

Our goal is to sample proper colorings of G uniformly at random. A coloring labels each vertex with a color $c \in Q$. A proper coloring is defined as one where no two vertices connected by an edge share the same color.

The decision problem for graph coloring involves deciding whether it is possible to construct a proper coloring for a given graph. Some important results about the graph coloring decision problem:

- If $q \geq \Delta + 1$, the answer to the decision problem is trivially true.
- if $q = \Delta$, the graph G has a proper coloring unless it contains a $(\Delta + 1)$-clique or is an odd cycle (in which case $\Delta = 2$); this is a classical result known as Brooks' Theorem.
- if $q < \Delta$, the decision problem is NP complete.

Obviously constructing a random coloring is at least as hard as solving the decision problem, so we shouldn't expect to be able to do so in polynomial time unless $q \geq \Delta$. In fact, the decision problem for $q = \Delta$ is also somewhat non-trivial, so we will generally be interested only in the case when $q \geq \Delta + 1$.

We are concerned with sampling proper colorings using a Markov chain. One natural Markov chain is as follows, where the current state is any proper coloring of G:

- Pick a vertex $v \in V$ uniformly at random and a color $c \in \{1, \ldots, q\}$ uniformly at random.
- Recolor v with c if this yields a proper coloring, else do nothing.

This Markov chain is symmetric and aperiodic, but is not always irreducible. In particular, if $q \leq \Delta + 1$, we can have "frozen" colorings in which no move is possible (even though other proper colorings do exist). One very simple such example is given in figure 5.3, in which $\Delta = 2$ and $q = 3$.

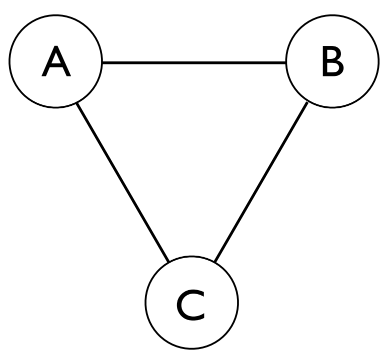

Figure 5.3: Proper coloring, $\Delta = 2$, $q = 3$.

However, it turns out that this problem cannot occur when the number of colors is at least $\Delta + 2$:

Important Exercise: show that the above Markov chain is irreducible if $q \geq \Delta + 2$.

Two interesting conjectures regarding proper graph coloring:

1. Random sampling of proper colorings can be done in polynomial time whenever $q \geq \Delta + 1$.
2. The Markov chain outlined above has mixing time $O(n \ln n)$ whenever $q \geq \Delta + 2$.

Unfortunately, we cannot yet prove either of these conjectures. However, we will now begin to approach the second conjecture by showing that, if $q \geq 4\Delta + 1$, then the mixing time is $O(n \log n)$. Note that intuitively we would expect that, the more colors we have, the easier it should be to prove that the the Markov chain is rapidly mixing. This is because with more colors the vertices are "more independent."

Claim 5.3 *Provided that $q \geq 4\Delta + 1$, the mixing time is $O(n \log n)$.*

This theorem (in the stronger version $q \geq 2\Delta + 1$) is due to Jerrum [Je95], and independently to Salas and Sokal [SS97]. Note that $O(n \log n)$ mixing time is the best we could hope for, since by coupon collecting it takes that long before all the vertices have a chance to be recolored. (Actually the coupon-collecting analogy is not strictly accurate; however, the $\Omega(n \log n)$ lower bound can be proved by a more careful argument [HS05].)

Proof: We will apply coupling to two copies X_t and Y_t of the chain. Our coupling procedure is very simple: Let X_t and Y_t choose the same vertex v and the same color c at each step, and recolor if possible. Let us now analyze this coupling.

Define $d_t := d(X_t, Y_t)$ to be the number of vertices where the colors disagree. For every step in our chain a vertex v and a color c are chosen, which could result in:

- "Good moves" (d_t decreases by 1): the chosen vertex v is a disagreeing vertex, and the chosen color c is not present at any neighbor of v in either X_t or Y_t. In this case, the coupling will recolor v to the same color in both X_t and Y_t, thus eliminating one disagreeing vertex. Since the neighbors of v have at most 2Δ distinct colors in X_t and Y_t, there are at least $d_t(q - 2\Delta)$ good moves available.
- "Bad moves" (d_t increases by 1): the chosen vertex v is not a disagreeing vertex but is a neighbor of some disagreeing vertex v', and the chosen color c is one of the colors of v' (in either X_t or Y_t). In this case, v will be recolored in one of the chains but not the other, resulting in a new disagreement at v. Since each disagreeing vertex v' has at most Δ neighbors, and there are only two corresponding "bad" colors c, the number of bad moves available is at most $2d_t\Delta$.
- "Neutral moves" (d_t is unchanged): all choices of (v, c) that do not fall into one of the above two categories lead to no change in d_t.

Note that the difference between the numbers of good and bad moves is (at least) $d_t(q - 4\Delta)$, so we expect the distance to decrease when $q \geq 4\Delta + 1$. We now make this argument precise.

Since each move has the same probability, namely $\frac{1}{qn}$, we can compute the expected change in d_t under one step of the coupling:

$$E[d_{t+1}|X_t, Y_t] \leq d_t - \frac{d_t(q - 2\Delta)}{qn} + \frac{2d_t\Delta}{qn} = d_t\left(1 - \frac{q - 4\Delta}{qn}\right).$$

Iterating this we get that

$$E[d_t|X_0, Y_0] \leq \left(1 - \frac{q - 4\Delta}{qn}\right)^t d_0 \leq \left(1 - \frac{q - 4\Delta}{qn}\right)^t n \leq \left(1 - \frac{1}{qn}\right)^t n. \tag{5.1}$$

This implies that

$$\begin{aligned}
\Pr[d_t > 0|X_t, Y_t] &= \Pr[d_t \geq 1|X_t, Y_t] \\
&\leq \mathrm{E}[d_t|X_t, Y_t] && \text{(Markov's Ineq.)} \\
&\leq \left(1 - \frac{1}{qn}\right)^t n.
\end{aligned}$$

To ensure that $\Delta(t) \leq \epsilon$, it is sufficient to let $t = qn(\ln n + \ln \epsilon^{-1})$. Thus in particular the mixing time is $\tau_{\text{mix}} = O(qn \log n)$. ■

Remark: The fact that the above mixing time bound increases with q is an artefact of our crude approximation $q - 4\Delta \geq 1$ in the last step of line (5.1). Clearly the value q in the final bound can be replaced by $\frac{q}{q-4\Delta}$.

Exercise (harder): Devise an improved coupling that achieves a similar $O(n \log n)$ mixing time for $q \geq 2\Delta + 1$.

References

[Al83] D. Aldous, "Random walks on finite groups and rapidly mixing Markov chains," *Séminaire de Probabilites XVII*, pages 243- 297. Springer Verlag, 1983. Lecture Notes in Mathematics 986.

[DS81] P. Diaconis and M. Shahshahani, "Generating a random permutation with random transpositions," *Zeitschrift für Wahrscheinlichketistheorie und verwandte Gebiete* **57** (1981), pp. 159–179.

[Gr78] D. Griffeath. "Coupling methods for Markov processes," in *Studies in Probability and Ergodic Theory*, G.-C. Rota ed., Academic Press, 1978, pp. 1–43.

[HS05] T. Hayes and A. Sinclair, "A general lower bound for mixing of single-site dynamics on graphs," *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005, pp. 511–520.

[Je95] M.R. Jerrum, "A very simple algorithm for estimating the number of k-colourings of a low-degree graph," *Random Structures and Algorithms* **7** (1995), pp. 157–165.

[HV03] T.P. Hayes and E. Vigoda, "A non-Markovian coupling for randomly sampling colorings," *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp. 618–627.

[KR99] V.S. Anil Kumar and H. Ramesh, "Markovian coupling vs. conductance for the Jerrum-Sinclair chain," *Proceedings of the 40th FOCS*, 1999, pp. 241–250.

[SS97] J. Salas and A.D. Sokal, "Absence of Phase Transition for Antiferromagnetic Potts Models via the Dobrushin Uniqueness Theorem," *Journal of Statistical Physics* **98** (2000), p. 551.