**CS294-2 Markov Chain Monte Carlo: Foundations & Applications** **Fall 2002**

# Lecture 7: September 24

*Lecturer: Alistair Sinclair* *Scribes: Barbara Engelhardt and Xuanlong Nguyen*



## 7.1 Review from Lecture 6

Last lecture, we discussed *path coupling* as a method of simplifying difficult coupling arguments. A **premetric** $d$ on $\Omega$ is a connected graph with positive edge weights such that every edge is a shortest path between the two nodes the edge connects (i.e., there are no "redundant" edges). We call such nodes *adjacent*. (Note that there is no requirement that the premetric be associated with the transitions in the Markov chain, though in practice there will be some connection if the premetric is to be useful.) A premetric extends in the obvious way to a metric on $\Omega$ (just take shortest path distances in the graph).

We proved the following path coupling lemma:

**Theorem 7.1** *If we can define a coupling $(X, Y) \mapsto (X'Y')$ for adjacent pairs $(X, Y)$ such that*

$$E[d(X', Y')|X, Y] \leq (1 - \alpha)d(X, Y) \tag{7.1}$$

*for some $\alpha \in [0, 1]$, then the coupling can be extended to all pairs $(X, Y)$ such that equation (7.1) still holds.*

If equation (7.1) holds with $\alpha > 0$, then $d$ has a drift towards 0 and it is easy to check (see Problem Set 2) that the mixing time is bounded by $O\left(\frac{1}{\alpha} \log D\right)$, where $D = \max_{X,Y} d(X, Y)$ is the maximum distance in the metric.

In the boundary case $\alpha = 0$, the distance does not decay geometrically, but by comparison with unbiased random walk on the integers $[0, \ldots, D]$ we would expect the distance to hit 0 in $O(D^2)$ steps. To be precise, we need one more condition: that the probability of a change in distance in the coupling at each time step is at least $\beta > 0$. Then we can conclude (see Problem Set 2) that the mixing time is $O\left(\frac{D^2}{\beta}\right)$.

## 7.2 Linear Extensions of a Partial Order

We now give another application of path coupling, due to Bubley and Dyer [BD98].

**Input**: A partial order $\prec$ on a set of $n$ elements (see figure 7.1).

**Goal:** Sample uniformly at random from $\Omega = \{\text{linear extensions of } \prec\}$.

In other words, sample u.a.r. from the set of total orderings that are consistent with the given partial order. This, and the associated problem of counting the number of such orderings, has applications to (e.g.) the analysis of sorting algorithms, since $\Omega$ represents our state of knowledge of some underlying total order after some pairwise element comparisons have been performed. The counting problem is difficult in general ($\#P$-complete [BW91]). Nonetheless, we will see how to construct a Markov chain that samples from $\Omega$

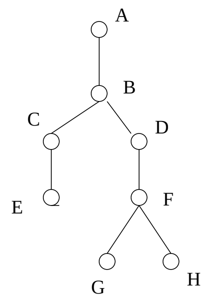


Figure 7.1: Partial order on elements $A, B, C, D, E, F, G, H$. Note that $E$ and $F$, for example, are incomparable in this partial order.

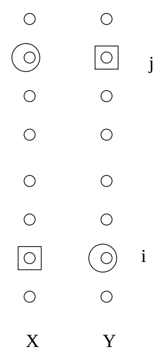


Figure 7.2: Illustration of two adjacent linear orders in the premetric. Their distance is $|j - i|$, where $i$ and $j$ are the locations of the two positions at which $X$ and $Y$ differ; at all other positions $X$ and $Y$ are equal.

in polynomial time. As a byproduct, this yields a polynomial time algorithm for approximating $|\Omega|$ (using general techniques discussed in the next lecture).

**Markov Chain:**

1. Pick a position $p \in \{1, 2, ..., n-1\}$ in the linear extension uniformly at random.
2. Switch the elements at positions $p$ and $p+1$ if possible (as determined by the partial order); else do nothing.

**Exercise:** Show that this Markov chain is irreducible, with diameter $\leq \binom{n}{2}$ (i.e., any linear extension can be reached from any other by a sequence of at most $\binom{n}{2}$ adjacent transpositions).

Make the chain lazy (and hence aperiodic) by adding self-loops of $p = \frac{1}{2}$ at each step; this ensures ergodicity. The Markov chain is symmetric, so the stationary distribution $\pi$ is uniform.

Now we would like to analyze the mixing time of this Markov chain. Since it is difficult to think about a coupling for an arbitrary pair of linear orders, we will introduce a convenient premetric and use path coupling.

**Premetric:** Call two linear orders *adjacent* if they differ by exactly one transposition (see, e.g., figure 7.2).

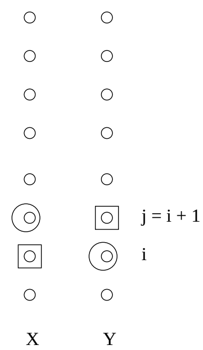


Figure 7.3: Illustration of the case when $j = i + 1$.

For an adjacent pair $X, Y$, define $d(X, Y) = |j - i|$, where $j$, $i$ are the positions where the states $X$ and $Y$ differ. It is not hard to check (**Exercise**) that every such edge is a shortest path, so that this is indeed a valid premetric. As usual we can extend this premetric to a metric in the obvious way.

**Coupling on adjacent pairs:** Let $i < j$ be the positions where $X$ and $Y$ differ.

If $j > i + 1$ then

- With probability $\frac{1}{2}$, do a self-loop in both $X$ and $Y$.
- Else, pick the same $p$ in $X$ and $Y$.

If $j = i + 1$ (see figure 7.3) then

- With probability $\frac{1}{2(n-1)}$, do a self-loop in $X$ and pick $i$ in $Y$.
- With probability $\frac{1}{2(n-1)}$, do a self-loop in $Y$ and pick $i$ in $X$.
- Else, do the same (either a self-loop or pick the same $p$) in both $X$ and $Y$.

**Analysis of Path Coupling**

There are four cases to check.

*Case 1:* $p \neq i - 1, i, j - 1, j$. No change in the distance between $X$ and $Y$.

*Case 2:* $p = i - 1$ or $p = j$. By symmetry these cases are equivalent, so assume $p = i - 1$.

With probability $\frac{1}{2}$, there is no change (self-loop).

Otherwise, the distance increases by at most 1 (check this!).

So in this case $E[d(X', Y') - d(X, Y)|X, Y] \leq \frac{1}{2}$.

*Case 3:* $p = j - 1$ or $p = i$, and $j \neq i + 1$. Again these cases are equivalent by symmetry, so assume $p = i$.

With probability $\frac{1}{2}$, no change (self-loop).

With probability $\frac{1}{2}$, the distance definitely decreases by 1. To see this, note that since $X(j) = Y(i) = a$ and $X(i) = Y(j) = b$, all the elements between positions $i$ and $j$ in the two orders must be incomparable with $a$

and $b$. Hence when we attempt to switch the elements at $i$ and $i+1$, this operation will always be allowed, and as a result the distance between the two mismatched positions in $X$ and $Y$ will decrease by 1.

So in this case $E[d(X',Y') - d(X,Y)|X,Y] = -\frac{1}{2}$.

*Case 4:* $p = i$ and $j = i+1$. By the definition of the path coupling, in this case the distance always decreases to zero. Hence $E[d(X',Y') - d(X,Y)|X,Y] = -1$.

Combining the above four cases, and using the fact that all choices of $p$ are equally likely, we have $E[d(X',Y')|X,Y] \leq d(X,Y)$ for any pair $X$ and $Y$. Thus equation (7.1) holds with $\alpha = 0$. We could now proceed to get a lower bound on the probability that the distance decreases, and thus a bound on the mixing time, but we leave this as an exercise and instead improve the Markov chain slightly so that (7.1) holds with some $\alpha > 0$.

**Exercise:** Use the above analysis to show that the mixing time is $O(n^5)$.

We now improve the Markov chain so that in the above analysis we get a negative drift in the distance. The intuition is that, rather than choosing the switch position $p$ uniformly, the chain should choose positions towards the center of the ordering with greater probability, since such switches are less likely to be permitted.

**Modified Markov Chain:** Same as before, except that instead of picking a position $p \in \{1, 2, \ldots, n-1\}$ u.a.r., we pick $p$ with probability $\frac{Q(p)}{Z}$, where $Z = \sum_p Q(p)$ and $Q(p) > 0$. Notice that this Markov chain is still symmetric (why?), so the stationary distribution is still uniform.

The path coupling argument goes through exactly as before, except that different moves have different weights, so the expected change in distance becomes

$$E[d(X',Y') - d(X,Y)|X,Y] \leq \frac{1}{2Z}[Q(i-1) + Q(j) - Q(i) - Q(j-1)].$$

Since we want (7.1) to hold with $\alpha > 0$, our goal is to choose $Q$ such that $Q(i-1)+Q(j)-Q(i)-Q(j-1) \leq -\alpha(j-i)$.

Now notice that $Q(i-1) + Q(j) - Q(i) - Q(j-1)$ represents a difference of discrete derivatives at $j$ and $i$. Since we want this to be a linear function of $|j-i|$, we should take $Q(i)$ to be a quadratic function of $i$. Setting $Q$ to be quadratic, and choosing its coefficients to maximize the value of $\alpha$, leads us to the choice $Q(p) = p(n-p)$. This yields $Z = n\sum_{p=1}^{n-1} p - \sum_{p=1}^{n-1} p^2 = \frac{1}{6}(n^3 - n) \leq \frac{n^3}{6}$.

Plugging this $Q$ into the above expectation, we get:

$$\begin{aligned} E[d(X',Y') - d(X,Y)|X,Y] &\leq \frac{1}{2Z}[Q(i-1) + Q(j) - Q(i) - Q(j-1)] \\ &= -\frac{2}{2Z}(j-i) \\ &\leq -\frac{6}{n^3}(j-i). \end{aligned}$$

Thus (7.1) holds with $\alpha = \frac{6}{n^3}$. Since also the maximum distance is $D = \binom{n}{2}$, we conclude that the mixing time is $\tau_{mix} = O(\alpha^{-1}\log(D)) = O(n^3 \log n)$.

David Wilson [Wi] has shown that the same bound of $O(n^3 \log n)$ holds for the original Markov chain (in which $Q$ is uniform). It is apparently still an open question whether the above non-uniform choice of $Q$ actually improves the mixing time (as opposed to just easing the analysis). Wilson also provides a lower bound showing that $O(n^3 \log n)$ is in fact tight for the original Markov chain. Interestingly, this lower bound holds even for the degenerate case where $\prec$ is the empty partial order (i.e., no constraints). In this case the Markov chain just corresponds to card shuffling by random adjacent transpositions.

## References

[BW91] G. Brightwell and P. Winkler, "Counting linear extensions," *Order* **8**, 1991, pp. 225–242.

[BD98] R. Bubley and M. E. Dyer, "Faster random generation of linear extensions," *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, 1998, pp. 350–354.

[Wi01] David Wilson, "Mixing times of lozenge tiling and card shuffling Markov chains," preprint, Microsoft Research, 2001. To appear in *Annals of Applied Probability*, 2002.