Randomness and Computation

or, “Randomized Algorithms”

Mary Cryan

School of Informatics
University of Edinburgh

Markov chain and mixing times

On Tuesday we saw an example of a *Markov chain* on the state space Ω_{IS} of *Independent Sets* of a given graph $G = (V, E)$.

We showed that that Markov chain had a unique *stationary distribution* over the state space Ω_{IS}, and that this stationary distribution was the *uniform distribution* on Ω_{IS} (in the limit, as we run the chain for many many steps, we converge to a distribution where each individual IS is equally likely).

We showed a similar result for our contingency tables chain in cwk2.

However, for practical use (to draw a random sample) we need to know *How many steps of the Markov chain do we need to take before we are close to uniform?*

Mixing time

Definition (Definition 12.1)

Let D_1 be a probability distribution over the (countable) state space Ω, and let D_2 be another probability distribution over the same state space. We define the *variation distance* between D_1 and D_2 as

$$\|D_1 - D_2\| \;=\; \frac{1}{2}\sum_{x\in\Omega} |D_1(x) - D_2(x)|.$$

Note variation distance is sometimes defined without the $\frac{1}{2}$. I am being consistent with the book here.

When we run the Markov chain M starting from some fixed $x \in \Omega$, the *distribution of the "current state" after t steps* is the x-th row of M^t, often written as $M^t[x, \cdot]$.
We will want to know how large we need to take t in order to have the *variation distance* of $M^t[x, \cdot]$ within ϵ of the stationary distribution.

Mixing time

Definition (Definition 12.2)

Let M be an ergodic Markov chain over the state space Ω and let $\bar{\pi}$ be its stationary distribution. We define $\Delta_x(t), \Delta(t)$ as

$$\Delta_x(t) \;=\; \|M^t[x,\cdot] - \bar{\pi}\|, \qquad \Delta(t) \;=\; \max_{x\in\Omega} \Delta_x(t).$$

We also define

$$\tau_x(\epsilon) \;=\; \min\{t : \Delta_x(t) \le \epsilon\}, \qquad \tau(\epsilon) \;=\; \max_{x\in\Omega} \tau_x(\epsilon).$$

When we have an upper-bound for $\tau(\epsilon)$ (usually in terms of $\ln(\frac{1}{\epsilon})$ and a size parameter of our state space), we call $\tau(\cdot)$ the *mixing time*.

For any *ergodic* Markov chain, $\|M^{t+k}[x,\cdot] - \bar{\pi}\| \;\le \|M^t[x,\cdot] - \bar{\pi}\|$ for any $k \ge 1$ (Section 12.3 of book). Hence we stay within ϵ variation distance after $\tau(\epsilon)$ steps have been taken.

Mixing time

- As (theoretical) computer scientists, it is important to us to have sampling algorithms that run *in polynomial time* in the size of description of Ω and in $\ln(\frac{1}{\epsilon})$ - the FPAUS.
- If using a Markov chain, we need to show that its mixing time $\tau(\epsilon)$ is a polynomial function in the size of the description of Ω, and in $\ln(\frac{1}{\epsilon})$.

  If we can show this, the Markov chain is said to be *rapidly mixing* (even if the polynomial has high (constant) exponents :-)).
- There are two main techniques for upper-bounding mixing time: *coupling* (including *path coupling*) and *conductance/canonical paths*.
- Coupling gives nice tight bounds when we can design a coupling that achieves our result. Canonical paths/conductance gives worse bounds, but it tends to work on a larger pool of Markov chains.

Path Coupling

A simpler version of coupling called *path coupling* only requires the coupling to be designed for similar states of the Markov chain.

Lemma (Bubley and Dyer 1997)

Let M be a Markov chain on Ω and let d be an integer-valued metric on $\Omega \times \Omega$ taking values in $\{0, 1, \ldots, D\}$ for some D. Let S be a subset of $\Omega \times \Omega$ such that for all $(X(t), Y(t)) \in \Omega \times \Omega$ there is a path

$$X_0 = X(t), X_1, \ldots, X_\ell = Y(t)$$

such that $(X_i, X_{i+1}) \in S$ for all $i, 0 \leq i < \ell$ and $d(X(t), Y(t)) = \sum_{i=0}^{\ell-1} d(X_i, X_{i+1})$. Suppose we have a coupling $(X, Y) \to (X', Y')$ of M on all pairs in S such that

$$\mathrm{E}[d(X', Y')] \leq \beta d(X, Y).$$

Path Coupling

Lemma (Bubley and Dyer 1997 (cont'd))

Then if $\beta < 1$, *the mixing time* $\tau(\epsilon)$ *of M satisfies*

$$\tau(\epsilon) \;\leq\; \frac{\ln(D\epsilon^{-1})}{1-\beta}.$$

If $\beta = 1$ *and there is some* $\alpha > 0$ *such that* $\Pr[d(X', Y') \neq d(X, Y)] \geq \alpha$ *for all* $(X, Y) \in \Omega \times \Omega$, *then*

$$\tau(\epsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \lceil \ln(\epsilon^{-1}) \rceil.$$

- This version of coupling simplifies matters over standard coupling because we only have to get the coupling to work for pairs of similar states.
- To get an FPAUS have to show that β (or α) are “inverse polynomial" in size of the state space description.

New Markov chain for Independent sets

It is very difficult to show that the Markov chain for IS of Lecture 18 is rapid mixing, despite it's simplicity. Consider a new Markov chain M for Independent sets:

Algorithm GENERATEIS2$(G = (V, E))$

1. Start with an arbitrary IS X_0
2. **for** $i \leftarrow 0$ **to** "whenever"
3. Choose $e = (u, v)$ uniformly at random from E.
4. **with prob.** $\frac{1}{3}$, **set**
5. $X_{i+1} \leftarrow X_i \setminus \{u, v\}$
6. **with prob.** $\frac{1}{3}$, **set**
7. $X_{i+1} \leftarrow (X_i \setminus \{u\}) \cup \{v\}$ **if** this is an IS, **else** $X_{i+1} \leftarrow X_i$
8. **with prob.** $\frac{1}{3}$, **set**
9. $X_{i+1} \leftarrow (X_i \setminus \{v\}) \cup \{u\}$ **if** this is an IS, **else** $X_{i+1} \leftarrow X_i$

New Markov chain for Independent sets (coupling)

We will design a coupling for this new Markov chain, then apply the path coupling result of Bubley/Dyer.

For any two ISs, X, Y, we define $d(X, Y) = |X \oplus Y|$ (recall $X \oplus Y$ is the difference set of X, Y). We can construct a sequence of states of length $d(X, Y)$ connecting X to Y in our new Markov chain *exactly* the same way as we showed irreducibility of the Lecture 18 chain.

We define $S = \{(X, Y) : |X \oplus Y| = 1\}$.

For any pair of states X, Y (whether $(X, Y) \in S$ or not) we say a vertex $v \in V$ is *bad* if $v \in X \oplus Y$, and otherwise we say v is *good*.

Now consider X, Y such that $Y = X \cup \{x\}$ (for some $x \notin X$). These of course are the pairs of S.

We will show that applying the *näive coupling* (same edge and transition chosen for X and Y), that if the max-degree of G is 4, that

$$\mathrm{E}[d(X', Y')] \leq d(X, Y).$$

$(\beta = 1)$

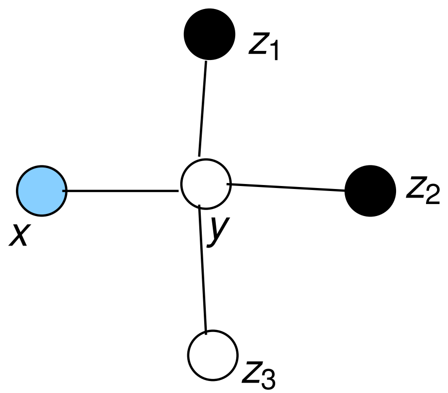
New Markov chain for Independent sets (coupling)

- If the edge e chosen has the “difference vertex" x as one of its endpoints, then we are guaranteed that X' will be equal to Y' (the same transitions are possible in X and Y, so we can “couple" them exactly, making X' identical to Y').
- If *neither* endpoint of the edge e chosen is adjacent to x, then the surrounding neighbourhoods of u, v are identical in X and Y, and hence can couple our actions exactly. However we will have $d(X', Y') = 1$ after this (since x won’t change).
- If the edge e chosen is *adjacent* to x, then there is a possibility that $d(X', Y')$ *could* increase on line 6,7 or 8,9 (since the transition might succeed in X but not in Y).

New Markov chain for Independent sets (coupling)

Consider $y \in V$, y a neighbour of x. Three cases. We will show the expected contribution to $d(X', Y')$ from y is 0, for each case.

case (a): y has *two or more* neighbours in the independent set X (and three or more in Y).

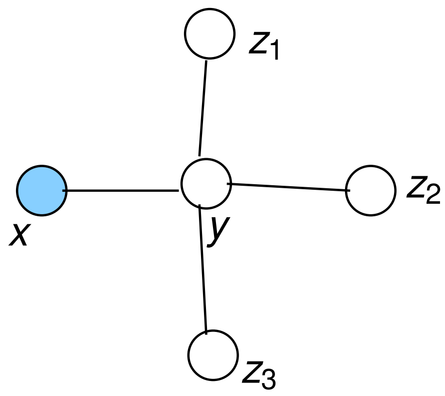


Then for this y, we have two adjacent neighbours in the IS for *both* X and Y. If we choose (y, z) for any of the neighbours (y, z), the move adding y is blocked. Hence y never changes, and these moves contribute 0 extra to $d(X', Y')$.

New Markov chain for Independent sets (coupling)

Consider $y \in V$, y a neighbour of x. Three cases.

case (b): y has *no* neighbours in the independent set X (and just one, x, in Y).

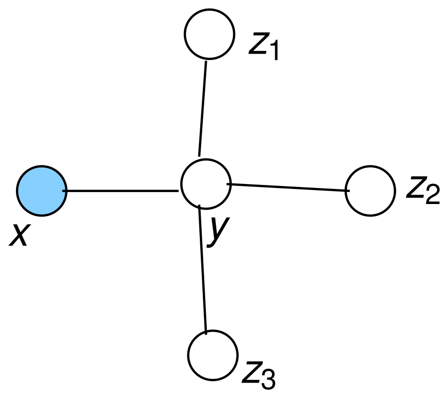


In this case, if we try $e = (y, z)$ for any $z \in Nbd(y) \setminus \{x\}$, then with probability $\frac{1}{3}$ we attempt the move to add y. This will *definitely fail* in X (x blocks it) but will definitely succeed in Y (no neighbours in the IS). So there is a contribution of $1.\frac{1}{3}$ to $d(X', Y')$ for each (y, z) adjacent to y, $z \neq x$.

New Markov chain for Independent sets (coupling)

Consider $y \in V$, y a neighbour of x. Three cases.

case (b) cont'd: y has *no* neighbours in the independent set X (and just one, x, in Y).



There are at most 4 neighbours for y, so at most 3 non-x neighbours, hence we have an extra expected contribution of 1 to $d(X', Y')$ from ys adjacent edges that are not (x, y).

However, we might alternatively choose $e = (x, y)$, and then we reduce $d(X', Y')$ by 1 with probability 1.

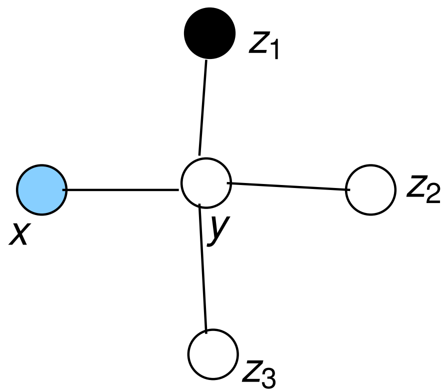
Hence the net contribution of edges adjacent to y to $d(X', Y')$ is 0.

New Markov chain for Independent sets (coupling)

Consider $y \in V$, y a neighbour of x. Three cases.

case (c): y has *exactly one* neighbour in the independent set X (and two in Y).

In this case, if we try $e = (x, y)$ as our edge, then we reduce $d(X', Y')$ by 1 with probability only $\frac{2}{3}$. This is because we can either drop $\{x, y\}$ identically in X, Y, and also can drop y, add x identically in X, Y, achieving “coupling" ($d(X', Y') = 1$).



However if we try to drop x, add y, this will *fail* in *both* X and Y, keeping $d(X', Y')$ as 1.
So overall on the edge (x, y) we have a $-\frac{2}{3}$ contribution to alter $d(X', Y')$.

New Markov chain for Independent sets (coupling)

Consider $y \in V$, y a neighbour of x. Three cases.

case (c) cont'd: y has *exactly one* neighbour in the independent set X (and two in Y).

For (y, z), z being the neighbour in X, we can cause *both* y and z to become bad if we choose (y, z) and attempt to add y and drop z (prob. $\frac{1}{3}$). This will succeed in X, but fail in Y. adding 2 (with probability $\frac{1}{3}$) extra to $d(X', Y')$.
For the other two options for (y, z), the move succeeds in both, adding 0 extra to $d(X', Y')$. Also the moves on (y, z') for $z' \notin Y$ have identical actions on X, Y, with 0 extra contribution to $d(X', Y')$.

Hence in case (c), we also have $d(X', Y') \leq (X, Y)$.

New Markov chain for Independent sets (coupling)

We have shown that for each $y \in Nbd(x)$, the expected contribution to $d(X', Y') - d(X, Y)$ from "edges adjacent to y" is 0.

We know that moves on edges with no endpoint in $Nbd(x)$ have 0 contribution to $d(X', Y') - d(X, Y)$.

Hence we have shown

$$\mathrm{E}[d(X', Y')] \leq d(X, Y),$$

giving $\beta = 1$ for path coupling on our S.

We can easily show that $\alpha \geq \frac{1}{3m}$ for our chain.
Hence Bubley-Dyer implies that the Markov chain can be used as an FPAUS for independent sets (when max degree of G is 4).

Reading and Doing

Reading:

- Sections 12.1 and 12.6 of the book relate to this lecture. Note that the argument in 12.6 ends by showing that the coupling on the S pairs can be extended to a coupling (which is given to us by Bubley/Dyer).
- Section 12.2 describes standard coupling (worth a read if you're interested) and gives the formal definition of "a coupling" (which I left out of these slides). Section 12.3 shows that variation distance is non-increasing with t for ergodic chains.

Doing:

- Show that today's new Markov chain on slide 8 *also* has the uniform distribution on Independent sets of G, in a similar way to how we did the original Markov chain on Tuesday.
- Can you think about a path coupling argument for contingency tables with two rows? (tricky)