Coupling methods

Teemu Hirsimäki

`teemu.hirsimaki@hut.fi`

A seminar talk based on

M. Jerrum, "Mathematical foundations of the Markov chain Monte Carlo method." In: M. Habib et al (eds.), *Probabilistic Methods for Algebraic Discrete Mathematics*. Springer-Verlag, Berlin, 1998.

What is coupling used for?

- The basic idea:
  - A random sample from a complex distribution is needed (e.g. a random colouring of a graph)
  - Create a clever Markov chain and run it long enough.
- To estimate how long is enough:
  - Run two copies of the chain **coupled** together.
  - Show that the chains **coalesce** with some probability.
  - The **Coupling Lemma** gives the mixing time for the original chain.

What is coupling?

- Suppose M is a countable and ergodic Markov chain with transition probabilities $P_M(\cdot|\cdot)$.
- Coupling $\big((X_t, Y_t) : t \in \mathbb{N}\big)$ is a Markov chain satisfying the following equations:

$$P(x'|x, y) = P_M(x'|x) \quad (1)$$

$$P(y'|x, y) = P_M(y'|y) \quad (2)$$

Examples of coupled chains

- Walking randomly among 2-colourings of two vertices by selecting a random vertex and a random color.
  - Running the chains separately satisfies Equations 1 and 2, but does not help much.
  - Instead, make the same change in both chains. The equations are still satisfied.
- Equations 1 and 2 allow that $P(x', y'|x, y) \neq P(x'|x)P(y'|y)$.
  - This important property makes it possible to create clever couplings, so that (X_t) and (Y_t) are forced to coalesce!

The Coupling Lemma

- The idea: If $P(X_t \neq Y_t)$ gets small fast enough when t grows, we get also a bound for the mixing time of the original chain.
- **Formally**: Suppose that M is a countable, ergodic Markov chain with transition probabilities $P(\cdot|\cdot)$, and let $\big((X_t, Y_t) : t \in \mathbb{N}\big)$ be a coupling, i.e. a Markov process satisfying (1) and (2). Suppose further that $t : (0,1] \to \mathbb{N}$ is a function such that $P(X_{t(\epsilon)} \neq Y_{t(\epsilon)}) \leq \epsilon$ for all $\epsilon \in (0,1]$, uniformly over the choice of initial state (X_0, Y_0). Then the mixing time $\tau(\epsilon)$ of M is bounded above by $t(\epsilon)$.

The mixing time

- Total variation distance is used to measure the distance between two distributions:

$$D_{tv}(\pi, \pi') := \max_{A \subseteq \Omega} \left|\pi(A) - \pi'(A)\right| = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)|$$

- The distance to the stationarity distribution:

$$\delta_x(t) := D_{tv}(P^t(x, \cdot), \pi)$$

- The mixing time of the chain:

$$\tau(\epsilon) = \min\left\{t : \delta_x(t') \leq \epsilon \text{ for all } t' \geq t\right\}$$

Toy example: An empty graph

- Colouring an empty graph $G = (V, \emptyset)$ with colours from Q.
- Each state of the Markov process is a valid colouring.
- The transition $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ in the coupling:
  1. Select a vertex $v \in V$, uniformly at random (u.a.r.)
  2. Select a colour $c \in Q$, u.a.r.
  3. Recolour vertex v in X_t and Y_t with c to get X_{t+1} and Y_{t+1}.

Toy example: Deriving the bound

- To utilise the Coupling Lemma, we need $t(\epsilon)$ for each ϵ to force $P(X_t \neq Y_t) \leq \epsilon$.
- Let D_t the set of vertices on which X_t and Y_t differ:

$$D_t = \{v \in V : X_t(v) \neq Y_t(v)\}$$

- Now $|D_t| > 0 \Leftrightarrow X_t \neq Y_t$, i.e.

$$P(X_t \neq Y_t) = P\big(|D_t| > 0\big)$$

Toy example: Deriving the bound

- How does D_t change in transitions?
  - $D_{t+1} = D_t \setminus v$, if the selected v is in D_t.
  - $D_{t+1} = D_t$, otherwise.
- To bound $P(|D_t| > 0)$, we can use the expected value. Since v is selected u.a.r.

$$E\big(|D_{t+1}| \,\big|\, D_t\big) = \left(1 - \frac{1}{n}\right)|D_t|$$

$$E\big(|D_t| \,\big|\, D_0\big) = \left(1 - \frac{1}{n}\right)^t |D_0|$$

Toy example: The bound

- Since $|D_t|$ is a non-negative integer random variable, Markov's Inequality tells us that

$$\begin{aligned} P\big(|D_t| > 0 \,\big|\, D_0\big) &\leq E\big(|D_t| \,\big|\, D_0\big) \\ &\leq n\left(1 - \frac{1}{n}\right)^t \\ &\leq n e^{-t/n} \end{aligned}$$

- Now, if $t \geq n \ln n\epsilon^{-1}$, we have $P(X_t \neq Y_t) < \epsilon$, and the Coupling Lemma says, that the mixing time of the Markov process $\tau_x(\epsilon) \leq n(\ln n + \ln \epsilon^{-1})$, independent of the starting state x.

Discussion

- Questions so far?
- What if want colourings from a distribution which is not uniform?
- What goes wrong if couple the chains more strongly than allowed by (1) and (2)?

Real example: Colouring a graph

- Colouring a low-degree graph of maximum degree Δ with $q \geq 2\Delta + 1$ colours.
- The transitions must be chosen more carefully.
  - More constraints are needed, because we want to walk among valid colourings.
  - But still, X_{t+1} must depend only on X_t.
- After selecting a common vertex v randomly, the colours c_X and c_Y are selected from a clever joint distribution.

Real example: Selecting colours

- Let Q be the set of colours, and denote by $\Gamma(v) \subseteq V$ the set of all neighbours of v in the graph G. Let $X_t(U) = \{X_t(u) : u \in U\}$, $q_X = |Q \setminus X_t(\Gamma(v))|$, $q_Y = |Q \setminus Y_t(\Gamma(v))|$, and $q_{XY} = |Q \setminus (X_t(\Gamma(v)) \cup Y_t(\Gamma(v)))|$.
- Select a vertex $v \in V$ u.a.r., and choose colours c_X and c_Y **jointly** so that the following conditions are met:
  - c_X is selected u.a.r. from the set $Q \setminus X_t(\Gamma(v))$
  - c_Y is selected u.a.r. from the set $Q \setminus Y_t(\Gamma(v))$
  - and the joint sample space satisfies

$$P(c_X = c_Y) = \frac{q_{XY}}{\max\{q_X, q_Y\}} \tag{3}$$

Real example: Deriving the bound

- Let A_t be the set of vertices on which the colourings agree, and D_t the set on which the colourings disagree.
- We know that $|D_{t+1}| - |D_t| \in \{-1, 0, 1\}$, and want to show that 1 is more probable than -1.
- Let's consider first the probability that $|D_{t+1}| = |D_t| + 1$. We know that v is in A_t, and the colours are not equal. Denote by $d'(v)$ the number of edges between A and D. Then we have

$$q_X - q_{XY} \leq d'(v) \tag{4}$$

$$q_Y - q_{XY} \leq d'(v) \tag{5}$$

$$q_{XY} \geq q - \Delta - d'(v) \tag{6}$$

where q is the number of colours.

Starting from (3) and using (4–6) we get

$$P(c_X = c_Y) = \frac{q_{XY}}{\max\{q_X, q_Y\}} \geq \frac{q_{XY}}{d'(v) + q_{XY}} \geq 1 - \frac{d'(v)}{q - \Delta}$$

$$P\big(|D_{t+1}| = |D_t| + 1\big) \leq \frac{1}{n} \sum_{v \in A} \frac{d'(v)}{q - \Delta} = \frac{m'}{(q - \Delta)n} \tag{7}$$

where $m' = \sum_{v \in D} d'(v) = \sum_{v \in A} d'(v)$, i.e. the edges between A and D.

- Similarly, for the case $|D_{t+1}| = |D_t| - 1$ we know that v must be in D and the selected colours must be equal. In this case we have the following inequalities:

$$\begin{aligned} q_X - q_{XY} &\leq \Delta - d'(v), \\ q_Y - q_{XY} &\leq \Delta - d'(v), \\ q_{XY} &\geq q - 2\Delta + d'(v) \end{aligned}$$

and we get

$$\begin{aligned} P(c_X = c_Y) &= \frac{q_{XY}}{\max\{q_X, q_Y\}} \\ &\geq \frac{q_{XY}}{\Delta - d'(v) + q_{XY}} \\ &\geq \frac{q - 2\Delta}{q - \Delta} + \frac{d'(v)}{q - \Delta'} \end{aligned}$$

Then we get the probability that $|D_t|$ decreases

$$P(|D_{t+1}| = |D_t| - 1) \geq \frac{1}{n} \sum_{v \in D} \left(\frac{q - 2\Delta}{q - \Delta} + \frac{d'(v)}{q - \Delta} \right)$$
$$= \frac{q - 2\Delta}{(q - \Delta)n} |D| + \frac{m'}{(q - \Delta)n} \tag{8}$$

- If we define

$$a = \frac{q - 2\Delta}{(q - \Delta)n} \qquad b = b(m') = \frac{m'}{(q - \Delta)n},$$

we can see that the size of the set D_t tends to decrease, because

$$P(|D_{t+1}| = |D_t| + 1) \leq b$$
$$P(|D_{t+1}| = |D_t| - 1) \geq a|D_t| + b$$

and we know that $a > 0$ since we assumed $q \geq 2\Delta + 1$.

Real example: The bound

- Finally, using (7) and (8) we get

$$\begin{aligned} E\big(|D_{t+1}| \,\big|\, D_t\big) &\le b(|D_t|+1) + (a|D_t|+b)(|D_t|-1) \\ &\quad + (1-a|D_t|-2b)|D_t| \\ &= (1-a)|D_t| \end{aligned}$$

  and

$$E\big(|D_t| \le (1-a)^t |D_0|\big) \le n(1-a)^t.$$

- Now because $|D_t|$ is a non-negative integer random variable, $P(|D_t| \neq 0) \le n(1-a)^t \le ne^{-at}$. So, $P(|D_t| \neq 0) \le \epsilon$, provided $t \ge a^{-1} \ln(n\epsilon^{-1})$.

About coupling methods

- The presented (direct) coupling has been used also to other problems.
  - Counting independent sets in a low-degree graph.
  - Estimating the volume of a convex body.
- The hardest thing is to design the coupling.
- The book presents two newer techniques
  - Path Coupling: coupling is defined only on pairs of adjacent states.
  - Exact sampling can be done with Coupling From the Past (CFTP).

Discussion

- Questions?
- Proof of the Coupling Lemma?
- What happens if we do not avoid invalid colourings?