Coupling
AMS Short Course

David A. Levin, *U. Oregon*

January 2010

Distance

If μ and ν are two probability distributions on a set Ω, then the **total variation** distance between μ and ν is

$$d_{\mathrm{TV}}(\mu, \nu) := \max_{A \subset \Omega} |\mu(A) - \nu(A)|$$
$$= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| .$$

Example. Let $\Omega = \{0, 1\}$, and set

$$\mu_p(0) = 1 - p, \quad \mu_p(1) = p .$$

Then

$$d_{\mathrm{TV}}(\mu_p, \mu_q) = \frac{1}{2}\Big[|(1 - p) - (1 - q)| + |p - q|\Big] = |p - q| .$$

Coupling

A *coupling* between two probability distributions μ and ν is a pair of random variables (X, Y) such that

- X and Y are defined on a common probability space,
- X has distribution μ, and
- Y has distribution ν.

Example. Let X_p be a random bit with

$$\mathbb{P}(X_p = 1) = p, \quad \mathbb{P}(X_p = 0) = 1 - p\,. \tag{1}$$

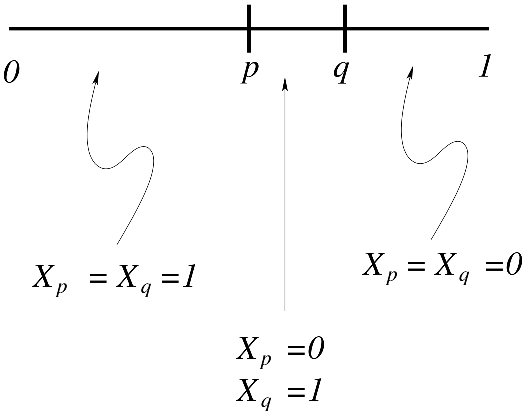
We can couple X_p and X_q as follows: Let U be a uniform random variable on $[0, 1]$, i.e., for $0 \leq a < b \leq 1$,

$$\mathbb{P}(a < U \leq b) = b - a\,.$$

Define

$$X_p = \begin{cases} 1 & \text{if } 0 < U \leq p \\ 0 & \text{if } p < U \leq 1 \end{cases}, \quad X_q = \begin{cases} 1 & \text{if } 0 < U \leq q \\ 0 & \text{if } q < U \leq 1 \end{cases}.$$

The random variable U *serves as a common source of randomness* for both X_p and X_q.


0
p
q
1
$X_p = X_q = 1$
$X_p = X_q = 0$
$X_p = 0$
$X_q = 1$

Example. Another coupling of μ_p with μ_q: Take X'_p and X'_q to be independent of each other.

Note that in this coupling,

$$\mathbb{P}(X'_p \neq X'_q) = p(1-q) + (1-p)q = p + q - 2pq\,.$$

In the coupling using the common uniform random variable,

$$\mathbb{P}(X_p \neq X_q) = |p - q|\,.$$

Assuming (without loss of generality) that $p < q$,

$$\mathbb{P}(X'_p \neq X'_q) - \mathbb{P}(X_p \neq X_q) = 2p(1-q) \geq 0$$

that is,

$$\mathbb{P}(X'_p \neq X'_q) \geq \mathbb{P}(X_p \neq X_q)\,.$$

Proposition

If μ and ν are two probability distributions, then

$$d_{\mathrm{TV}}(\mu,\nu) = \min_{(X,Y)\ \textit{couplings}} \mathbb{P}(X \neq Y)\,.$$

Example. For coin-tossing distributions μ_q and μ_p,

$$d_{\mathrm{TV}}(\mu_p,\mu_q) = \frac{1}{2}\left[|(1-p)-(1-q)| + |p-q|\right] = |p-q|,$$

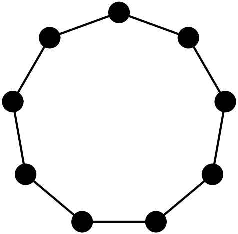
so the coupling using the uniform variable is *optimal*.

Couplings of Markov Chains

Let P be a transition matrix for a Markov chain. A coupling of a P-Markov-chain started at x and a P-Markov-chain started at y is a sequence $\{(X_n, Y_n)\}_{n=0}^{\infty}$ such that

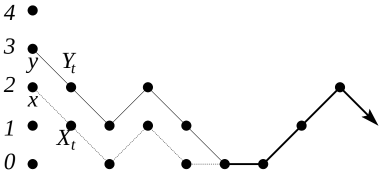
- all variables X_n and Y_n are defined on the same probability space,
- $\{X_n\}$ is a P-Markov-chain started at x, and
- $\{Y_n\}$ is a P-Markov-chain started at y.

Example: The lazy random walk on the n-cycle.



- This chain remains at its current position with probability $1/2$, and moves to each of the two adjacent site with probability $1/4$.
- Can couple the chains started from x and y as follows:
  - Flip a fair coin to decide if the X-chain moves or the Y-chain moves,
  - Move the selected chain to one of its two neighboring sites, chosen with equal probability.
- Both the x-particle and the y-particle are performing lazy simple random walks on the n-cycle.

- Let $(X_t, Y_t)_{t=0}^{\infty}$ be a coupling of a P-chain started from x and a P-chain started at y.
- Let
$$\tau := \min\{t \geq 0 \ : \ X_t = Y_t\} \, .$$
The coupling can always be redefined so that
$$X_t = Y_t \text{ for } t \geq \tau,$$
So, let us assume this.



- The pair (X_t, Y_t) (for given t) is a coupling of $P^t(x, \cdot)$ and $P^t(y, \cdot)$.

- Since X_t has distribution $P^t(x,\cdot)$ and Y_t has distribution $P^t(y,\cdot)$, using the coupling characterization of total variation distance,

$$\mathbb{P}(\tau > t) = \mathbb{P}(X_t \neq Y_t) \geq d_{\mathrm{TV}}(\mathbb{P}^t(x,\cdot), \mathbb{P}^t(y,\cdot))\,.$$

- Combined with the inequality

$$d_{\mathrm{TV}}(P^t(x,\cdot),\pi) \leq \max_{y\in\Omega} d_{\mathrm{TV}}(P^t(x,\cdot), P^t(y,\cdot))\,,$$

  if there is a coupling (X_t, Y_t) for every pair of initial states (x, y), then this shows that

$$\begin{aligned} d(t) = \max_{x\in\Omega} d_{\mathrm{TV}}(P^t(x,\cdot),\pi) &\leq \max_{x,y} d_{\mathrm{TV}}(P^t(x,\cdot), P^t(y,\cdot)) \\ &\leq \max_{x,y} \mathbb{P}_{x,y}(\tau > t)\,. \end{aligned}$$

Mixing for lazy random walk on the n-cycle

- Use the coupling which selects at each move one of the "particles" at random; the chosen particle is equally likely to move clockwise as counter-clockwise.
- The clockwise difference between the particles, $\{D_t\}$, is a simple random walk on $\{0, 1, \ldots, n\}$.
- When $D_t \in \{0, n\}$, the two particles have collided.
- If τ is the time until a simple random walk on $\{0, 1, \ldots, n\}$ hits an endpoint when started at k, then

$$\mathbb{E}_k \tau = k(n-k) \leq \frac{n^2}{4}.$$

RW on n-cycle, continued

- By Markov's inequality,

$$\mathbb{P}(\tau > t) \leq \frac{\mathbb{E}\tau}{t} \leq \frac{n^2}{4t}.$$

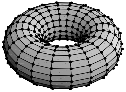
- Using the coupling inequality,

$$d(t) \leq \max_{x,y} \mathbb{P}(\tau > t) \leq \frac{n^2}{4t}.$$

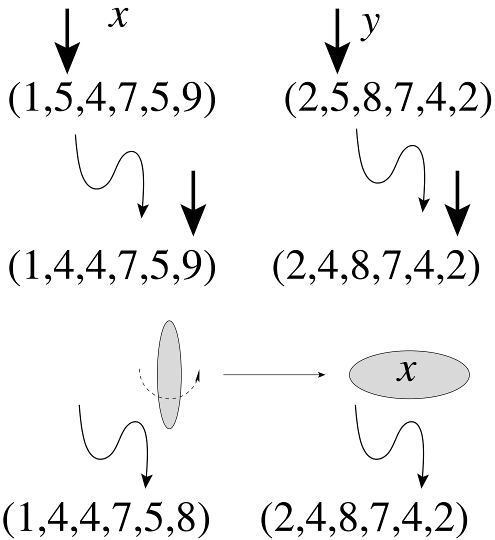
- Taking $t \geq n^2$ yields $d(t) \leq 1/4$, whence

$$t_{\mathrm{mix}} \leq n^2.$$

Random Walk on d-dimensional Torus



- $\Omega = (\mathbb{Z}/n\mathbb{Z})^d$. The walk remains at current position with probability $1/2$.
- Couple two particles as follows:
  - Select among the d coordinates at random.
  - If the particles agree in the selected coordinate, move the walks together in this coordinate. Thus both walks together either make a clockwise move, a counterclockwise move, or remain put.
  - If the particles disagree in the chosen coordinate, flip a coin to decide which walker will move. Move the selected walk either clockwise or counterclockwise, each with probability $1/2$.


x
y
(1,5,4,7,5,9)
(2,5,8,7,4,2)
(1,4,4,7,5,9)
(2,4,8,7,4,2)
x
(1,4,4,7,5,8)
(2,4,8,7,4,2)

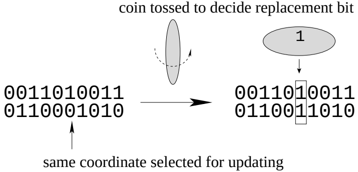
- Consider the clockwise difference between the i-th coordinate of the two particles. It moves at rate $1/d$, and when it does move, it performs simple random walk on $\{0, 1, \ldots, n\}$, with absorption at 0 and n. Thus the expected time to couple the i-th coordinate is bounded above by $dn^2/4$.
- Since there are d coordinates, the expected time for all of them to couple is not more than

$$d \times d\frac{n^2}{4} = \frac{d^2n^2}{4}\,.$$

- By the coupling theorem,

$$t_{\mathrm{mix}} \leq d^2n^2\,.$$

RW on hypercube



- Consider the lazy random walk on the hypercube $\{0, 1\}^n$. Sites are neighbors if they differ in exactly one coordinate.
- To update the two walks, first pick a coordinate at random. *The same coordinate is used for both walks.*
- Toss a coin to determine if the bit at the chosen coordinate is replaced by a 1 or a 0. *The same bit is used for both walks.*
- No matter the initial positions of the two walks, when every coordinate has been selected, the two walks agree.
- Reduces to a "coupon collector's" problem: how many times must a coordinate be drawn at random before every coordinate is chosen?

Coupon collector

- Let $A_k(t)$ be the event that the k-th coupon has *not* been collected by time t.
- Observe
$$\mathbb{P}(A_k(t)) = \left(1 - \frac{1}{n}\right)^t \le e^{-t/n} .$$
- Consequently,
$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k(t)\Big) \le \sum_{k=1}^{n} e^{-t/n} = ne^{-t/n} .$$
- In other words, if τ is the time until all coupons have been collected,
$$\mathbb{P}(\tau > n\log n + cn) = \mathbb{P}\left(\bigcup_{k=1}^{n} A_k(n\log n + cn)\right) \le e^{-c} .$$

Returning to the hypercube,

$$d(n \log n + cn) \leq \mathbb{P}(\tau > n \log n + cn) \leq e^{-c},$$

whence

$$t_{\mathrm{mix}}(\epsilon) \leq n \log n + n \log(1/\epsilon)\,.$$

Contractions

Suppose

- there is a metric ρ on Ω with

$$\rho(x, y) \geq \mathbf{1}\{x \neq y\}$$

- and for any two states x, y, there is a coupling (X, Y) of one step on the chain started from x with one step started from y satisfying

$$\mathbb{E}_{x,y}(\rho(X, Y)) \leq (1 - \alpha)\rho(x, y)\,.$$

Then we obtain a coupling $(X_t, Y_t)_{t=0}^{\infty}$ such that

$$\mathbb{E}_{x,y}(\rho(X_t, Y_t)) \leq (1 - \alpha)^t \text{diam}(\Omega)\,.$$

- We have $\mathbb{E}_{x,y}\rho(X_t, Y_t) \leq \mathrm{diam}(\Omega)e^{-\alpha t}$
- Thus,

$$d(t) \leq \max_{x,y} \mathbb{P}_{x,y}(\tau > t) = \max_{x,y} \mathbb{P}_{x,y}(\rho(X_t, Y_t) \geq 1)$$
$$\leq \max_{x,y} \mathbb{E}_{x,y}\rho(X_t, Y_t) \leq \mathrm{diam}(\Omega)e^{-\alpha t}\,.$$

- If
$$t \geq \frac{\log(\mathrm{diam}(\Omega))}{\alpha} + \frac{c}{\alpha}$$
then
$$d(t) \leq e^{-c}\,.$$
- In other words,
$$t_{\mathrm{mix}}(\epsilon) \leq \frac{\log(\mathrm{diam}(\Omega))}{\alpha} + \frac{\log(1/\epsilon)}{\alpha}\,.$$

Path coupling

Suppose that Ω has a path-metric: Ω is the vertex-set of a graph, and $\rho(x, y)$ is the graph distance between x and y.

Theorem (Bubley-Dyer)

If, for all x, y such that $\rho(x, y) = 1$ there exists coupling (X_1, Y_1) of one step of the chain started from x with one step started from y satisfying

$$\mathbb{E}_{x,y}\rho(X_1, Y_1) \leq (1 - \alpha)\rho(x, y) = (1 - \alpha),$$

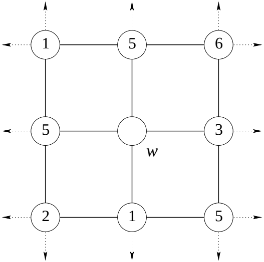
then

$$t_{\mathrm{mix}}(\epsilon) \leq \frac{\log(\mathrm{diam}(\Omega))}{\alpha} + \frac{\log(1/\epsilon)}{\alpha}.$$

Coloring a graph

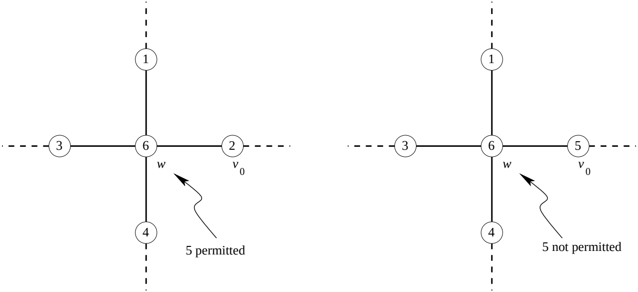
Fix a graph on n vertices.

- A proper q-coloring of a graph is an assignment of the integers $\{1, 2, \ldots, q\}$ to vertices such that adjacent vertices are assigned different values.
- Metropolis: pick a vertex v uniformly at random, and replace the color at vertex v by a random color, *if the color does not create a conflict*.

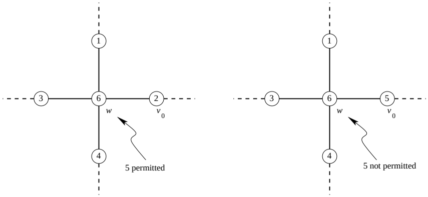


Colors: $\{\cancel{1}, 2, \cancel{3}, 4, \cancel{5}, 6\}$

- Suppose x and y are colorings of a graph differing only at vertex v_0.
- Update both x and y by selecting the same vertex w, and choosing the same color proposal, $K \in \{1, 2, \ldots, q\}$ to recolor w.
- Sometimes w will be rejected in x and accepted in y, or vice versa.



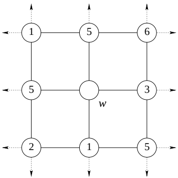
- Situation occurs only if w is a neighbor of v and for 2 out of the q possible color proposals.



- Increase the number of differing vertices with probability

$$\mathbb{P}(\text{select a neighbor of } v_0) \times \frac{2}{q} \leq \frac{2\Delta}{nq},$$

where Δ is the maximal degree of the graph.



Colors: {~~1~~, 2, ~~3~~, 4, ~~5~~, 6}

- Decrease the number of differing vertices with probability

$$\mathbb{P}(\text{select } v_0) \times \mathbb{P}(\text{pick a non-conflicting color}) \geq \frac{1}{n} \times \frac{q - \Delta}{q}.$$

where Δ is the maximal degree of the graph.

The expected distance after one step is:

$$\mathbb{E}_{x,y}\rho(X_1, Y_1) = 1 - \frac{q - \Delta}{nq} + \frac{2\Delta}{nq} = 1 - \frac{1 - 3\Delta/q}{n} .$$

If $q > 3\Delta$, then we have

$$\mathbb{E}_{x,y}\rho(X_1, Y_1) \le 1 - \frac{c(q, \Delta)}{n} .$$

Applying the path-coupling theorem,

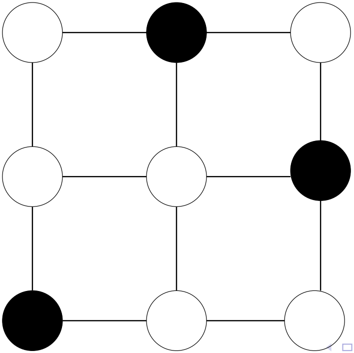
$$t_{\mathrm{mix}}(\epsilon) \le \frac{1}{c} n \log n + \frac{1}{c} n \log(1/\epsilon) .$$

Hardcore model

Fix a graph on n vertices.

- A *hardcore* configuration is a placement of particles on vertices of the graph so that no two particles are adjacent.
  Encode this by σ : vertices $\to \{0, 1\}$,

$$\sigma(v) = \begin{cases} 1 & \text{if } v \text{ is occupied,} \\ 0 & \text{otherwise} . \end{cases}$$



- For every hardcore configuration σ, let

$$\pi_\lambda(\sigma) = \frac{\lambda^{\sum_v \sigma(v)}}{Z(\lambda)}$$

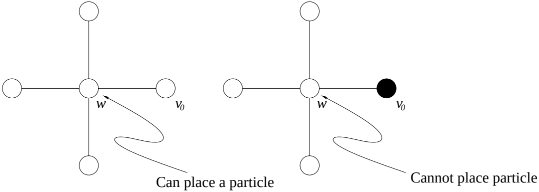
- Want to construct a Markov chain with stationary distribution π: Glauber dynamics.

The rule for updating a configuration σ is as follows:

- Draw a vertex w uniformly at random.
- Flip a coin which land heads with probability $\frac{\lambda}{1+\lambda}$.
- If tails, erase any particle at w.
- If heads, place a particle at w if possible.

This produces a Markov chain with stationary distribution π_λ.

Suppose x and y are two hardcore configurations differing at a single site, say v_0. Thus, y has a particle at v_0, while x does not have a particle at v_0.



- Pick the same vertex to update in x and y, and use the same coin.
- A new disagreement is introduced in the case when
  - a neighbor w of v_0 is selected,
  - v_0 is the only neighbor of w which is occupied, and
  - the coin is heads.
- Thus

$$\mathbb{P}(\text{introduce another disagreement}) \leq \frac{\Delta}{n}\frac{\lambda}{1+\lambda}$$

If v_0 is selected, the disagreement is reduced. We have

$$\mathbb{E}_{x,y}\rho(X_1, Y_1) \leq 1 - \frac{1}{n} + \frac{\Delta}{n}\frac{\lambda}{1+\lambda} = 1 - \frac{1}{n}\left[\frac{1-\lambda(\Delta-1)}{1+\lambda}\right].$$

If $\lambda > (\Delta - 1)^{-1}$, then

$$\mathbb{E}_{x,y}\rho(X_1, Y_1) \leq 1 - \frac{c(\lambda)}{n}$$

By the path-coupling theorem,

$$t_{\mathrm{mix}}(\epsilon) \leq \frac{n}{c(\lambda)}\left[\log n + \log(1/\epsilon)\right].$$