Lect 05

Moment Generating Functions

Def. random variable X. $M_X : \mathbb{R} \to \mathbb{R}$ given by $M_X(t) = \mathbb{E}[e^{tX}]$.

Prop. $M_X(t) = \mathbb{E}[e^{tX}] = 1 + \underline{\mathbb{E}[X]}\, t + \underline{\mathbb{E}[X^2]} \cdot \frac{t^2}{2!} + \underline{\mathbb{E}[X^3]} \frac{t^3}{3!} + \cdots$

moments of x. X幂次的期望.

$$\frac{d^k M_X(t)}{dt^k}(0) = \mathbb{E}\left[\frac{d^k}{dt^k} e^{tX}\Big|_0\right] = \mathbb{E}[x^k e^{0 \cdot x}] = \mathbb{E}[x^k].$$

if $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$ ($\exists \delta > 0$), then X and Y has the same distribution

E.X. $Z \leftarrow N(\mu, \sigma^2)$

$$M_Z(t) = \mathbb{E}[e^{Zt}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \cdots$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

sum of two independent Gaussians is again Gaussian:

$X \leftarrow N(\mu_1, \sigma_1^2)$ $\quad Y \leftarrow N(\mu_2, \sigma_2^2)$ $\quad X \perp Y$.

$$M_{X+Y}(t) = \mathbb{E}[e^{t(x+y)}] = \mathbb{E}[e^{tx}]\, \mathbb{E}[e^{ty}]$$

$$= e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} \cdot e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2}$$

$$= e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2) \cdot t^2}$$

$$\Rightarrow X+Y \leftarrow N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

E.X. X_i binary with $\Pr[X_i = 1] = p_i$ $\quad X_i \perp X_j$ for $i, j \in \{1, 2, \cdots, n\}$.

Let $X = \sum_i X_i$

$$M_{X_i}(t) = \mathbb{E}[e^{x_i t}] = p_i \cdot e^t + (1-p_i) = 1 + p_i(e^t - 1)$$

$$M_X(t) = \mathbb{E}[e^{\sum_i x_i t}] = \mathbb{E}\left[\prod_i e^{tx_i}\right] = \prod_i [1 + p_i(e^t - 1)]$$

Chernoff Bounds

→ apply Markov's inequality to moment generating

E.X. Coin Flipping. X is #heads after n times fair coin.

Markov: $\Pr[X \geq \frac{3}{4}n] \leq \frac{\frac{1}{2}n}{\frac{3}{4}n} = \frac{2}{3}$

Chebyshev: $\Pr[X \geq \frac{3}{4}n] \leq \Pr[|X - \frac{1}{2}n| \geq \frac{1}{4}n] = \frac{\frac{1}{4} \cdot n}{(\frac{1}{4}n)^2} = \frac{4}{n}$

Chernoff bound: $\Pr\{X \geq \frac{3}{4}n\} = \Pr\{X \geq (1+\frac{1}{2}) \mathbb{E}X\} \leq \left(\frac{e^{\frac{1}{2}}}{(\frac{3}{2})^{\frac{3}{2}}}\right)^{\frac{n}{2}} \approx (0.899)^{\frac{n}{2}}$ $= \exp(-\Omega(n))$

$\mu = \frac{n}{2}$, $\delta = \frac{1}{2}$

Easier bound: $\leq e^{-\frac{n}{2} \cdot (\frac{1}{2})^2 / 3} = e^{-n/24}$

Def. For any $t>0$, $\mathbb{P}\{X\ge c\} = \mathbb{P}\{e^{tX}\ge e^{tc}\} \le \dfrac{\mathbb{E}[e^{tX}]}{e^{tc}}$

For any $t<0$, $\mathbb{P}\{X\le c\} = \mathbb{P}\{e^{tX}\ge e^{tc}\} \le \dfrac{\mathbb{E}[e^{tX}]}{e^{tc}}$ 选一个t效果最好的.

Thm. Let $X=\sum_i X_i$, where $\begin{cases} X_i \leftarrow \text{Bernouli}(p_i) \\ X_i \perp X_j,\ \forall i,j \\ \mu = \sum_i p_i = \mathbb{E}(X). \end{cases}$

Chernoff Bound

Then:

For any $\delta>0$, $\mathbb{P}\{X\ge(1+\delta)\mu\} \le \left(\dfrac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$

For any $\delta>0$, $\mathbb{P}\{X\le(1-\delta)\mu\} \le \left(\dfrac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$

Prof. $\mathbb{E}[e^{tX}] = \prod_i(1+p_i(e^t-1)) \le \prod_i e^{p_i(e^t-1)} = e^{\mu(e^t-1)}$ When equal?

(upper bound) $\mathbb{P}[X\ge c] \le \underbrace{\dfrac{\mathbb{E}[e^{tX}]}{e^{tc}}}_{\text{Markov}} \le e^{\mu(e^t-1)-tc}$ for $t>0$

$P[X\ge(1+\delta)\mu] \le e^{\mu(e^t-1)-t\cdot(1+\delta)\mu}$ ← choose $t=\log(1+\delta)$

$= e^{\mu[\delta-(1+\delta)\log(1+\delta)]}$

$= \left(\dfrac{e^{\delta}}{(1+\delta)e^{(1+\delta)}}\right)^{\mu}$

lower bound is the same

$P[X\le(1-\delta)\mu] \le \left(\dfrac{e^{\delta}}{(1-\delta)e^{(1-\delta)}}\right)^{\mu}$ ← choose $t=\log(1-\delta)$

$\because \delta>0,$
$\therefore t<0.$ ✓

Col. Easier-to-parse Bounds

for any $\delta\in(0,1]$, $\mathbb{P}[X\ge(1+\delta)\mu] \le e^{-\mu\cdot\delta^2/3}$

偏移小时 $\mathbb{P}[X\le(1-\delta)\mu] \le e^{-\mu\cdot\delta^2/2}$

for $c\ge 6$, $\mathbb{P}[X\ge c\cdot\mu] \le 2^{-c\mu}$

偏移大时.

$e^{\mu[\delta-(1+\delta)\log(1+\delta)]} \le e^{-\mu\delta^2/3}$

$\delta-(1+\delta)\log(1+\delta) \le \delta^2/3$

Col. if you only have a bound on $\mathbb{E}(X)$, but you don't know $\mathbb{E}(X)$

replace μ with c, for any $c\ge\mu$, thm still holds

For any $\delta > 0$, $\mathbb{P}\{X \geq (1+\delta)c\} \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{c}$ for any $c \geq \mu$

For any $\delta > 0$, $\mathbb{P}\{X \leq (1-\delta)c\} \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{c}$ for any $c \leq \mu$

Prof. Let $c \geq \mu$. Let $Y_i = \mathrm{Ber}(q_i)$ so that $\sum_i q_i = c - \mu$. 可能大于1要加多个

Let $Z = \sum_i X_i + \sum_i Y_i$

$$\mathbb{P}\{X \geq (1+\delta)c\} \leq \mathbb{P}\{Z \geq (1+\delta)c\} \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{c}$$

Other useful Chernoff-like bounds

$\forall \delta > 0$, $\Pr[X > \mathbb{E}(X) + \delta] \leq e^{-2\delta^2/n}$

Hoeffding $X_i \in [a_i, b_i]$. $\forall \delta > 0$. $\Pr[|X - EX| \geq \delta] \leq 2 \cdot \exp\left(\frac{-2\delta^2}{\sum_i (a_i - b_i)^2}\right)$

Bernstein X_i : mean zero with $|X_i| \leq M$. $\forall \delta > 0$, $\Pr[X > \delta] \leq \exp\left(\frac{-\delta^2/2}{\sum_i \mathbb{E}[X_i^2] + \frac{1}{3}M\delta}\right)$

useful when var(X) is small

Lect 06

Balls and Bins

n balls

m bins

e.x. Birthday Paradox

365 bins 23 people

$$\Pr[\text{No two balls collide}] = 1 \times \left(\frac{m-1}{m}\right) \times \left(\frac{m-2}{m}\right) \times \cdots \left(\frac{m-n+1}{m}\right)$$

$$= \frac{(m-1)(m-2)\cdots(m-n+1)}{m^{n-1}}$$

If $m=365$, $n=23$ ≈ 0.4927.

e.x. Max Load

n balls $\to$ n bins. maximum number of balls in any bin

Prop. There's some constant c so that with high prob the max load will be at most $\frac{c\log n}{\log\log n}$, for sufficiently large n.

Prof. Step 1. For any $k > \frac{3\log n}{\log\log n}$, $\Pr[\text{Bin1 has load exactly } k] = o\left(\frac{1}{n^2}\right)$

$$\Pr[\text{Bin 1 has load exactly } k] = \binom{n}{k}\left(\frac{1}{n}\right)^k\left(\frac{n-1}{n}\right)^{n-k} \qquad k > \frac{3\log n}{\log\log(n)}$$

$$\le \left(\frac{n^k}{k!}\right)\left(\frac{1}{n}\right)^k \cdot 1$$

$$= \frac{1}{k!} \le \left(\frac{e}{k}\right)^k \le \left(\frac{e\log\log(n)}{3\log(n)}\right)^{\frac{3\log n}{\log\log n}}$$

$$\le \left(\frac{\log\log(n)}{\log(n)}\right)^{\frac{3\log n}{\log\log n}}$$

$$= \exp\left[\log\left(\frac{\log\log n}{\log n}\right)\cdot\frac{3\log n}{\log\log n}\right]$$

$$= \exp\left(\frac{3\log n}{\log\log n}\left[\log\log\log n - \log\log n\right]\right)$$

$$= n^{-3+\underbrace{3\log\log\log n/\log\log n}_{o(1)}}$$

$$= o(1/n^2)$$

接近 $\frac{1}{n^3}$, 所以是 $o\left(\frac{1}{n^2}\right)$

△ Stirling Approximation

$$\ln(k!) = \sum_{i=1}^{k}\ln i \approx \int_1^k \ln x\,dx = k\ln k - k + 1$$

$$k! \approx e^{k\ln k - k} = \left(\frac{k}{e}\right)^k$$

Step 2. Then we'll take the union bound over n bins, all $<n$ relevant values of k and conclude

$$\Pr\left[\text{Any bin has load} > \frac{3\log n}{\log\log n}\right] = o(1)$$

The Poisson Distribution

Def. $Poi(\lambda)$, for $\lambda \geq 0$. if $X \leftarrow Poi(\lambda)$, then $Pr[X=k] = \frac{e^{-\lambda}\lambda^k}{k!}$

Another def: limit of $Binomial(\frac{\lambda}{n}, n)$ as $n \to \infty$

Why are these the same?

$$\underbrace{\binom{n}{k}}_{\text{Stirling } \left(\frac{en}{k}\right)^k} \left(\frac{\lambda}{n}\right)^k \underbrace{\left(1-\frac{\lambda}{n}\right)}_{e^{-\lambda/n}}{}^{n-k} \quad \text{because } n \to \infty \qquad \text{v.s.} \qquad \frac{e^{-\lambda}\lambda^k}{k!}$$

$$\left(\frac{en}{k}\right)^k \left(\frac{\lambda}{n}\right)^k e^{-(\lambda/n)(n-k)}$$

$$= \underbrace{\left(\frac{e}{k}\right)^k}_{1/k!} \lambda^k \underbrace{e^{-\lambda(1-k/n)}}_{e^{-\lambda}}$$

$$= (e^{-\lambda} \cdot \lambda^k)/k!$$

=

Prop. ① If $X \leftarrow Poi(\lambda)$. then $\mathbb{E}[X] = Var[X] = \lambda$

② If $X \leftarrow Poi(\lambda_1)$, $Y \leftarrow Poi(\lambda_2)$, then $X+Y \leftarrow Poi(\lambda_1+\lambda_2)$

③ If $X \leftarrow Poi(\lambda)$, then for any $c>0$, $Pr[|X-\lambda| \geq c] \leq 2e^{\frac{-c^2}{2(c+\lambda)}}$

when c is large: $\Omega(e^{-c})$

e.g. Application to balls and bins

Setting: $k \leftarrow Poi(n)$ Toss k balls randomly to m bins.

- \# balls in bin i is distributed as $Poi(\frac{n}{m})$ for all i
- \# balls in bin 1, 2, $\cdots$ bin m are all independent! unlike before

Proof.

Assume 2 bins (m bins leave for exercise)

Let X_1, X_2 be # balls in bin 1 bin 2.

Claim: $Pr[X_1 = i, X_2 = j] = Pr[Poi(\frac{n}{2}) = i] \, Pr[Poi(\frac{n}{2}) = j]$

$$\mathbb{P}[X_1 = i, X_2 = j] = \mathbb{P}[k = i+j] \cdot \mathbb{P}[Binomial(i+j, \tfrac{1}{2}) = i]$$

\# balls ~ $Poi(n)$

$$= e^{-n} n^{i+j}/(i+j)! \cdot \binom{i+j}{i} \cdot \left(\tfrac{1}{2}\right)^{i+j}$$

$$= e^{-n/2} \cdot e^{-n/2} \cdot n^i \cdot n^j \cdot \frac{1}{i!\,j!} \left(\tfrac{1}{2}\right)^i \left(\tfrac{1}{2}\right)^j$$

$$= \mathbb{P}[Poi(n/2) = i] \, \mathbb{P}[Poi(n/2) = j]$$

Claim: $\Pr[X_1 = i] = \Pr[\mathrm{Poi}(\frac{n}{2}) = i]$, same for X_2. trivial 看 marginal

$$\mathbb{P}[X_1 = i] = \sum_j \mathbb{P}[X_1 = i, X_2 = j]$$
$$= \sum_j \mathbb{P}[\mathrm{Poi}(n/2) = i] \cdot \mathbb{P}[\mathrm{Poi}(n/2) = j] = \mathbb{P}[\mathrm{Poi}(n/2) = i]$$

Claim: X_1 and X_2 are independent. ✓

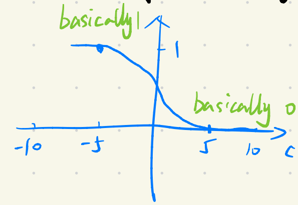
$$\mathbb{P}[X_1 = i, X_2 = j] = \mathbb{P}[X_1 = i]\,\mathbb{P}[X_2 = j].$$

Poissonization

n balls into m bins $\overset{\text{分析右边}}{\underset{\#\text{balls 独立}}{\Longrightarrow}}$ $k \leftarrow \mathrm{Poi}(n)$ balls into m bins.

e.x. Coupon Collecting

说明 X 接近 nlogn



Claim: For any (possibly negative) constant c,

$$\lim_{n\to\infty} \Pr[X \geq n\log n + cn] = 1 - e^{-e^{-c}}$$

Poissonization:

$n\log n + cn$ balls into n bins $\longrightarrow$ $k \leftarrow \mathrm{Poi}(n\log n + cn)$

Step 1. $\lim_{n\to\infty} \Pr[\text{See every type of coupon after } k] = 1 - e^{-e^{-c}}$

Poissonify

Suppose we open $k \leftarrow \mathrm{Poi}(n\log n + cn)$ boxes.

$\Pr[\mathrm{Poi}(\lambda) = 0] = e^{-\lambda}$

$\#$ coupons for each type $\leftarrow \mathrm{Poi}(\log n + c)$.

$$\Pr[\text{see all coupons after } k] = (1 - \mathbb{P}[\mathrm{Poi}(\log n + c) = 0])^n$$
$$= (1 - e^{-(\log n + c)})^n$$
$$= (1 - e^{-c}/n)^n \longrightarrow e^{-e^{-c}} \quad \square$$

De-Poissonify

because $(1 - \frac{1}{x})^{-x} \to e$, $(1 - \frac{1}{x})^x \to \frac{1}{e}$

Step 2. $\Pr[\text{See every type of coupon after } n\log n + cn] = \Pr[\cdots \text{ after } k] + o(1)$

Let $X = \#$boxes needed to get all coupons

$$\mathbb{P}[X \geq b + n^{0.9}] = \mathbb{P}[X \geq b - n^{0.9}] + \mathbb{P}[\ \ ]$$
$$\leq \mathbb{P}[X \geq b - n^{0.9}] + \frac{2 \cdot n^{0.9}}{n}$$
$$= \mathbb{P}[X \geq b - n^{0.9}] + o(1)$$

$\Rightarrow$ Union Bound of $2 \cdot n^{0.9}$ box

find the last coupon

$$\Pr[|k - (n\log n + cn)| \geq n^{0.9}] \leq \frac{\mathrm{Var}(k)}{n^{1.8}} = \frac{n\log n + cn}{n^{1.8}} = o(1)$$

$$\Rightarrow \lim_{n\to\infty} \Pr[X \geq n\log n + cn] = 1 - e^{-e^{-c}}$$

Maximum Load

Step 1. Draw $k \leftarrow \mathrm{Poi}(\frac{n}{2})$ ☆不是 n! Each bin : $\mathrm{Poi}(\frac{1}{2})$

Let $b = \frac{c \log n}{\log \log n}$. $\mathbb{P}[\text{no bin with load} \geq b] \leq \mathbb{P}[\text{no bin with load} = b]$

$$= \left(1 - \frac{e^{-(\frac{1}{2})}(\frac{1}{2})^b}{b!}\right)^n$$

$$\frac{1}{\sqrt{e}\, 2^b b!}$$

Stirling

$$\approx \exp(-b\log b + b + O(\log b) - b\ln 2 - \tfrac{1}{2})$$

$$= \exp\left(-\frac{c\log n}{\log\log n} \cdot \log\left(\frac{c\log n}{\log\log n}\right) + [\text{stuff}]\right)$$

$$= \exp(-c\log n + [\text{stuff}])$$

$$= \Omega(n^{-c})$$

$$= (1 - \Omega(n^{-c}))^n$$

$$\leq e^{-\Omega(n/n^c)} = o(1) \text{ if } c < 1$$

$\Pr[\text{max load after } k \text{ is } \geq \frac{c\log n}{\log\log n}] \geq 1 - o(1)$

Step 2. For $k \leq n$, then

$\Pr[\text{max load after } n \geq \frac{c\log n}{\log\log n}] \geq \Pr[\text{max load after } k \geq \frac{c\log n}{\log\log n}] \geq 1 - o(1)$.

$\mathbb{P}\{k \leq n\} \leq \mathbb{P}\{|k - E(k)| \geq \frac{n}{2}\} \leq \frac{\mathrm{Var}(k)}{(\frac{n}{2})^2} = \frac{n/2}{n^2/4} = o(1)$

Lect 07

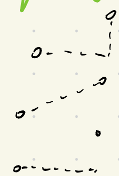
Metric Embeddings

Def Metric Space (X, d) X is a set, $d: X \times X \Rightarrow \mathbb{R}$ is the metric, which satisfies:

- For all $x \in X$, $d(x,x)=0$.
  $x, y \in X$ with $x \neq y$, $d(x,y) = d(y,x) > 0$
  $x,y,z \in X$, $d(x,y) \le d(x,z) + d(z,y)$ triangle inequality

E.X. ℓ_p. $(\mathbb{R}^k, d_p)$ $p \ge 1$.

$$d_p(x,y) = \left(\sum_{i=1}^{k} (x_i - y_i)^p \right)^{\frac{1}{p}} = \|x-y\|_p.$$

l_1 
l_2
l_∞ $\max_i |x_i - y_i|$

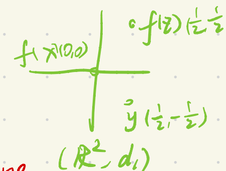
E.X. Graph Metric Space (V, d_G) positive edge weights

$d_G(x,y)$ = length of shortest path between x and y in G.

Any metric space (X,d) where X is finite can be represented as a graph metric!

Def. Isometric Embeddings

(X, d_X) (Y, d_Y) metric space. a function $f: X \to Y$ is a Isometric Embedding if for all $x, x' \in X$. $d_Y(f(x), f(x')) = d_X(x, x')$ Equal!



Any finite vector space can be isometrically embedded into l_∞

$X = \{x_1, \cdots, x_n\}$, metric d. define $f: X \to \mathbb{R}^n$ by. Fréchet Embedding

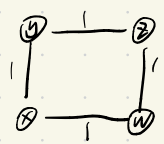
$f(x_i) = (d(x_1, x_i), d(x_2, x_i), \cdots, d(x_n, x_i))$.

Claim: $d(x_i, x_j) = d_\infty(f(x_i), f(x_j))$

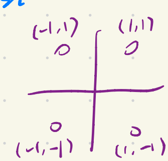
$d_\infty(f(x_i), f(x_j)) = \max_k \left| d(x_i, x_k) - d(x_j, x_k) \right|$ $\le d(x_i, x_j)$ $k=i/j$时取max

But isometric embeddings don't always exist

e.g.


(V, d_G)

$\xrightarrow{?}$ l_2 space



$\beta = \sqrt{2}$

$\alpha = \sqrt{2}$

Def. Low-Distortion Embeddings

(X, d_X) (Y, d_Y) metric space. a function $f: X \to Y$ is a embedding of X to Y with distortion α and scaling factor β, if for all $x, x' \in X$.

$$\beta\, d_X(x, x') \le d_Y(f(x), f(x')) \le \alpha\beta\, d_X(x, x')$$

Bourgain's embedding

Thm. Given any finite (X,d) with $|X|=n$, there exists an embedding of (X,d) into $(\mathbb{R}^k, d_1)$ where $\begin{cases} k=O(\log^2 n) \\ \text{distortion is } O(\log n)\end{cases}$

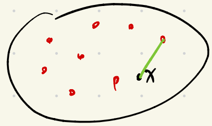
d_1: $\sum |x_i - y_i|$

Actually this embedding works for any d_p $(p \geq 1)$

For $i = 1, 2, \cdots \log n$ — Different Density
  For $j = 1, 2, \cdots c\log n$ — c: some constant we'll choose; j: Different tries
    Choose $S_{ij} \subseteq X$ at random. so every $x \in X$ is contained in S_{ij} with prob 2^{-i} （S_{ij}：一个集合）

Define
$$f(x) = (d(x, S_{1,1}), d(x, S_{1,2}), \cdots, d(x, S_{\log n \cdot c\cdot \log n}))$$
$$d(x,S) = \min_{y\in S} d(x,y)$$

Intuition
The embedding

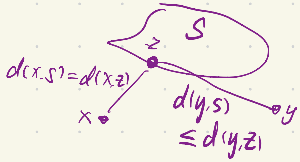


X

random set get more sparse
each time find the nearest …

Consider $f(x) = d(x,S)$.

① The map is non-expanding.
$$|f(x) - f(y)| \leq d(x,y) \text{ for all } x,y\in X$$



② $|f(x) - f(y)| \geq \Delta$. not too shrinking

Bourgain's Embedding Proof

The Embedding

For $i = 1, 2, \cdots \log n$
  For $j = 1, 2, \cdots c\log n$
    Choose $S_{ij} \subseteq X$ at random. so every $x \in X$ is contained in S_{ij} with prob 2^{-i}

Define
$$f(x) = (d(x, S_{1,1}), d(x, S_{1,2}), \cdots, d(x, S_{\log n \cdot c\cdot \log n}))$$

Thm. Let $k = c\log^2 n$. There is some constant b so that for all $x, y \in X$,
$$\frac{k}{b\log n} d(x,y) \leq \|f(x) - f(y)\|_1 \leq k\cdot d(x,y)$$
即 $(\mathbb{R}^k, d_1)$, distortion $O(\log n)$, k: $O(\log^2 n)$

Proof. Step 1 $\|f(x)-f(y)\|_1 \le k\cdot d(x,y)$

- Enough to show that $|d(x,S)-d(y,S)| \le d(x,y)$ for every S. (there are k different S in total)

  (口语化)
  a.k.a. $x \mapsto d(x,S)$ is non-expanding

  Def. for every $x,y\in X$. $f(x)=d(x,S)$ the map for a single dimension
  $|f(x)-f(y)|\le d(x,y)$

- $\|f(x)-f(y)\|_1 = \sum_{i,j} |d(x,S_{ij})-d(y,S_{ij})|$
  $\le k\cdot d(x,y)$ ☐

We show that already!

Say $d(x,z)=d(x,S)$, $z\in S$. then

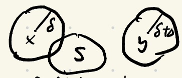
选出离x最近的点

$d(y,S)\le d(y,z)$

$f(y)-f(x) = d(y,S)-d(x,S)$
$\le d(y,z)-d(x,z)$
$\le d(x,y)$ triangle.

Switch x,y, also have $f(x)-f(y)\le d(x,y)$
So $|f(y)-f(x)|\le d(x,y)$.

Step 2. $\frac{k}{b\log n} d(x,y) \le \|f(x)-f(y)\|_1$

Intuition: Choose lots of δ, show that  happens with decent prob.

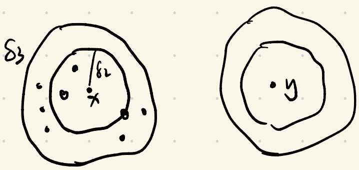
$S\cap\{x+\delta\}\ne\phi$ $\quad S\cap\{y+\delta+\Delta\}=\phi$.

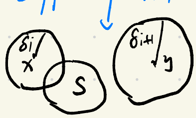
$B(x,\delta) := \{z\in X: d(x,z)\le\delta\}$ 圆

Fix x,y. Choose $0=\delta_0<\delta_1<\delta_2<\dots<\delta_t$ so that δ_i is the smallest value so that $B(x,\delta_i)$ and $B(y,\delta_i)$ both have $\ge 2^i$ points in them.

Stop at the largest t so that $\delta_t < \frac{d(x,y)}{3}$

Let $\delta_{t+1} = \frac{d(x,y)}{3}$



Claim 1. Suppose  holds. then $|d(x,S)-d(y,S)|\ge\delta_{i+1}-\delta_i$

$d(x,S)\le\delta_i$
$d(y,S)\ge\delta_{i+1}$ ☐.

Claim 2. It's decently likely that occurs.

Either there are $<2^{i+1}$ points in $B(x,\delta^{i+1})$ or $B(y,\delta^{i+1})$
Suppose it's y.

Why? 如果两个边界都有点呢?

$\mathbb{P}[x\in S_{ij}] = 2^{-i}$ for all $x\in X$

$\mathbb{P}[\ \cdot\ ] = \mathbb{P}[\ \cdot\ ]\times\mathbb{P}[\ \cdot\ ]$

$\ge (1-\underbrace{(1-1/2^i)^{2^i}}_{\text{all } x\in X \text{ not in } S_{ij}})\times\underbrace{(1-1/2^i)^{2^{i+1}}}_{\text{some } x\in X \text{ not in } S_{ij}}$

$\ge (1-\frac{1}{e})\times(1-1/2)^{2^2} \ge 1/2^5$

Fix i. $\Pr[\ge 2^{-6}$ fraction of the S_{ij} have:  $]$

$= 1-\mathbb{P}[\underbrace{\sum_{j=1}^{c\log n}\mathbb{1}\{\ \cdot\ \}}_{\mathbb{E}\text{ would be }\frac{c\log n}{2^5}}\le\frac{1}{2}\cdot\frac{c\log n}{2^5}]$

chernoff bound

$> 1-\exp(\frac{-c\log n}{8\cdot 2^5}) \ge 1-1/n^3$. if we chose $c>3\cdot 2^8$.

Putting it together. Union Bound over n^2 pairs x,y and all $\le\log n$ choices of i.

with high prob whp. $\forall i,x,y.$ $\ge\frac{c\log n}{2^6}$ S_{ij}'s have $|d(x,S_{ij})-d(y,S_{ij})|\ge\delta_{i+1}-\delta_i$

$$\|f(x)-f(y)\|_1 = \sum_{ij} |d(x,S_{ij}) - d(y,S_{ij})| \geq \sum_{ij} \left(\frac{c\log n}{2^b}\right)(\delta_{i+1}-\delta_i)$$

$$= \frac{c\log n}{2^b}\cdot \delta_{t+1} = \frac{c\log n}{3\cdot 2^b} d(x,y)$$

$$\|f(x)-f(y)\|_1 \geq \frac{k}{\underbrace{3\cdot 2^b}_{b}\log n}\cdot d(x,y).$$

$$\therefore \frac{k}{b\log n} d(x,y) \leq \|f(x)-f(y)\|_1 \leq k\cdot d(x,y)$$

这里是随机的吗??

w.h.p embedding ✓

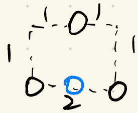
不成立的话重新生成?

quiz.

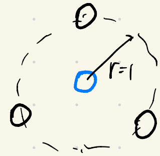
Q1 $d(x_0, x_{123}) = 1$

$d(x_i, x_j) = 2 \quad ij > 0$

d_1

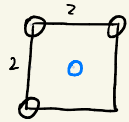


d_2



X. 不可能两两连成直径.

d_∞

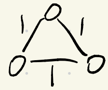


Q2. $d(x_i, x_j) \overset{?}{=} d_\infty(f(x_i), f(x_j))$

$$\overset{\parallel}{\max_k |d(x_i, x_k) - d(x_j, x_k)|} \quad \checkmark = d(x_i, x_j)$$

$d_1(f(x_i), f(x_j))$

$$\overset{\parallel}{\sum_k |d(x_i, x_k) - d(x_j, x_k)|}$$



A: (0, 1, 1)
B: (1, 0, 1)
C: (1, 1, 0)

$d_\infty = 1$
$d_1 = 2$

DRW	1h tech ✗				
Citadel	1h tech ✗				
Flatiron	1h tech ✓	2h tech ✓	offer		
Palantir	hr call ✓	1h tech ✓	2h tech ✓	1h hm ✓	offer
Two Sigma	1h tech ✓	2h tech ✓	3h bq ✓	offer	
HRT	1h math ✓	1h tech ✗			
Akuna	1h tech ✓	2h tech	withdraw		
Pinterest	hr call ✓	1h tech ✓	0.5h bq ✓	offer	
eBay	2h tech	withdraw			

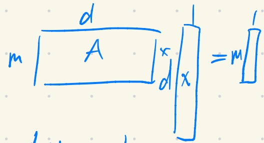
Johnson - Lindenstrauss Lemma

☆ Given any $\epsilon \in (0,1)$ and a set $X \subset \mathbb{R}^d$ with $|X| = n$, there exists a linear map $f: \mathbb{R}^d \to \mathbb{R}^m$ with $m = O(\frac{\log n}{\varepsilon^2})$ that embeds (X, d_2) into $(\mathbb{R}^m, d_2)$ with distortion at most $(1+\varepsilon)$.

Proof construct a random f → w.h.p f satisfies → such f exists.

Let $A \in \mathbb{R}^{m \times d}$ be a random matrix with entries chosen independently from $N(0, \frac{1}{m})$

Define $f(x) = A \cdot x$

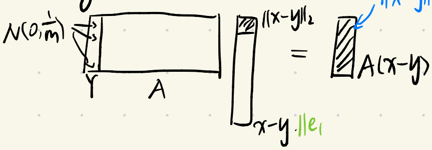


① Fix $x, y \in \mathbb{R}^d$. we'll show that w.h.p. $\|f(x) - f(y)\|_2 = \|A(x-y)\|_2 = (1 \pm \varepsilon)\|x-y\|_2$.

② Union bound over all n^2 (x,y) pairs.

Without loss of generality, suppose $x - y = \|x-y\|_2 \cdot e_1$ $\quad e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

New goal:



want to show: $\|A(x-y)\|_2 = (1 \pm \varepsilon)\|x-y\|_2$

$\Updownarrow$

$\|Y\|_2 = 1 \pm \varepsilon$.

Y is a random vector with $Y_i \sim N(0, \frac{1}{m})$

Let $Z_1, \cdots Z_m \sim N(0,1)$ $\quad$ ↗ choose later

$\Pr[\sum_i Z_i^2 > (1+\varepsilon)m] = \mathbb{P}[e^{t\sum_i Z_i^2} > e^{t(1+\varepsilon)m}]$

$\mathbb{E}e^{tZ_i^2} = 1/\sqrt{1-2t}$ for $t < \frac{1}{2}$.

$\leq \frac{\prod_i \mathbb{E} e^{tZ_i^2}}{e^{tm(1+\varepsilon)}}$

$= (\frac{1}{1-2t})^{\frac{m}{2}} \cdot \frac{1}{e^{tm(1+\varepsilon)}}$ for $t < \frac{1}{2}$

$\frac{1}{2}\log(\frac{1}{1-2t}) \leq t + 2t^2$ for $0 < t < \frac{1}{2}$

$= \exp(m[\frac{1}{2}\log(\frac{1}{1-2t}) - t(1+\varepsilon)])$

$\leq \exp(m(2t^2 - \varepsilon t))$ ← choose $t = \frac{\varepsilon}{4}$.

$= \exp(-m\varepsilon^2/8)$

$\mathbb{P}[\|Y\|_2 > 1+\varepsilon] = \mathbb{P}[\|Y\|_2^2 > (1+\varepsilon)^2]$

$\leq \mathbb{P}[\|Y\|_2^2 > 1+\varepsilon] = \mathbb{P}[\sum_i Z_i^2 > m(1+\varepsilon)] \leq e^{-\frac{m\varepsilon^2}{8}}$

$\mathbb{P}[\|A\cdot(x-y)\|_2 \geq (1+\varepsilon)\|x-y\|_2] \leq e^{-m\varepsilon^2/8}$

$\mathbb{P}[\exists x, y \in X, \text{ s.t. } \|A(x-y)\|_2 \geq (1+\varepsilon)\|x-y\|_2] \leq n^2 \cdot e^{-m\varepsilon^2/8}$

$m = O(\frac{\log n}{\varepsilon^2}) \Longrightarrow \quad \leq 1/\text{poly}(n)$

$1-\varepsilon$ Very Similar.

Nearest Neighbors Search.

Def. $S \subset \mathbb{R}^n$ has size n. Goal: Given $x \in \mathbb{R}^d$, find the $y \in S$ that's closest to x.

Space $n^{O(d)}$
Time $d^{O(1)} \log n$

Def. (c-Approximate Nearest Neighbors)

- $S \subset \mathbb{R}^n$ has size n.
- Construct a data structure that can answer questions in this form:

  Given $x \in \mathbb{R}^d$, find any $y \in S$ s.t. $\|x-y\|_2 \le c \cdot \min_{z \in S} \|x-z\|_2$

Next Idea: JL Transform!

Data Structure: Let $f: \mathbb{R}^d \to \mathbb{R}^m$ be a JL Transform. $\varepsilon = \frac{1}{2}$. $m = O(\frac{\log n}{\varepsilon^2})$

Query Algorithm: Use the "Space $n^{O(m)}$, Query Time $m^{O(1)} \log n$" scan for exact NN, on $\{f(y): y \in S\}$.

Space: $n^{O(m)} = n^{O(\log n)}$ Query time: $m^{O(1)} \log n = \operatorname{poly}\log(n)$

quiz.

Q2. ① $x=3$. Mindis $=1$. ② $1 \le x' \le 5$ 3 Answers

② $x=7$ 0 ⓪ 1

③ $x=7.3$ 0.2 0.4 $6.9 \le x' \le 7.7$ 2

(×2)

P balls $\to N$ bins.

$C \le \frac{N}{2} \leftrightarrow$ Empty D. $D \ge \frac{N}{2}$

$K \leftarrow Poi(\frac{3}{4}n)$

each bin: $X_i \leftarrow Poi(\frac{3}{4})$

Indicator of $X_i = 0 \Rightarrow Y_i$

$\mathbb{E}(Y_i) = e^{-\frac{3}{4}}$

$$Pr(D \ge \tfrac{N}{2}) = Pr(\sum Y_i \ge \tfrac{N}{2}) \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} = \left(\frac{e^{\delta}}{\left(\frac{e^{\frac{P}{N}}}{2}\right)^{1+\delta}}\right)^{N\cdot e^{-\frac{P}{N}}}$$

$$\boxed{\left(\frac{e^{\frac{e^{\frac{3}{4}}}{2}-1}}{\left(\frac{e^{\frac{3}{4}}}{2}\right)^{\frac{e^{\frac{3}{4}}}{2}}}\right)^{N\cdot e^{-\frac{3}{4}}}}$$

$\downarrow \log$

$\mu = N e^{-\frac{3}{4}}$

$(1+\delta)\mu = \frac{N}{2}$

$$\delta = \frac{1}{2 \cdot e^{-\frac{3}{4}}} - 1 = \frac{e^{\frac{3}{4}}}{2} - 1$$

$N \cdot e^{-\frac{3}{4}}[\cdots]$. $\cancel{-N \cdot e^{-\frac{4}{5}} \cdot \frac{(\frac{e^{\frac{4}{5}}}{2} - 1)^2}{3}}$

$$\frac{N}{e}\left[\frac{e}{2} - 1 - \frac{e}{2}\log\frac{e}{2}\right] = N\left[\frac{1}{2} - \frac{1}{e} - \frac{1}{2}\log\frac{e}{2}\right] = \Omega(N)$$

Claim:

$$Pr(D \ge \tfrac{N}{2} \text{ with } k \text{ balls}) = Pr(D \ge \tfrac{N}{2} \text{ with } N \text{ balls}) + o(e^{-\Omega(N)}).$$

$$Pr(|X - \lambda| \ge c) \le 2e^{-\frac{c^2}{c+\lambda}} \qquad \lambda = \tfrac{3}{4}n. \quad c = \tfrac{1}{4}n.$$

$$2e^{-\frac{\frac{1}{16}n^{\cancel{2}}}{\cancel{n}}} = 2e^{-\frac{1}{16}n}$$

$t > 1$

$t(1 - \log t) < 1$

$1 - \log t < \frac{1}{t}$.

$\frac{1}{t} + \log t - 1 > 0$ ($t=1$)

$e^P > 2$. $P > \ln 2$

$$\frac{e^P}{2}\left[1 - \log\frac{e^P}{2}\right] < 1$$

$$e^{\frac{e^P}{2} - 1} < \left(\frac{e^P}{2}\right)^{\frac{e^P}{2}}$$

$$\frac{e^P}{2} - 1 < \frac{e^P}{2}\log\frac{e^P}{2}$$

De-Poissonify

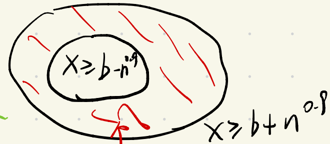
Step2. $\Pr[\text{See every type of coupon after } n\log n + cn] = \Pr[\cdots \text{ after } k] + o(1)$

Let X = # boxes needed to get all coupons

$P[X \geq b + n^{0.9}] = P[X \geq b - n^{0.9}] + P[\quad]$

$\leq P[X \geq b - n^{0.9}] + \frac{2 \cdot n^{0.9}}{n}$ $\Rightarrow$ Union Bound of $2 \cdot n^{0.9}$ box

$= P[X \geq b - n^{0.9}] + o(1)$



$\Pr[|k - (n\log n + cn)| \geq n^{0.9}] \leq \frac{Var(k)}{n^{1.8}} = \frac{n\log n + cn}{n^{1.8}} = o(1)$

$\Rightarrow \lim_{n\to\infty} \Pr[X \geq n\log n + cn] = 1 - e^{-e^{-c}}$

Lect 09 Compressed Sensing and the RIP

Restricted Isometry Property

Def. (Sparsity) For $x \in \mathbb{R}^n$, $\|x\|_0 =$ # nonzero coordinates of x.
x is k-sparse if $\|x\|_0 = k$.

Embedding sparse vectors

k-sparse n-dim ℓ_2 space $\xrightarrow{\text{distortion } 1+\delta}$ ℓ_2 space of dim $O(\frac{k \cdot \log n}{\delta^2})$

infinite set!

Def (RIP) Matrix $A \in \mathbb{R}^{m\times n}$ has the Restricted Isometry Property with para k, δ if for any $x \in \mathbb{R}^n$ with $\|x\|_0 \le k$,

$$(1-\delta)\|x\|_2 \le \|Ax\|_2 \le (1+\delta)\|x\|_2$$

Application:

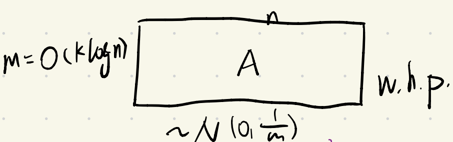
We can recover x from Ax if A has RIP.
Say A has RIP with paras $2k$, $\delta = 0.1$
Claim: $\exists$ unique k-sparse $x \in \mathbb{R}^n$, given Ax.
Proof: if x, x' are k-sparse, $Ax = Ax'$, then
$\|x - x'\|_2 \le \frac{1}{1-\delta}\|A(x-x')\|_2 = 0$ Conflict.

Recover Alg:

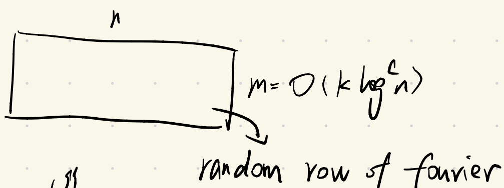
inefficient $\hat{x} = \arg\min_y \|y\|_0$ s.t. $Ay = Ax$
efficient $\hat{x} = \arg\min_y \|y\|_1$ s.t. $Ay = Ax$

if A has $2k$-RIP they're the same

What matrices have the RIP?



. . .

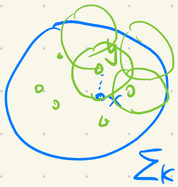


best deterministic constructions: $m = O(k^{1.99} \log n)$

Random Gaussian Matrices have the RIP whp.

Thm. Let $\delta \in (0,1)$ and choose $k<n$. There is some $m = O(\frac{k \log n}{\delta^2})$ so that a matrix $A \in \mathbb{R}^{m\times n}$ with independent entries $A_{ij} = N(0, \frac{1}{m})$ has the RIP with prob $1 - O(1)$.

Proof. Let $\Sigma_k = \{x \in \mathbb{R}^n : \|x\|_2 = 1, \|x\|_0 \le k\}$ → 其它等比例 distortion 不影响结果



1. Show that for any fixed $x \in \Sigma_k$. $\|Ax\|_2 \approx \|x\|_2$ whp.

~~2. Union bound~~ infinite x

2. Define a finite set $Y \in \Sigma_k$ that "covers" Σ_k pretty well.
3. Union bound over all $y \in Y$
4. Argue that this is enough.

Def. (ε-coverings) An ε-covering of $X \subseteq \mathbb{R}^n$ is a subset $Y \subseteq X$ so that for all $x \in X$, $\exists y \in Y$ s.t. $\|x - y\|_2 \le \varepsilon$.

Lemma. For $\varepsilon \in (0,1)$, $X = \{x \in \mathbb{R}^k : \|x\|_2 = 1\}$ has an ε-covering of size $\le \left(\frac{3}{\varepsilon}\right)^k$

Proof. Consider greedily choosing Y:

$$\|y - y'\|_2 \ge \varepsilon. \quad \forall y \ne y' \in Y$$

$$B(y; \varepsilon/2) \subseteq B(0; 1 + \tfrac{\varepsilon}{2}) \quad \forall y \in Y$$

$$\Rightarrow \sum_{y \in Y} \mathrm{Vol}(B(y, \varepsilon/2)) \le \mathrm{Vol}(B(0; 1 + \tfrac{\varepsilon}{2}))$$

$$\mathrm{Vol}(B(0, \ell)) = C_k \ell^k$$

$$|Y| \le \frac{C_k (1 + \varepsilon/2)^k}{C_k (\varepsilon/2)^k} = \left(\frac{2(1 + \frac{\varepsilon}{2})}{\varepsilon}\right)^k$$

$$\le \left(\frac{3}{\varepsilon}\right)^k.$$

Corollary:

Set Σ_k has an ε-covering Y of size at most $\binom{n}{k} \cdot \left(\frac{3}{\varepsilon}\right)^k$ → 每个用 $\left(\frac{3}{\varepsilon}\right)^k$

n 维里选 k 维不是。

W.h.p. $\forall y \in Y$. $\|y\|_2 = (1 \pm \varepsilon) \|Ay\|_2$.

Fix $y \in Y$.

Claim: $\Pr\left[\, \big|\, \|y\|_2 - \|Ay\|_2 \big| \ge \varepsilon \right] \le 2 \cdot \exp(-c\varepsilon^2 m)$ → constant.

Proof. See proof of JL. ✓

Union Bound:

$$\mathbb{P}\left[\exists y \in Y \text{ s.t. } \big|\|y\|_2 - \|Ay\|_2\big| \ge \varepsilon\right] \le \underbrace{\binom{n}{k}}_{\le n^k} \left(\frac{3}{\varepsilon}\right)^k \cdot 2\exp(-c\varepsilon^2 m)$$

$m=\Theta(\frac{k\log n}{\varepsilon^2})$ $\quad \le \exp(k\log n + k\log(\frac{3}{\varepsilon}) + \ln 2 - c\varepsilon^2 m)$

$\le \exp(-\Omega(k\log n))$

W.h.p. $\forall x\in X$ $\|x\|_2=(1\pm\varepsilon)\|Ax\|_2$.

Suppose for $\forall y\in Y$, $\|y\|_2=(1\pm\varepsilon)\|Ay\|_2$

Let δ^* be the smallest value so that $\forall x\in\Sigma_k$, $\|x\|_2=(1\pm\delta^*)\|Ax\|_2$.

Pick $x\in\Sigma_k$. let $y\in Y$ s.t. $\|x-y\|_2\le\varepsilon$ and $\|x-y\|_0\le k$. $\because$ y is constructed from $\binom{n}{k}$ spheres that is k-sparse. 选对应x k维的就好.

Then $\big|\|Ax\|_2-\|x\|_2\big| \le \underbrace{\big|\|Ay\|_2-\|y\|_2\big|}_{\le\varepsilon} + \underbrace{\|x-y\|_2}_{\le\varepsilon}+\underbrace{\|A(x-y)\|_2}_{\le(1+\delta^*)\|x-y\|_2}$ $x-y$代入

$\le(1+\delta^*)\varepsilon$

$$\delta^*\le 3\varepsilon+\delta^*\varepsilon \Rightarrow \delta^*\le\frac{3\varepsilon}{1-\varepsilon}\le 4\varepsilon = \delta.$$

选 $\varepsilon=\frac{\delta}{4}$? ✓

How to check if A satisfies the constraint?

上课 Single Pixel Camera.

$G(n,p)$. random graph with n vertices. each $\binom{n}{2}$ edge exists with prob $p=\frac{c}{n}$

Thm If $c<1$ w.h.p. the largest connected component has size $O(\log n)$.

If $c>1$ w.h.p. the largest — — — $f(c)\cdot n$

second largest — — — $O(\log n)$

Proof. Start with vertex $\underline{v}$. Check its neighbors.

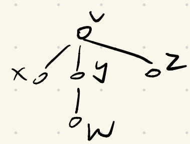
$\underline{v}$ is expected to have c neighbors. What's the distribution? Binomial.

Neighbors: $(n-1)$ possible, each with $p=\frac{c}{n}$ prob.

v's connected component

| v | x | y | z | w | - - - - |

BFS



Claim: for $k\in[100\log n,\ 0.1n]$ with prob $>1-\frac{1}{n^3}$

given that we have "explored" k vectors. we still have ≥1 unexplored vertices

For each "exploration", we expect to add $\underline{c-1}$ "unexplored" vertices

因为k很小,可以忽略?

Lect 10

The Probablistic Method and Ramsey Numbers

Def. (The probabilistic Method)

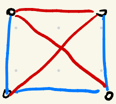
Does some object C exist? $\begin{cases} 1.\ \text{Define random variable } X \\ 2.\ \text{Show that } Pr[X=C]>0 \end{cases}$

If sth. exists with positive prob, it exists!

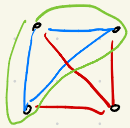
Def. (Ramsey Numbers)

Two-colorings of the complete graph and monochromatic k-cliques

k-cliques: k vertices that are all connected (e.g. K_4)



a 2-coloring of K_4



⇒ a monochromatic 3-clique.

Ramsey Numbers:

How many vertices n do you need before there must be a monochromatic k-clique?

monochromatic k-clique: K'th Ramsey Number. R_k.

e.g. $R_3=6$

$R_1=1,\ R_2=2,\ R_3=6,\ R_4=18,\ 43\le R_5\le 48,\ \cdots,\ 798\le R_{10}\le 23{,}556$

Thm. $R_k\in(2^{k/2},2^{2k})$

Proof. ① $R_k>2^{k/2}$. Show there's a coloring on $n=2^{k/2}$ vertices with no monochromatic k-clique.

Say $n=2^{k/2}$, and color the graph at random.

$Pr[\text{a given set of } k \text{ vertices is monochromatic}]=(\frac{1}{2})^{\binom{k}{2}-1}$

$Pr[\exists \text{ monochromatic } k \text{ clique}]\le\binom{n}{k}(\frac{1}{2})^{\binom{k}{2}-1}\le\frac{n^k}{k!}2^{-\frac{k^2}{2}+\frac{k}{2}+1}=\frac{2^{1+\frac{k}{2}}}{k!}<1$ for $k\ge3$.

plug in $n=2^{k/2}$

用 probablistic Method 妙!

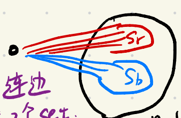
$\therefore Pr[\nexists \text{ monochromatic k-clique}]>0$. so $\exists$ coloring w/ no monochromatic k-clique.

② $R_k<2^{2k}$ Show any coloring on $n\ge2^{2k}$ vertices must have a monochromatic k-clique.

Some Def. Let $R_{r,b}$ be the minimal number of n so that the complete graph on n vertices has either a red r-clique or a blue b-clique.

Claim. $R_{r,b}\le R_{r-1,b}+R_{r,b-1}+1$.

Suppose $n=R_{r-1,b}+R_{r,b-1}+1$.



按这个点连边颜色分成2个set.

n-1 vertices

$|S_r|+|S_b|=R_{r-1,b}+R_{r,b-1}$

So Either $|S_r|\ge R_{r-1,b}$ ⇒ OR $|S_b|\ge R_{r,b-1}$

Suppose it's $|S_r|\ge R_{r-1,b}$.

Either: S_r has a blue b-clique ✓

S_r has a red r-1 clique ✓ plus the left side vertex.

Induction. $R_{r,b}<2^{r+b}$ for all $r+b\le c$. Check the base case!

注意严格

Say $r+b=c+1$.

$R_{r,b}\le R_{r-1,b}+R_{r,b-1}+1\le(2^{r+b-1}-1)+(2^{r+b-1}-1)+1<2^{r+b}$ ∎

By induction, $R_{r,b}<2^{r+b}\ \forall r,b \Rightarrow R_k=R_{k,k}<2^{2k}$

Independent Sets

Def. (Independent set) in a graph $G=(V,E)$, is a set $S \subseteq V$ so that no two vertices in S are connected by an edge.

Finding the size of the biggest independent set is NP-hard.

Thm. For any $G=(V,E)$ with $|V|=n$ and $|E|=m$, and $m \geq \frac{n}{2}$, there is an independent set in G of size at least $\frac{n^2}{4m}$.

e.g. $m=\binom{n}{2}$ Complete Graph. $\frac{n^2}{4m} = \frac{n^2}{4\binom{n}{2}} \approx \frac{1}{2}$.

$m=\binom{n}{2}/100$. Kind of Sparse Largest I.S. has size $\geq$ 50 ish.

$m = 100n$ Really Sparse $\geq n/400$

One way to find an independent set:

- For each edge $(u,v) \in E$, remove one of u,v. $\Rightarrow$ Always return an I.S. but not always a big one.
- Return what's left.

A better way

- For each vertex v. with prob $1-\frac{n}{2m}$, remove v and all the edges attached to it.
- For each remaining edge $(v,u) \in E$. remove one of u,v. (either one).
- Return what's left.

Analysis

Let X be # vertices that survives the first step
Y edges second

$\mathbb{E}(X) = n \cdot \left(\frac{n}{2m}\right) = \frac{n^2}{2m}$

$\mathbb{E}(Y) = \sum_{\text{edge }(u,v)} \mathbb{P}[\text{both } u,v \text{ survived}] = m \cdot \left(\frac{n}{2m}\right)^2 = \frac{n^2}{4m}$

Size of I.S. returned $\geq X - Y$.

Because each edge kills one remaining vertex at most.

$$\mathbb{E}(\text{size of I.S. returned}) \geq \mathbb{E}(X) - \mathbb{E}(Y) = \frac{n^2}{2m} - \frac{n^2}{4m} = \frac{n^2}{4m} \quad \blacksquare$$

Since $\mathbb{E}(\cdots) = \frac{n^2}{4m}$. At least one I.S. returned has size $\geq \frac{n^2}{4m}$.

p: S_i is mono.

$Pr[\exists i \text{ s.t. } S_i \text{ is mono}] \le p \cdot m < 1$

$\therefore Pr[\forall i\ S_i \text{ isn't mono}] > 0$.

$m < \frac{1}{p}$

Q3. $\frac{n^2}{4m} = \frac{n^2}{40n} = \frac{n}{40}$.

$C_x + \frac{n - C_x}{b} = E[count(x)]$

$Pr[count(x) \ge C_x + \varepsilon n]$

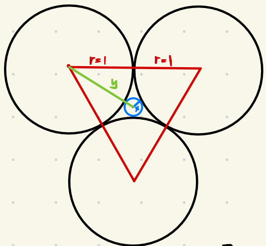
$\le \frac{\frac{b-1}{b} C_x + \frac{n}{b}}{C_x + \varepsilon n} \le \frac{1}{b} \cdot \frac{(b-1)C_x + n}{C_x + \varepsilon n}$

$\le \frac{1}{\varepsilon}$?

$C_x + \frac{n - C_x}{b} = \frac{n}{b}$.

$C_x = 0$???

$[(b-1)C_x + n]\varepsilon \overset{?}{\le} C_x + \cancel{\varepsilon n}$



$x = y - r = \frac{2}{\sqrt{3}} - 1 = \frac{2\sqrt{3}}{3} - 1$

$\log N + \log N \underline{\log 2c} \le \log 0.01$

$N \cdot (2c)^{\log N} \le 0.01$

$\log N \cdot \log 2c \le \log\left(\frac{0.01}{N}\right)$

$\log 2c \le \frac{\log\left(\frac{0.01}{N}\right)}{\log N}$.

$2c \le e^{\frac{\log\left(\frac{0.01}{N}\right)}{\log N}}$

$= \left(\frac{0.01}{N}\right)^{\frac{1}{\log N}}$

$N \cdot \left(\frac{2k}{b\varepsilon}\right)^T \le 0.01$

$\log$

$bT = \frac{ck \log N}{\varepsilon}$

$b = \frac{k}{\varepsilon} \quad T = c \log N$.

$b = c \cdot \frac{k}{\varepsilon}, \quad T = \log N$.

$N \cdot \left(\frac{2k}{b\varepsilon}\right)^T \le 0.01$

$N \cdot \left(\frac{2}{c}\right)^{\log N} \le 0.01$

$100N \le \left(\frac{c}{2}\right)^{\log N}$

$\log N \log\left(\frac{c}{2}\right) \ge \log(100N)$

$\log\left(\frac{c}{2}\right) \ge \boxed{\frac{\log 100 + \log N}{\log N}}$.

$\frac{\log 100 + \log N}{\log N} \le 2$. ($N > 100$时)

$\therefore$ 当$N>100$时，只需让

$\log \frac{c}{2} > 2$

即 $c > 2e^2$.

即可满足 $\log \frac{c}{2} \ge \frac{\log 100N}{\log N}$

$\Updownarrow$

$Pr \le 0.01$

Lect 11 # The Second-Moment Method

Thm. (Second-Moment Method)

Let X be a real-valued random variable. Then:

$$Pr[X=0] \le \frac{Var(X)}{(\mathbb{E}(X))^2}$$

Proof. $\mathbb{P}[X=0] \le \mathbb{P}[|X-\mathbb{E}(X)| \ge |\mathbb{E}(X)|] \le \frac{Var(X)}{(\mathbb{E}(X))^2}$ Chebyshev

E.X. Is there likely a 4-clique in $G_{n,p}$?

Random graph on n vertices
Each edge exists with prob p, independently

Thm. There are costants c_1, c_2 so that for sufficiently large n,

① If $p \le c_1 n^{-\frac{2}{3}}$, then $\Pr[G_{n,p}$ contains a $K4] < 0.1$
② If $p \ge c_2 n^{-\frac{2}{3}}$, then $\Pr[G_{n,p}$ contains a $K4] > 0.9$

can change if you change c_1/c_2

Proof. ① Let X be the number of 4-cliques in $G_{n,p}$

$\mathbb{E}[X] = \binom{n}{4} \cdot p^6$.

Markov: $\Pr[X \ge 1] \le \mathbb{E}X \le \binom{n}{4}(c_1 n^{-\frac{2}{3}})^6 \le n^4 c_1^6 \cdot n^{-4} = c_1^6 \le 0.1$ for small enough c_1.

② For $S \in \{1, 2, \cdots, n\}$ with $|S| = 4$, let $X_S = \begin{cases} 1 & S \text{ forms a } K4 \\ 0 & \text{otherwise} \end{cases}$

4 vertices

$X = \sum_S X_S$.

Pretend the X_S are independent for now!

$Var(X) = \sum_S Var(X_S) = \sum_S p^6(1-p^6) \le \binom{n}{4} p^6$

$\mathbb{P}[\text{no } K4] = \mathbb{P}[X=0] \le \frac{Var(X)}{(\mathbb{E}(X))^2} \le \frac{\binom{n}{4}p^6}{(\binom{n}{4}p^6)^2} = \frac{1}{\binom{n}{4}p^6} \le \frac{n^4}{\binom{n}{4}c_2^6} \le \frac{4^4}{c_2^6} \le 0.1$ for large enough c_2.

Fix the proof:

Dependencies: If $|S \cap S'| \le 1$, then $X_S, X_{S'}$ are independent. (no mutual edge)

$Var(X) = \mathbb{E}((\sum_S X_S)^2) - (\mathbb{E}X)^2$

$= \mathbb{E}[\sum_{S,S'} X_S X_{S'}] - [\mathbb{E}X]^2$

$= \sum_S \mathbb{E}X_S^2 + \sum_{\substack{S \ne S' \\ |S \cap S'| \le 1}} \mathbb{E}[X_S X_{S'}] + \sum_{\substack{S \ne S' \\ |S \cap S'| \ge 1}} \mathbb{E}[X_S X_{S'}] - (\mathbb{E}X)^2$

the only difference

$\Theta(n^8)$ terms here $\gg$ $O(n^6)$ terms here

$\lesssim$ what it would be if everything is independent $\ll (\mathbb{E}X)^2$ if c_2 is big.

The Lovasz Local Lemma

Motivation: K-SAT.

k literals per clause

$\varphi = (x_1 \vee \overline{x_2} \vee x_7) \wedge (\overline{x_6} \vee x_2 \vee x_9) \wedge \cdots \wedge (\underbrace{\overline{x_1} \vee x_4 \vee \overline{x_8}})$ m clauses.

Is φ satisfiable? $\exists$ assignment of $x_1, x_2, \cdots$ so that $\varphi = \text{True}$

Probablistic Method.

1. Choose a random assignment, x
2. Show that $Pr[\varphi(x) = \text{True}] > 0$

First try.

Let A_i be the event that the ith clause is unsatisfied.

$\mathbb{P}r[A_i] = p = (\frac{1}{2})^k$

By the union bound, $\mathbb{P}r[\text{none of the } A_i] \geq 1 - mp \geq 0$

If $m < \frac{1}{p}$, there exists a satisfying assignment

If $m > \frac{1}{p}$?

Second try.

Pretend A_i's are independent.

$Pr[\text{none of the } A_i] = (1-p)^m > 0 \Rightarrow$ exists a satisfying assignment ✗

Thm. Lovász Local Lemma (LLL)

Let $A_1, A_2, \cdots, A_m$ be "bad" events s.t. for all i:

- $Pr[A_i] \leq p$.
- A_i is mutually independent of all but d other events.

不满足的 d 个 subset

$\exists S \subseteq \{1,2,\cdots,m\} \setminus \{i\}, |S| \leq d.$

s.t. $\forall T \subseteq \{1,2,\cdots,m\} \setminus \{i\}$.

with $T \cap S = \phi$,

$\mathbb{P}[A_i | \bigcap_{j \in T} A_j] = \mathbb{P}[A_i]$

其它的 subset 都和 A_i 独立.

Then:

(Version I). If $pd \leq \frac{1}{4}$, then

$$\mathbb{P}r[\text{None of the } A_i \text{ occur}] \geq (1-2p)^m > 0$$

(Version II) If $p(d+1) \leq \frac{1}{e}$, then

$$\mathbb{P}r[\text{None of the } A_i \text{ occur}] \geq (1 - \frac{1}{d+1})^m > 0$$

Proof. (Version I)

Lemma. (same conditions as LLL)

For any set $S \subset \{1,2,\cdots,m\}$, and any $i \notin S$, $Pr[A_i | \cap_{j \in S} \overline{A_j}] \leq 2p$

This implies LLL!

$\mathbb{P}[\cap_{i=1}^m \overline{A_i}] = \mathbb{P}[\overline{A_1} | \cap_{j=2}^m \overline{A_j}] \cdot \mathbb{P}[\overline{A_2} | \cap_{j=3}^m \overline{A_j}] \cdot \cdots \cdot \mathbb{P}[\overline{A_m}] \geq (1-2p)^m$ ✓

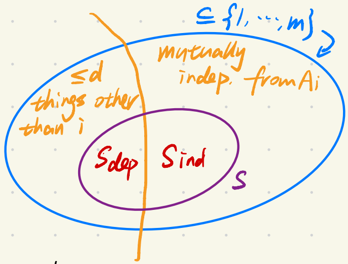
So Let's prove the Lemma using induction on the size of S.

Base case: $|S| = 0$. $\mathbb{P}[A_i | \text{Nothing}] = \mathbb{P}[A_i] \leq p \leq 2p$. ✓

Assume the statement holds for all S with $|S| \leq k$. Now choose S with $|S| = k+1$.

- Case 1: $|S_{ind}| = k+1$ $(S = S_{ind})$

$$\mathbb{P}[A_i \mid \cap_{j\in S} \overline{A_j}] = \mathbb{P}[A_i] \leq p \leq 2p \checkmark$$



- Case 2: $|S_{ind}| \leq k$.

$$\mathbb{P}[A_i \mid \cap_{j\in S} \overline{A_j}] = \frac{\mathbb{P}[A_i \cap (\bigcap_{j\in S_{dep}} \overline{A_j}) \mid \bigcap_{\ell\in S_{ind}} \overline{A_\ell}]}{\mathbb{P}[\bigcap_{j\in S_{dep}} \overline{A_j} \mid \bigcap_{\ell\in S_{ind}} \overline{A_\ell}]}$$

分母 Denominator: $\geq 1 - \underbrace{\sum_{j\in S_{dep}}}_{\substack{\leq d \text{ things} \\ \leq 1/4p \text{ things}}} \underbrace{\mathbb{P}[A_j \mid \cap_{\ell\in S_{ind}} \overline{A_\ell}]}_{\leq 2p \text{ by induction}} \geq 1 - (\frac{1}{4p})\cdot 2p = \frac{1}{2}$

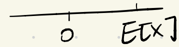
不是Ai是Aj
所以没有独立性.

分子 Numerator: $\leq \mathbb{P}[A_i \mid \cap_{\ell\in S_{ind}} \overline{A_\ell}] = \mathbb{P}[A_i] \leq p$.

$$\therefore \mathbb{P}[A_i \mid \cap_{j\in S} \overline{A_j}] \leq \frac{p}{1/2} = 2p \checkmark$$

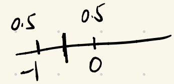
quiz.

$$\Pr[X=0] \le \Pr[|X-E(X)|^2 > |E(X)|^2] \le \frac{E[(X-E(X))^2]}{(E(X))^2}$$



$$\Pr[X=0] \le \Pr[|X-E(X)|^4 > |E(X)|^4] \le \frac{E[(X-E(X))^4]}{(E(X))^4}$$

$$\Pr[X=0] \le \Pr[\underbrace{|X-E(X)|^3}_{\text{positive}} > |E(X)|^3] \qquad \frac{E[(X-E(X))^3]}{E(X)^3}$$



$E(X) = -\frac{1}{2}$

$E[(X-E(X))^3] = 0$

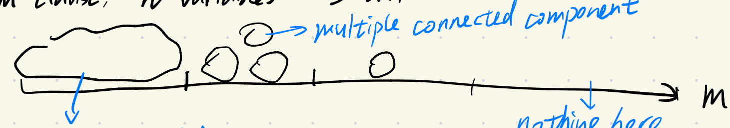
$$E(X_i) = \frac{1}{4} - \frac{3}{4} = -\frac{1}{2}.$$

$$Var(X_i) = \left(\frac{3}{2}\right)^2 \cdot \frac{1}{4} + \left(\frac{1}{2}\right)^2 \cdot \frac{3}{4} = \frac{12}{16} = \frac{3}{4}.$$

$E(X) = -\frac{1}{2}n.$ $\quad Var(X) = \frac{3}{4}n.$ $\qquad \Pr[X=0] \le \frac{Var(X)}{(E(X))^2} = \frac{\frac{3}{4}n}{\frac{1}{4}n^2} = \frac{3}{n}.$

$n \le n_0 \qquad R_4 > n_0.$

m clause, n variables 3-SAT.



exp things that satisfy φ

all connected (edge: flip a single variable)

Lecture 12 — Algorithmic LLL

Notation:

- V is an underlying set of variables.
- $\mathcal{A} = \{A_1, A_2, \cdots, A_m\}$ is a collection of bad events determined by variables.
- $Vbl(A_i)$ is the set of variables involved in A_i
- $\Gamma(A_i)$ is the set of all A_j so that $Vbl(A_i) \cap Vbl(A_j) \neq \phi$

Algorithm:

- Given V and $\mathcal{A}$:
  - choose a random assignment σ_v for each $v \in V$
  - While there's some $A \in \mathcal{A}$ so that $A(\sigma) = 1$ → some bad event happens under this assignment
    - Choose arbitrarily an event A with $A(\sigma) = 1$ ↙ change variables related to that
    - Update σ by re-selecting $\{\sigma_v : v \in Vbl(A)\}$ randomly

<u>Thm.</u> (Algorithmic LLL)

- Let $\mathcal{A}$ be a collection of bad events determined by random variables in V.
- Suppose that for all $A \in \mathcal{A}$.
  - $|\Gamma(A)| \leq d+1$ ← A is independent of all but d other events.
  - $\Pr[A] \leq \frac{1}{e(d+1)}$
- Then, eventually, the algorithm will find an assignment s.t. none of $\mathcal{A}$ occurs.
- The expected number of "re-randomizations" of the algorithm is $O(\frac{|\mathcal{A}|}{d+1})$.

<u>Thm.</u> (Algorithmic LLL) (more general version)

- Let $\mathcal{A}$ be a collection of bad events determined by random variables in V.
- Suppose that there exists a mapping $x : \mathcal{A} \to (0,1)$ s.t. for all $A \in \mathcal{A}$

$$\Pr[A] \leq x(A) \prod_{B \in \Gamma(A) \setminus \{A\}} (1 - x(B))$$

- Then, eventually, the algorithm will find an assignment s.t. none of $\mathcal{A}$ occurs.
  The expected number of "re-randomizations" of the algorithm is bounded by $\sum_{A \in \mathcal{A}} \frac{x(A)}{1-x(A)}$

General version → Simpler version

Set $x(A) = \frac{1}{d+1} \Rightarrow x(A) \prod_{B \in \Gamma(A) \setminus \{A\}} (1 - x(B)) \geq \frac{1}{d+1} \cdot (1 - \frac{1}{d+1})^d \geq \frac{1}{e(d+1)}$

Proof of the Algorithmic LLL

Special case for k-SAT.

φ is a k-CNF formula on n variables with m clauses.

Each clause C in φ shares variable with at most $d+1 = 2^{k-c}$ clauses (including C itself), for some constant c.

Then φ is satisfiable and we can find randomized algorithm in runtime poly in n and m.

Proof idea

$$\text{Formula } \varphi \rightarrow \boxed{\text{Algorithm}} \rightarrow \text{assignment } \sigma$$

random bits $\downarrow$ (into Algorithm); $\downarrow$ print statements (out of Algorithm)

If the algo runs for too long, it'll output $\ll N$ bits worth of info in print and you can recover all random bits from print and there's a contradiction.

Algorithm

- Define global variables $t = 0$ and T.

FindSat ($\varphi = C_1 \wedge \cdots \wedge C_m$):
- Choose random assignment σ
- For each C_i = False:
  - Print "Fixing clause i"
  - $\sigma \leftarrow$ Fix (φ, i, σ)
- Return σ

Fix (φ, i, σ): Recursive
- $t \leftarrow t+1$.
- if $t > T$:
  - print "Too many iterations! My current assignment is σ."
  - Halt.
- Update σ by rerandomize every variable in C_i
- Let $C_{i_1}, C_{i_2}, \ldots, C_{i_{d+1}}$ be the clauses that share variable with C_i
- For $j = 1, \ldots, d+1$ — rerandomization might fail to fix C_i, so C_i is included in this len$(d+1)$ list.
  - If C_{i_j} is violated:
    - Print "Trying to fix the j'th child" — only need $\log(d+1)$ bits
    - $\sigma \leftarrow$ Fix (φ, i_j, σ)
- Print "All done. Moving up a level of recursion"
- Return σ

3位的

Observation:

- Given the print statements, we can recover: 均:
  - the original random vector σ
  - All the random bits flipped by the algorithm.

恢复 $\boxed{n + k \cdot t}$

Claim:

- Consider the first time Fix is called from FindSat.
- The running time of that call is poly(n,m) w.h.p.

Implication

- The running time of the call the FindSat is poly(n,m) w.h.p.

  Union Bound over $\leq m$ calls to fix

- If it finishes, FindSat returns a satisfying assignment.

Proof the claim.

- Suppose the algorithm runs for T steps, and fails to find a sat assignment.
- Random bits in: $\underbrace{n}_{\text{for init } 6} + \underbrace{k \cdot T}_{T \text{ calls, each time flip } k \text{ times}}$
- Print Statements: $\underbrace{\log m}_{\text{"fixing clause } i\text{"}} + T(\underbrace{c}_{\text{"trying to fix ... child"}} + \underbrace{\log(d+1)}_{\text{"j-th"}}) + \underbrace{Tc}_{\text{"all done"}} + \underbrace{c}_{\text{"too many iterations"}} + \underbrace{n}_{\text{current assignment}}$

get a contradiction if:

$$n + kT \gg \log m + T(c + \log(d+1)) + Tc + n + c$$

$$T \gg \frac{\log m + c'}{k - \underbrace{\log(d+1)}_{\text{we've chosen } d+1 = 2^{k-c''}} - c'} = \frac{\log m + c'}{c'' - c'}$$

If it happens likely, then it's unlikely to recover much more random bits from print.

Lect 13 Markov Chains

Def (Markov Chain) X_0, X_1, X_2 is a ~ if for all t, for any sequence of values $c_0, c_1, \cdots$

$$\Pr[X_t = c_t \mid X_0 = c_0, X_1 = c_1, \cdots, X_{t-1} = c_{t-1}] = \Pr[X_t = c_t \mid X_{t-1} = c_{t-1}]$$ can depend on time t

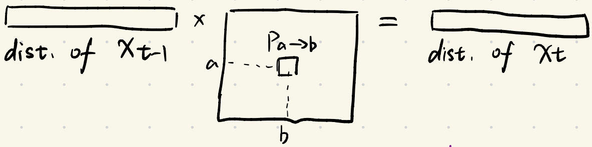
Def. (Time Homogeneous) A Markov Chain is ~ if. for all values a, b, and all $t \geq 0$,

$$\Pr[X_t = a \mid X_{t-1} = b] = P_{b \to a}$$ for some $P_{b\to a} \in [0,1]$ that does not depend on t.

can be represented as Directed Graph

(memorylessness).

Transition Matrix for time homo Markov.



Randomized 2-SAT. (application of Markov Chain)

e.g. $\varphi = (x_1 \vee \overline{x_2}) \wedge (\overline{x_1} \vee x_3) \wedge \cdots$

Algo.
- Pick any assignment σ_0
- For $t = 0, 1, \cdots, cn^2$.
  - if $\varphi(\sigma_t) = 1$. return σ_t
  - choose arbitrarily any unsatisfied clause.
  - suppose x_i and x_j in this clause.
  - Obtain σ_{t+1}: $\begin{cases} \text{with prob} = \frac{1}{2}, \text{ flip } x_i \\ \text{otherwise flip } x_j. \end{cases}$
- return "not satisfiable"

Thm.
- if φ is not satisfiable: algo returns "not satisfiable" with prob 1 ✓
- if φ is satisfiable : algo returns an assignment with prob $\geq 1 - \frac{1}{2^{c/2}}$

Proof. Fix any satis.. assignment σ^*.

For $t = 0, 1, \cdots$; define X_t to be #variables in σ_t that agrees with σ^*.

Claim: $\Pr[X_t = X_{t-1} + 1 \mid \sigma_{t-1}] \geq \frac{1}{2}$

If $X_t \neq X_{t-1} + 1$, then $X_t = X_{t-1} - 1$.

Proof. Suppose $(x_i \vee x_j)$ is unsat. in σ_{t-1}

CASE 1: $\sigma^*(x_i) = T$. $\sigma_{t-1}(x_i) = F$

$\sigma^*(x_j) = T$ $\sigma_{t-1}(x_j) = F$ ✓ $+1$ $p = 1$

CASE 2: $\sigma^*(x_i) = T$

$\sigma^*(x_j) = F$ ✓ $+1$ $p = \frac{1}{2}$

-1 $p = \frac{1}{2}$

Is $X_0, X_1, X_2, \cdots$ a Markov Chain? No!!

有后效性的,想一想.

Claim: There exists a Markov Chain $Y_0, Y_1, \dots$ s.t.

- $Y_0 = X_0$
- For all $t \geq 0$, $Y_t \leq X_t$. $\longrightarrow$ $\underbrace{Y_t \geq n}_{\text{show that this happens w.h.p.}} \Rightarrow X_t \geq n \Leftrightarrow \sigma^* = \sigma_t$
- If $Y_{t-1} \in \{1, \dots, n-1\}$, then $\begin{cases} \Pr[Y_t = Y_{t-1}+1 \mid Y_{t-1}] = \frac{1}{2} \\ \Pr[Y_t = Y_{t-1}-1 \mid Y_{t-1}] = \frac{1}{2} \end{cases}$
- $\Pr[Y_t = 1 \mid Y_{t-1} = 0] = 1$.

$$0 \underset{1/2}{\overset{1}{\rightleftarrows}} 1 \underset{1/2}{\overset{1/2}{\rightleftarrows}} 2 \underset{1/2}{\overset{1/2}{\rightleftarrows}} 3 \cdots$$

Proof. Define $Y_0 = X_0$.

Define Y_t as a function of $\sigma_{t-1}, Y_{t-1}, X_t$:

If $Y_{t-1} = 0$, $Y_t = 1$.

Let $p_t = \Pr[X_t = X_{t-1}+1 \mid \sigma_{t-1}] \geq \frac{1}{2}$.

If $X_t = X_{t-1} - 1$, then $Y_t = Y_{t-1} - 1$

If $X_t = X_{t-1} + 1$, then

with prob $\frac{1}{2p_t}$, set $Y_t = Y_{t-1} + 1$

with prob $1 - \frac{1}{2p_t}$, set $Y_t = Y_{t-1} - 1$.

$\Pr[Y_t = Y_{t-1} - 1 \mid Y_{t-1}]$

$= (1-p_t)\cdot 1 + p_t \cdot \left(1 - \frac{1}{2p_t}\right) = \frac{1}{2}$.

$\Pr[Y_t = Y_{t-1} + 1 \mid Y_{t-1}]$

$= (1-p_t)\cdot 0 + p_t \cdot \frac{1}{2p_t} = \frac{1}{2}$ ✓

Lemma. Suppose $Y_0, \dots$ is this Markov Chain: $0 \underset{1/2}{\overset{1}{\rightleftarrows}} 1 \underset{1/2}{\overset{1/2}{\rightleftarrows}} 2 \underset{1/2}{\overset{1/2}{\rightleftarrows}} \cdots$

Then for all i,

$$\mathbb{E}[\underbrace{\min\{t : Y_t = n\}}_{\text{steps until satisfying}} \mid Y_0 = i] \leq \mathbb{E}[\min\{t : Y_t = n\} \mid Y_0 = 0] = n^2$$

Proof. Let $r_i = \mathbb{E}[\min\{t : Y_t = n\} \mid Y_0 = i]$.

Then
$$r_0 = 1 + r_1$$
$$r_1 = 1 + \tfrac{1}{2} r_0 + \tfrac{1}{2} r_2$$
$$\vdots$$
$$r_{n-1} = 1 + \tfrac{1}{2} r_{n-2} + \tfrac{1}{2} r_n$$
$$r_n = 0$$

Claim. The sol satisfies $r_i = r_{i+1} + 2i + 1$.

Pf. induction.

base: $r_0 = r_1 + 1$ ✓

induction: suppose $r_{i-1} = r_i + 2i - 1$.

$r_i = 1 + \frac{1}{2} r_{i-1} + \frac{1}{2} r_{i+1}$

$= 1 + \frac{1}{2} r_i + i - \frac{1}{2} + \frac{1}{2} r_{i+1}$

$\Rightarrow r_i = 1 + r_{i+1} + 2i$ ✓

$\Rightarrow r_i = r_{i+1} + 2i + 1$

$r_n = 0$

$r_{n-1} = 2(n-1) + 1$

$r_{n-2} = 2(n-1) + 1 + 2(n-2) + 1$

$\vdots$

$r_{n-i} = \sum_{j=1}^{i} [2(n-j)+1]$

$\leq \sum_{j=1}^{n} [2(n-j)+1] = n + 2\sum_{\ell=0}^{n-1} \ell = n^2$. ✓

$\mathbb{E}[\cdots] = n^2$

$\Pr[\cdots \geq 2n^2] < \frac{1}{2}$.

Finishing the proof of Thm:

$\Pr[\text{none of } X_0, \dots, X_{cn^2-1} \text{ are } = n] \leq \Pr\{\text{none of } Y_0, \dots, Y_{cn^2-1} \text{ are } = n\}$

$= \Pr\{\text{none of } Y_0, \dots, Y_{2n^2-1} \text{ are } = n\} \times \Pr\{\text{none of } Y_{2n^2}, \dots, Y_{4n^2-1} \text{ are } = n \mid \underline{Y_{2n^2-1}}\} \times \cdots$

$= \mathbb{P}\{\min(t : Y_t = n) \geq 2n^2\} \times \mathbb{P}\{\min(t : Y_t = n) \geq 2n^2 \mid Y_0 = \quad\} \times \cdots$

$\leq \left(\frac{1}{2}\right)^{c/2}$ ■

$f_0 + 3 = 4$

$f_i = 2^{i+2} - 3$

$1, 5, 13,$

$r_i = 1 + \frac{2}{3} r_{i-1} + \frac{1}{3} r_{i+1}$

$f_0 = 1$

$r_i = 1 + \frac{2}{3}(f_{i-1} + r_i) + \frac{1}{3} r_{i+1}$

$f_i = 2 f_{i-1} + 3$

$f_i + 3 = 2(f_{i-1} + 3)$

$\frac{1}{3} r_i = \cancel{1} + \frac{1}{3} f_i \cancel{- 1} + \frac{1}{3} r_{i+1}$

$r_n = 0$

$r_i = f_i + r_{i+1}$

$r_{n-1} = f_{n-1} + 0$

$r_{n-2} = f_{n-2} + f_{n-1}$

$r_0 = \sum_{i=0}^{n-1} f_i = O(2^n)$

Made with Goodnotes

Lect 14 Some Markov Chain Definitions

Def (Irreducible Markov Chain)

A time homo.. Markov chain $X_0, X_1, \cdots$ is irreducible if for all pairs of state i, j,

这不是state,是第t时刻的state的random variable!!

$$\sum_{t\geq 0} Pr[X_t = j \mid X_0 = i] > 0.$$

即从i走到j是可能的.

irreducible 指可以运行很长时间在任意state都可能

reducible 应该是说运行一会后会困在某些到不了其它地方的 absorbing states.

Def (Transient and Recurrent States)

Let $r_i = \sum_{t\geq 1} Pr[X_t = i, \text{ and } X_s \neq i \text{ for all } 1\leq s<t \mid X_0 = i]$

- A state i is transient if $r_i < 1$. ← there's some prob that we'll never return to that state
- A state i is recurrent if $r_i = 1$.

$\mathbb{E}[\#\text{times we visit } i] = \frac{1}{1-r_i}$ (geometry distribution)

$\mathbb{E}[\min\{t: X_t = i\} \mid X_0 = i]$ 一样的. 不同说法

- if i is transient, ↗ is ∞
- if the Markov Chain is finite, i is recurrent $\Leftrightarrow \mathbb{E}[\#\text{timesteps between visits to } i] < \infty$
- if infinite?

如果是>1维就不行了.

因为x轴到0的时候y不一定是0. 或者看圆里大概有d^2点没法都走完.

不对. 好像也可以???

$P(X=0) \approx \frac{1}{d}$. $P(X=0, y=0) \approx \frac{1}{d^2}$.

$\mathbb{E}[X=0, y=0] = \frac{1}{d^2} \cdot d^2 = c$. 然后再出发.

三维: $\frac{1}{d} + \frac{1}{2d} + \frac{1}{4d} + \cdots$ 半径×2

e.g. $\cdots \leftrightarrows (-1) \underset{1/2}{\overset{1/2}{\leftrightarrows}} (0) \underset{1/2}{\overset{1/2}{\leftrightarrows}} (1) \cdots$ State 0 is recurrent.

Recall the proof in randomized 2-SAT.

$\forall t, \mathbb{P}[\exists t' \in \{t+1, \cdots, t+2n^2\} \text{ s.t. } |X_{t'} - X_t| \geq n] \geq \frac{1}{2}$

很可能游走到n步以外 考虑另一方向

$\Rightarrow$ no matter how far you get (e.g. n), $\mathbb{P}[\text{return to } 0 \text{ after } 2n^2 \text{ steps}] \geq \frac{1}{4}$

$\frac{1}{4}$: return to 0

$\frac{3}{4}$: wander to n'

$\Rightarrow \mathbb{P}[\text{never return to } 0] = 0$

0   n n'

$\mathbb{P}[\text{never return to } 0]$

$= \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdots\cdots = (\frac{3}{4})^\infty = 0$

The Gambler's Ruin

"will eventually return to 0 dollars"

However, $\mathbb{E}[\#\text{timesteps between visits to state } 0] = \infty$. NOT ANALYZING THIS NOW :(

Def (Periodic Markov Chains)

A Markov Chain $X_0, X_1, \cdots$ is periodic if $\exists$ state i s.t.

$$\gcd(\{t : Pr[X_t = i \mid X_0 = i] > 0\}) \neq 1$$

Otherwise it's aperiodic. e.g. Last example $t=0 \Rightarrow \gcd(\cdots) = 2$.

Fact: Any irreducible Markov Chain that has a self-loop is aperiodic. trivial!

(i) ~ (j) suppose j has a self-loop.

∵ irreducible $\gcd(x, x+1, \cdots) = 1$.

Def (Stationary Distributions)

Let π be the distribution on the states with transition matrix P.

π is a ~ if $\pi P = \pi$. (distribution on X_{t+1} equals X_t).

Fundamental Theorem of Markov Chains

Thm. Let $X_0, X_1, \dots$ denote a Markov Chain so that: 不在课上证了.

① the state space is finite

② the chain is irreducible

③ the chain is aperiodic

Then $\exists$ unique stationary distribution π s.t.

- For any state $i \neq j$, $\lim_{t\to\infty} \Pr[X_t = i \mid X_0 = j] = \pi_i$

  (π_0 任意) no matter where we started. when $t\to\infty$, dist. of $X_t \to \pi$

- For any state i, $\pi_i = \dfrac{1}{\mathbb{E}[\min_{t>0}\{t : X_t = i\} \mid X_0 = i]}$ Why? 怎么从直觉上理解?

How to understand the assumptions?

① if state space is infinite. but chain is irreducible & aperiodic

Either there's a stationary dist. s.t. $\lim_{t\to\infty} \Pr[X_t = i \mid X_0 = j] = \pi_i > 0$ for all i,j ✓

how? add up to >1?

Or $\lim_{t\to\infty} \Pr[X_t = i \mid X_0 = j] = 0$ for all i,j

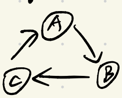
② if chain is not irreducible

$\Rightarrow$ decompose it into irreducible components.

chain will eventually stuck in one of the component

each of these components (if aperiodic) will have their own dist.

③ if chain is aperiodic



There's still a unique stationary dist. π

still have $\pi_i = \dfrac{1}{\mathbb{E}[\min\{t : X_t = i\} \mid X_0 = i]}$.

But! no longer the case that

$\lim_{t\to\infty} \Pr[X_t = i \mid X_0 = j] = \pi_i$ not converging! t mod gcd

Prop. Suppose $X_0, X_1 \cdots$ is an irreducible, aperiodic Markov Chain with a finite state space.

If the transitions are symmetric then the stationary dist. is uniform.

Proof. enough to show π (uniform dist.) satisfy $\pi P = \pi$.

$$(\pi P)_i = \sum_j \pi_j P_{ji} = \frac{1}{|S|} \cdot \sum_j P_{ji} = \frac{1}{|S|} \underbrace{\sum_j P_{ij}}_{\sum \text{ prob of state } i \to \text{ state } j = 1} = \frac{1}{|S|} \cdot 1 = \pi_i$$

E.X. shuffle: n cards.

Repeat:
- Choose 2 cards independently, with replacement
- Swap them.

This is a aperiodic, irreducible MC with finite state space.

∵ self-loop if chosen same cards

∵ can get to any ordering of the n cards

all the possible ordering $n!$

X_t = order of the n cards after t steps.

⇒ Then the stationary dist. is uniform!

As $t \to \infty$, dist. on orderings → uniform fair game.

Prop. Let $X_0, X_1, \cdots$ be a random walk on a connected, undirected, non-bipartite graph $G=(V,E)$ (每一次 uniformly 选 X_t 的 neighbor)

Then there's a unique stationary dist. π s.t. $\pi_v = \frac{\deg(v)}{2|E|}$ 只取决于degree!

Proof. All we need to show is $\pi P = \pi$.

$$\pi \cdot P = \pi P, \quad P_{uv} = \begin{cases} 1/\deg(u) & \text{if } (u,v)\in E \\ 0 & \text{otherwise} \end{cases}$$

$$(\pi P)_v = \sum_u P_{uv}\pi_u = \sum_{\substack{u \text{ s.t.} \\ (u,v)\in E}} \left(\frac{1}{\deg(u)}\right)\cdot\left(\frac{\deg(u)}{2|E|}\right)$$

$$= \frac{\deg(v)}{2|E|} = \pi_v \ ✓$$

connected → irreducible ✓

undirected & non-bipartite → aperiodic

exists an odd cycle?

Markov Chain Monte Carlo

Problem: How to uniformly sample from a proper k-coloring of G (given Graph G)?

Strategy: MCMC. Construct a Markov Chain whose stationary distribution is the dist. you care about. Then run it for a while.

Example..

Markov Chain: if X_{t-1} is a coloring, get to X_t by:
- choose a random vertex and color it a random color
- if that's a legit coloring, do that
- if not, stay where you are

⇒ symmetric, aperiodic, irreducible if $k \ge [\text{max degree}]+2$ ✓ ⇒ uniform!

easy to check ✓ self loop ✓

左染 $d+1$ 右冲突改 $d+2$. 左改目标色.

Example. Sampling funkier dist.

Graph G. choose a random proper k-coloring of G s.t. a coloring with $<k$ colors is twice as likely as a coloring with k colors.

Metropolis Algorithm.

If we want to sample from dist. π.

Suppose we have connected $G=(V,E)$, where

- V is the support of π
- It's computationally easy to find a vertex v's neighbor
- If $(u,v)\in E$, we can compute $\frac{\pi(v)}{\pi(u)}$

Consider MC with states V and transition matrix P:

$$P_{ij}=\begin{cases} 0 & \{i,j\}\notin E \\ \frac{1}{d}\min(1, \pi(j)/\pi(i)) & \{i,j\}\in E,\ i\neq j \\ 1-\sum_{l\neq i} P_{il} & i=j \end{cases}$$

d = any constant that's larger than max-deg(G)

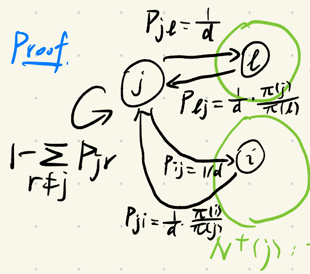
Thm. The MC is irreducible, aperiodic, and has stationary dist. π.

irreducible: G connected

aperiodic: $\because d>$ max-deg $\therefore P_{ii}>0$, self loop

stationary dist.: want to show $\pi P=\pi$.

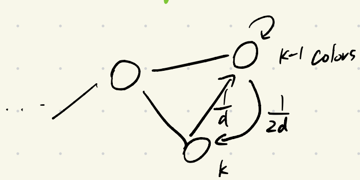
Proof.



break j's neighbor into two sets.

$$(\pi P)_j=\sum_{l\in N^-(j)}\pi(l)\frac{1}{d}\left(\frac{\pi(j)}{\pi(l)}\right)+\sum_{i\in N^+(j)}\pi(i)\cdot\frac{1}{d}$$
$$+\pi(j)\cdot\left(1-\sum_{l\in N^-(j)}\frac{1}{d}-\sum_{i\in N^+(j)}\frac{1}{d}\cdot\frac{\pi(i)}{\pi(j)}\right)=\pi(j)\ \checkmark$$

Return to example.



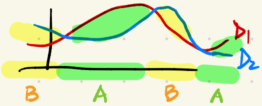
$d>k\cdot n$

Question: How long?

Total variation distance

Def. (Total variation distance) Suppose D_1 and D_2 are dist. with support on some countable set S. The ~ between D_1 and D_2 is:

$$\|D_1 - D_2\| = \underbrace{\frac{1}{2}\sum_{s\in S}|D_1(s)-D_2(s)|}_{\text{half of the } L_1 \text{ norm}} = \underbrace{\max_{A\subseteq S}\left(\Pr_{D_1}[A] - \Pr_{D_2}[A]\right)}_{A:\text{ an event}}$$



$\|D_1 - D_2\| =$ (green) $=$ (yellow)

$= \mathbb{P}_{D_1}(A) - \mathbb{P}_{D_2}(A)$

Fact. Let J be a joint distribution over X and Y. where the marginal dist. of X is D_1, and the marginal dist of Y is D_2. Then $\|D_1 - D_2\| \le \Pr[X \ne Y]$

Further, $\exists J^*$ s.t. equality holds.

e.g. D_1: fair coin flip, D_2: heads with prob 0.6

$\mathbb{P}[X\ne Y] = 0.1$

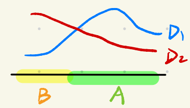
$\|D_1 - D_2\| = \frac{1}{2}(|0.5-0.6| + |0.5-0.4|) = 0.1$

	Y=H	Y=T	D_1
X=H	0.5	0	⇒ 0.5
X=T	0.1	0.4	⇒ 0.5
D_2	⇓ 0.6	⇓ 0.4	

Note the diff. between X, Y and D_1, D_2

Proof. Let $A = \{x: D_1(x) \ge D_2(x)\}$, $B = \{x: D_1(x) < D_2(x)\}$

Define $p = \sum_x \min(D_1(x), D_2(x))$.



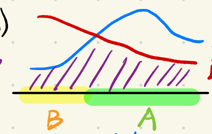
△ **Claim:** $1 - p = \|D_1 - D_2\|$

$\because \|D_1 - D_2\| = \sum_{x\in A}(D_1(x) - D_2(x))$

$1 - p = 1 - \sum_{x\in A} D_2(x) - \sum_{x\in B} D_1(x)$

$= \left(\sum_{x\in A} D_1(x) + \cancel{\sum_{x\in B} D_1(x)}\right) - \sum_{x\in A} D_2(x) - \cancel{\sum_{x\in B} D(x)}$

$= \sum_{x\in A}(D_1(x) - D_2(x)) = \|D_1 - D_2\|$ ✓.

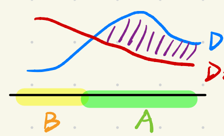


△ **Construct a dist. J^* on (X,Y) s.t. $\Pr[X\ne Y] = 1-p$.**

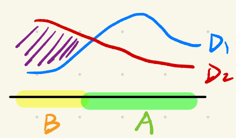
with prob. p: Draw $x \in S$ w/ prob $\frac{1}{p}\cdot\min(D_1(x), D_2(x))$ ⇒ does sum to 1! valid dist.

Set $X = Y = x$.

with prob $1-p$: $X \leftarrow x$ with prob $\begin{cases}\propto D_1(x) - D_2(x) & \text{if } x\in A\\ 0 & \text{otherwise}\end{cases}$



$Y \leftarrow y$ with prob $\begin{cases}\propto D_2(y) - D_1(y) & \text{if } y\in B\\ 0 & \text{otherwise}\end{cases}$



Check: marginal for X, Y is indeed D_1, D_2!

and $\Pr[X\ne Y] = 1-p$.

△ **Any J on (X,Y) has $\Pr[X\ne Y] \ge 1-p$:**

In J^*, $\forall x$ $\Pr[X = Y = x] = \min(D_1(x), D_2(x))$. according to def

If we make this any larger, we'll change the marginals.

e.g. $\mathbb{P}[X = Y = x] \ge \min(D_1(x), D_2(x)) + \varepsilon$ ← WLOG, $D_1(x)$ is the min

$\mathbb{P}[X = x] = \mathbb{P}[X = Y = x] + \mathbb{P}[X = x \wedge Y \ne x] \ge D_1(x) + \varepsilon$

$\therefore \mathbb{P}[X=Y=x]$ can't be made any larger for any x.
$\Rightarrow \mathbb{P}[X \neq Y]$ can't be made any smaller.

Mixing Times

Def. (Distance to Stationary Distribution)

Let $X_0, X_1 \cdots$ be a finite irreducible aperiodic MC with stationary dist. π

Let $\Delta(t) = \max_s \|\pi - P_s^t\|$, where P_s^t is the dist. of $X_t | X_0 = s$ iterate all starting points.

Def. (Mixing times) 命名意思是不同开局dist.开始混合(都接近π)的时间?

The ~ of a MC $X_0, X_1, \cdots$ with stationary dist. π, is denoted as τ_{mix}.

$$\tau_{mix} = \min\{t : \Delta(t) \leq \frac{1}{2e}\}$$

how fast an MC approaches π.

Fact (not obvious):

- For any f⋯, i⋯, a⋯ MC:
  $\Delta(t)$ is non-increasing
  $\Delta(c \cdot \tau_{mix}) \leq e^{-c}$

Coupling

Fact. $\Delta(t) = \max_s \|P_s^t - \pi\| \leq \max_{s,s'} \|P_s^t - P_{s'}^t\| \leq 2\Delta(t)$

(dist of $X_t | X_0 = s$) (之后分析这个)

Proof. $\pi = \sum_w \pi(w) P_w^t$ start from dist. π, run MC for t steps $\Rightarrow$ dist. π

$$\Rightarrow \max_s \|P_s^t - \pi\| = \max_s \|P_s^t - \underbrace{\sum_w \pi(w) P_w^t}_{\text{weighted average}}\|$$

$$\leq \max_{s,s'} \|P_s^t - \underline{P_{s'}^t}\|$$ 可以取s'让它更大.

$$\|P_s^t - P_{s'}^t\| \leq \|P_s^t - \pi\| + \|\pi - P_{s'}^t\| \leq 2\Delta(t)$$

Idea Behind Couplings

Bound $\max_{s,s'} \|P_s^t - P_{s'}^t\|$

- For any s, s' construct a joint dist. (X_t, Y_t)
- The marginal dist. are $X_t \sim P_s^t$, $Y_t \sim P_{s'}^t$
- $\Pr[X_t \neq Y_t]$ is small $\Rightarrow$ 上一节里 $\|D_1 - D_2\| \leq \Pr[X \neq Y]$.

Def. (Coupling)

Given a Markov Chain with transition matrix P, a ~ is a joint process $(X_0, Y_0)(X_1, Y_1), \cdots$ s.t.

① $\forall s, s'$, $\mathbb{P}[X_t = s | X_{t-1} = s'] = \mathbb{P}[Y_t = s | Y_{t-1} = s'] = P_{s,s'}$

② $\forall t$, if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$

Prop. Given a coupling $(X_0, Y_0), (X_1, Y_1) \cdots$ let

$$T_{s,s'} = \min\{t : X_t = Y_t \mid X_0 = s, Y_0 = s'\}.$$

Then $\Delta(t) \leq \max_{s,s'} \Pr[T_{s,s'} > t]$

$\because$ if marginal of X, Y is D_1, D_2 then $\|D_1 - D_2\| \leq \Pr[X \neq Y]$.

Proof. $\Delta(t) \leq \max_{s,s'} \|P_s^t - P_{s'}^t\| \leq \max_{s,s'} \mathbb{P}[X_t \neq Y_t | X_0 = s, Y_0 = s'] \leq \max_{s,s'} \mathbb{P}[T_{s,s'} > t]$

E.X. sample uniformly random k-coloring of G. ↗ through this prop

Strategy via Coupling: { come up with a coupling that "couple"s ASAP
this will imply $\Delta(t)$ is small for reasonably small values of t.
aka. the mixing time is small.

Thm:

Let $G=(V,E)$ be a graph with n vertices and max-deg d. no matter where the starting point is if we run t steps, the dist. is not far from the uniform dist. we wanted. ($\Delta t \le \frac{1}{2e}$)

Suppose $k \ge 4d+1$

Then the mixing time of MC is bounded by $kn(2+\log n)$

Our coupling:

X and Y starts at different colorings but follows the same decision. (pick randomly the same vertex and the same color)

Analyze when X_t and Y_t are likely to collide

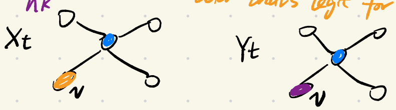
Let Z_t be # vertices on which X_t and Y_t differ. $Z_{t+1} = \begin{cases} Z_t \\ Z_t - 1 \\ Z_t + 1 \end{cases}$

- When does $Z_{t+1} = Z_t - 1$? Choose a vertex on which X_t and Y_t differ Z_t such vertices

  $\mathbb{P}[Z_{t+1} = Z_t - 1 \mid Z_t] \ge \frac{Z_t(k-2d)}{nk}$ Choose a coloring that's legit for both colorings at most $2d$ colors that are not legit.

- When does $Z_{t+1} = Z_t + 1$? Choose a vertex on which X_t and Y_t are the same. Pick v then choose its neighbor

  $\mathbb{P}[Z_{t+1} = Z_t + 1 \mid Z_t] \le \frac{Z_t \cdot 2d}{nk}$ Choose a color that's legit for one coloring but not for the other. color should be one of the two colors of v.



$\le Z_t \cdot d \cdot 2$

$$\mathbb{E}[Z_{t+1} \mid Z_t] \le Z_t + \frac{2d Z_t}{nk} - \frac{Z_t(k-2d)}{nk} = Z_t\left[1 - \frac{k-4d}{nk}\right] \quad \because k \ge 4d+1$$
$$\le Z_t[1 - 1/nk]$$

$$\mathbb{E}[Z_{t+1}] \le \mathbb{E}[Z_t](1 - 1/kn)$$
$$\le \mathbb{E}[Z_0](1 - 1/kn)^{t+1} \le n \cdot \exp\left(-\frac{t+1}{kn}\right) \quad \because \mathbb{E}[Z_0] \le n. \quad (1-x) \le \exp(-x)$$

$\Rightarrow \mathbb{E}[Z_t] \le n \cdot \exp\left(\frac{-t}{kn}\right)$

If $t \ge kn(2+\log n)$, then $\mathbb{E}[Z_t] \le \frac{1}{e^2} \le \frac{1}{2e}$.

$\Rightarrow \Delta(t) \le \max_{s,s'} \mathbb{P}[T_{s,s'} > t] \le \mathbb{P}[Z_t \ge 1] \le \mathbb{E}[Z_t] \le \frac{1}{2e}$.

By the def of mixing time, $\tau_{mix} \le kn(2+\log n)$

Martingales

Def. (Martingale) Let $Z_0, Z_1, \cdots$ be a sequence of real-valued random variables.
$X_0, X_1, \cdots$ random variables.

$\{Z_t\}$ is a martingale with respect to $\{X_t\}$ if for $\forall t$:

① Z_t is a function of $X_0, \cdots, X_t$

② $\mathbb{E}[|Z_t|] < \infty$

③ $\mathbb{E}[Z_t \mid X_0, \cdots, X_{t-1}] = Z_{t-1}$

Ex. X_i = Head or Tail. Head: +1 Tail: −1
Z_t = amount of money at time t.

Def. Let $Y_t = Z_t - Z_{t-1}$. $\{Y_t\}$ is called martingale differences.
If $\{X_t\}$ is a martingale w.r.t. itself, we just say $\{X_t\}$ is a martingale.

Def. (Doob Martingale) Let A be a random variable.
Let $\{X_t\}$ be a sequence of random variable.
The ~ of A w.r.t. $\{X_t\}$ is the sequence $\{Z_t\}$ given by:
$$Z_t = \mathbb{E}[A \mid X_0, \cdots, X_t]$$

A是固定的.
逐渐给了 $X_0 \cdots X_t$ 的信息算期望.
e.g. $A = \sum_{i=1}^{n} X_i$ ✓
$Z_t = \sum_{i=1}^{t} X_i$ ✗ 不一定是 martingale

Proof. ① ✓ ② Assume it's true first.
③ $\mathbb{E}[Z_t \mid X_0, \cdots X_{t-1}] = \mathbb{E}[\,\mathbb{E}[A \mid X_0, \cdots X_t] \mid X_0, \cdots, X_{t-1}]$
$= \mathbb{E}[A \mid X_0, \cdots, X_{t-1}] = Z_{t-1}$

$\forall x, y, z$, $\mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Y] = \mathbb{E}[X \mid Y]$. the outer expectation integrate away Z.

E.X: Balls and Bins.
m balls into n bins.
A = #empty bins after m balls have been dropped.
Let X_t = location of t^{th} ball.
$Z_t = \mathbb{E}[A \mid X_1, \cdots, X_t]$.
$Z_0 = \mathbb{E}[A]$. → expected value
$\vdots$
$Z_m = \mathbb{E}[A \mid X_1, \cdots, X_m]$ → actual number

E.X. Edge Exposure Martingale
$G \sim G_{n,p}$.
A = the chromatic number of G ↗ 最少几种颜色染点.
$X_1, \cdots X_{\binom{n}{2}}$: $X_t = \mathbb{1}\{\text{edge } t \text{ exists}\}$.
$Z_t = \mathbb{E}[A \mid X_1, \cdots, X_t]$.
gradually expose each edge.

E.X. Vertex Exposure Martingale.
$Y_1, \cdots, Y_n$: Y_t = neighborhood of node i.
e.g. Y_t can take on 2^{n-1} possible value.
$Z_t = \mathbb{E}[A \mid Y_1, \cdots, Y_t]$. gradually expose each vertex

Azuma-Hoeffding Bound

Thm. Let $\{Z_t\}$ be a martingale w.r.t $\{X_t\}$. suppose there are constants $c_1, \cdots c_n$ s.t. $|Z_i - Z_{i-1}| \le c_i$. $\forall i$

For $\forall \lambda > 0$, $\Pr[\,|Z_n - Z_0| \ge \lambda] \le 2 \cdot \exp\left(\frac{-\lambda^2}{2\sum_i c_i^2}\right)$

Lemma. Let $Y \in [-c, c]$ s.t. $\mathbb{E}[Y] = 0$. Then for all t, $\mathbb{E}[e^{tY}] \le e^{t^2c^2/2}$ ✓

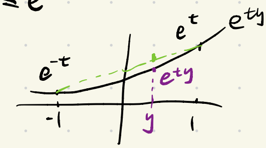
Proof. Enough to prove it for $c=1$. (Define $Y' = \frac{1}{c}Y$. Then $Y' \in [-1,1]$. $\mathbb{E}[e^{tY'}] \le e^{t^2/2}$ $\forall t$

$\Rightarrow \mathbb{E}[e^{tY/c}] \le e^{t^2/2}$

$\Rightarrow \mathbb{E}[e^{sY}] \le e^{s^2c^2/2}$ $s \leftarrow t/c$

Next. assume $c=1$. Try to proof $\mathbb{E}[e^{tY}] \le e^{t^2/2}$

Claim $\mathbb{E}[e^{tY}] \le \frac{e^{-t}+e^{t}}{2} \le e^{t^2/2}$



$\forall y \in [-1,1]$. $e^{ty} \le \left(\frac{1-y}{2}\right)e^{-t} + \left(\frac{1+y}{2}\right)e^{t}$ by convexity of e^{ty}

$\mathbb{E}[e^{ty}] \le \mathbb{E}\left[\left(\frac{1-Y}{2}\right)e^{-t} + \left(\frac{1+Y}{2}\right)e^{t}\right] = \frac{1}{2}(e^{-t}+e^{t})$.

Taylor Expansion time!

$\frac{1}{2}(e^t + e^{-t}) = \frac{1}{2}[(1+t+t^2/2+t^3/3!+\cdots)+(1-t+t^2/2-t^3/3!+\cdots)]$

$\wedge\!|$ $= 1 + t^2/2 + t^4/4! + t^6/6! + \cdots$

$e^{t^2/2} = \sum_{j\ge 0} \frac{t^{2j}}{2^j \cdot j!} = 1 + t^2/2 + t^4/4\cdot 2 + t^6/2^3\cdot 3! + \cdots$ compare

Proof thm. the other direction is the same

$\mathbb{P}[Z_n - Z_0 \ge \lambda] = \mathbb{P}[e^{t(Z_n - Z_0)} \ge e^{t\lambda}]$

$\le \mathbb{E}[e^{t(Z_n-Z_0)}]\, e^{-t\lambda}$. $\forall t > 0$

$\mathbb{P}[X > 1] \le \mathbb{E}[X]$. ✓

$\mathbb{E}[e^{t(Z_n-Z_0)}] = \mathbb{E}[e^{t[Z_n - Z_{n-1} + Z_{n-1} - Z_0]}]$

constant!

$= \mathbb{E}[\,\mathbb{E}[e^{t[Z_n - Z_{n-1} + Z_{n-1} - Z_0]} \mid X_1, \cdots, X_{n-1}]]$

$= \mathbb{E}\left[\boxed{\mathbb{E}[e^{t(\underbrace{Z_n - Z_{n-1}}_{Y})} \mid X_1, \cdots, X_{n-1}]} \cdot e^{t(Z_{n-1}-Z_0)}\right]$ by def of martingale, this has zero mean.

$\le \mathbb{E}[e^{t^2c_n^2/2} \cdot e^{t(Z_{n-1}-Z_0)}]$

$= e^{t^2c_n^2/2} \cdot \mathbb{E}[e^{t(Z_{n-1}-Z_0)}]$.

$\Rightarrow \mathbb{E}[e^{t(Z_n - Z_0)}] \le \exp\left(t^2 \cdot \frac{\sum c_i^2}{2}\right)$

$\mathbb{P}[Z_n - Z_0 \ge \lambda] \le \exp\left(\frac{t^2 \sum c_i^2}{2} - t\lambda\right)$

Choose $t = \lambda / \sum_i c_i^2$. and we get the upper tail.

$\exp\left(\frac{\lambda^2}{2\sum_i c_i^2} - \frac{\lambda^2}{\sum_i c_i^2}\right) = \underset{\uparrow}{2}\exp\left(\frac{-\lambda^2}{2\sum_i c_i^2}\right)$.

two directions. $\begin{cases} Z_n - Z_0 \ge \lambda \\ Z_0 - Z_n \ge \lambda \end{cases}$.

$\Rightarrow 2\exp\left(\frac{-\lambda^2}{2\sum_i c_i^2}\right)$

Plan.

2 week.

① Lect ② Class ④ Homework.

1	2	3	4	5	6	7
					1-3	4-6
7-9	10-12	13-15	16-18	practice exam & prepare cheat sheet & Quiz		
	exam 3:30-6:30pm.					

① Mini Lect & Lect Notes
② Class Agenda
③ Quiz
④ Homework
⑤ Practice Exams.

Exam:

Will be the most imporcant things to remember in the Lecture
Slightly easier than the practice exam.

Lect 17 — Stopping Times and the Martingale Stopping Theorem

Observation Suppose $Z_0, Z_1, \cdots$ is a martingale w.r.t. $X_0, X_1, \cdots$

Then for any t, $\mathbb{E}[Z_t] = \mathbb{E}[Z_0]$.

Note if T is a random variable, it's not true that $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$.

Def. (Stopping time) ↙ does not need to be martingale!!

Given a discrete time process $\{X_t\}$, an integer random variable T is called a ~ for $\{X_t\}$.
if the event $T=i$ is mutually independent of all of the events $\{X_j \mid X_0, \cdots, X_i\}$ for $\forall j > i$

Thm. (Martingale Stopping Theorm)

主要的问题是T取值可能有无穷个
就会影响 $\mathbb{E}[Z_T]$ ← weighted average

Let $\{Z_t\}$ be a martingale w.r.t. $\{X_t\}$.
Let T be a stopping time for $\{X_t\}$.
Then $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$ if any of the following hold:

① $\exists$ constant c s.t. $\forall i$, $|Z_i| < c$

② $\exists$ constant c s.t. with prob 1, $T < c$

③ $\exists$ constant c s.t. $\forall i$, $\mathbb{E}[|Z_{i+1} - Z_i| \mid X_0, \cdots, X_i] < c$, and $\mathbb{E}[T] < \infty$.

e.g. Let T be the first time t s.t. $|Z_t| = 10$.

Let $\tilde{Z}_t = \begin{cases} Z_t & t \le T \\ 10 & t > T \end{cases}$ $\quad \mathbb{E}[\tilde{Z}_T] = \mathbb{E}[Z_T] = \mathbb{E}[Z_0]$

↓
satisfy ① ($c=11$)

check: $\mathbb{E}[Z_T] = \frac{1}{2} \times 10 + \frac{1}{2} \times (-10) = 0$.

Let T' be the first time t s.t. $Z_t = 10$.
None of 1, 2, 3 applies!
(because we can wander to $-\infty$ without ever hitting $+10$)

Hitting Times of Random Walks (application of stopping thm)

E.X. A random walk.

$$\cdots \; (-2) \underset{1/2}{\overset{1/2}{\rightleftharpoons}} (-1) \underset{1/2}{\overset{1/2}{\rightleftharpoons}} (0) \underset{1/2}{\overset{1/2}{\rightleftharpoons}} (1) \underset{1/2}{\overset{1/2}{\rightleftharpoons}} (2) \; \cdots$$

$Z_0 = 0$.

Let T be the first time that Z_t reaches either $-a$ or b.

☆ What's $\Pr[Z_T = b]$? $\frac{a}{a+b}$

$\tilde{Z}_i$ that stops at T ↙

Martingale Stopping Thm ① applies. $|\tilde{Z}_i| < \max\{a, b\}$.

$\Rightarrow \mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0. = \mathbb{P}[Z_T = b] \cdot b + \mathbb{P}[Z_T = -a] \cdot (-a)$

$\Rightarrow 0 = \mathbb{P}[Z_T = b] \cdot (b+a) - a \Rightarrow \mathbb{P}[Z_T = b] = \frac{a}{a+b}$.

☆ What's $\mathbb{E}[T]$? $a \cdot b$

Define a new martingale: $Y_t = Z_t^2 - t$.

① Prove that $\{Y_t\}$ is a martingale w.r.t. $\{Z_t\}$.

$$\mathbb{E}[Y_t \mid Z_0, \cdots, Z_{t-1}] = \tfrac{1}{2}(Z_{t-1}+1)^2 + \tfrac{1}{2}(Z_{t-1}-1)^2 - t$$
$$= Z_{t-1}^2 - (t-1) = Y_{t-1} \quad \checkmark$$

② The Martingale Stopping Thm 3 applies to Y_t.

Claim: $\mathbb{E}[T] < \infty$. $\begin{cases} \text{At any } Z_t,\ \mathbb{P}[\text{stop after } a+b \text{ additional steps}] \geq \frac{1}{2^{a+b}} \\ \Rightarrow \mathbb{E}[T] \leq 2^{a+b}\cdot(a+b) \end{cases}$ (geometric dist.) (Every $a+b$ steps, prob. $\frac{1}{2^{a+b}}$).

$$\mathbb{E}\left[|Y_{i+1} - Y_i| \,\middle|\, Z_0, \cdots, Z_i\right] \leq 1 + 2|Z_i| \leq 1 + 2\cdot\max\{a,b\} \quad \checkmark$$

Computing $\mathbb{E}[T]$:

$$\mathbb{P}[Z_T = b] = \frac{a}{a+b}.$$
$$0 = \mathbb{E}[Y_0] = \mathbb{E}[Y_T] = \mathbb{E}[Z_T^2] - \mathbb{E}[T].$$
$$\mathbb{E}[T] = \mathbb{E}[Z_T^2] = \frac{b}{a+b}\cdot a^2 + \frac{a}{a+b}\cdot b^2 = a\cdot b \quad \text{对!}$$

E.x. Another random walk.

$$\begin{tikzcd} \cdots & -2 \arrow[r, shift left, "p"] & -1 \arrow[l, shift left, "1-p"] \arrow[r, shift left, "p"] & 0 \arrow[l, shift left, "1-p"] \arrow[r, shift left, "p"] & 1 \arrow[l, shift left, "1-p"] \arrow[r, shift left, "p"] & 2 \arrow[l, shift left, "1-p"] & \cdots \end{tikzcd}$$

$Z_0 = 0$. T is the first time that X_t reaches either $-a / b$.

Thm. Say $p \neq \frac{1}{2}$. then:

- $\Pr[Z_T = -a] = 1 - \Pr[Z_T = b] = \dfrac{1-c^b}{c^{-a} - c^{-b}}$, where $c = \frac{1}{p} - 1$. 记作 P_a 吧.
- $\mathbb{E}[T] = \dfrac{-a\cdot P_a + b\cdot(1-P_a)}{2p-1}$

Proof. $\{Z_t\}$ isn't a martingale!

Let $Y_t = c^{Z_t}$ for $c = ?$ ~~X~~ $\frac{1}{p} - 1$

$$\mathbb{E}[Y_t \mid Z_0, \cdots, Z_{t-1}] = p(c^{Z_{t-1}+1}) + (1-p)(c^{Z_{t-1}-1})$$
$$= c^{Z_{t-1}}\underbrace{\left(cp + \frac{1-p}{c}\right)}_{\text{this should be } 1.}$$

Now apply the martingale stopping theorem: Why it applies? 1 applies

$$1 = \mathbb{E}[Y_0] = \mathbb{E}[Y_T] = \mathbb{P}[Z_T = -a]\cdot c^{-a} + \mathbb{P}[Z_T = b]\cdot c^b$$
$$1 = \mathbb{P}[Z_T = -a]\cdot(c^{-a} - c^b) + c^b$$
$$\Rightarrow \mathbb{P}[Z_T = -a] = \frac{1-c^b}{c^{-a} - c^b} \quad \checkmark$$

Define $Q_t = Z_t - \alpha t$ for $\alpha = ?$ $2p-1$

$$\mathbb{E}[Q_t \mid Z_0, \cdots, Z_{t-1}] = p(Z_{t-1}+1) + (1-p)(Z_{t-1}-1) - \alpha t$$
$$= Z_{t-1} + \underbrace{(2p - 1 - \alpha t)}_{\text{want this to be } -\alpha(t-1)}$$

Now apply the martingale stopping theorem: Why it applies? 3 applies

$$0 = \mathbb{E}[Q_0] = \mathbb{E}[Q_T] = \mathbb{P}[Z_T = -a]\cdot(-a - \alpha\mathbb{E}[T]) + \mathbb{P}[Z_T = b]\cdot(b - \alpha\mathbb{E}[T]).$$

Solve for $\mathbb{E}[T]$... $\alpha\cdot\mathbb{E}[T] = \mathbb{P}[Z_T = -a]\cdot(-a) + (1 - \mathbb{P}[Z_T = -a])\cdot b$

$$\Rightarrow (2p-1)\mathbb{E}[T] = \mathbb{P}[Z_T = -a]\cdot(-a) + (1 - \mathbb{P}[Z_T = -a])\cdot b \quad \checkmark$$

13. 证明
17 class work.
18 class work.
☆ 19 class
compare bounds & Poisson