

Discrepancy Minimization
via Gaussian random walks

Lecture 1

Probability

Events, probabilities,
conditional probabilities,
union bound

Random variables,
expected value,
linearity of expectation
Markov's inequality

Algorithms

Approx. 3-SAT,

Big idea: we only want the average

$$\begin{aligned} \text{Linearity of Expectation} \\ E[X+Y] &= E[X] + E[Y] \\ \text{"average sum = sum of averages"} \end{aligned}$$

let $Y_i = \begin{cases} 1 & \text{if } i\text{th clause satisfied} \\ 0 & \text{otherwise} \end{cases}$ ($i=1, \dots, m$)

$$E[Y_i] = 7/8 \quad (1 \text{ clause case})$$

$$E[\text{\# clauses satisfied}] = E\left[\sum_{i=1}^m Y_i\right] = \sum_{i=1}^m E[Y_i] = 7/8 m$$

let $Y_i = \begin{cases} 1 & \text{if } i\text{th clause satisfied} \\ 0 & \text{otherwise} \end{cases}$

$$Z = \sum_{i=1}^m Y_i = \text{\# clauses satisfied}$$

$$E[Z] = \frac{1}{2} E[Z | x_1 = \text{true}] + \frac{1}{2} E[Z | x_1 = \text{false}]$$

$\Rightarrow$ either $E[Z | x_1 = \text{true}]$ or $E[Z | x_1 = \text{false}]$ is $\geq E[Z]$!

QuickSort,

Random divide and conquer

① pick random 'pivot'



② Divide A into two parts



Expected running time of QuickSort

For each $i, j \in [n]$ w/ $i < j$,

$X_{ij} = \begin{cases} 1 & \text{if } i\text{th smallest element is compared to } j\text{th smallest} \\ 0 & \text{otherwise} \end{cases}$

$\sum_{i,j} X_{ij}$ = total # comparisons made by the algorithm
 $\approx$ total running time

$E[\sum_{i,j} X_{ij}]$ = average running time

$$E[\sum_{i,j} X_{ij}] = \sum_{i,j} E[X_{ij}]$$

now analyze $E[X_{ij}]$ in isolation

$$\begin{aligned} E[X_{ij}] &= \Pr[X_{ij} = 1] \\ &= \Pr[i\text{th element is compared to } j\text{th}] \end{aligned}$$



i th elem is compared w/ j th
 $\Leftrightarrow$ i th or j th elem selected before any element in between happens w/ probability $\frac{2}{j-i+1}$

$$E[X_{ij}] = \frac{2}{j-i+1}$$

$$E[\sum_{i,j} X_{ij}] = \sum_{i,j} E[X_{ij}]$$

$$= \sum_{i,j} \frac{2}{j-i+1}$$

$$= \sum_{i,j} \frac{2}{j-i+1} = \sum_{i,j} \frac{2}{j-i+1}$$

$$\leq 2n \left(\frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right)$$

$$\leq O(n \log n) \quad \text{with harmonic series}$$

Quickselect (HW)

Lecture 2

Probability

Hash Functions,

Universal Hash Functions

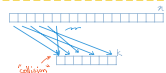
$h: [n] \rightarrow k$ s.t. for all pairs $i, j \in [n]$,
 $\Pr[h(i) = h(j)] \leq \frac{1}{k}$ (much weaker than ideal hash)

e.g. $h(x) = (ax + b) \bmod p \bmod k$

where • p prime, $> k$

• a in $\{1, \dots, p-1\}$ unif. random

• b in $\{0, \dots, p-1\}$ unif. random



Markov's Inequality

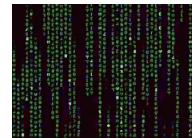
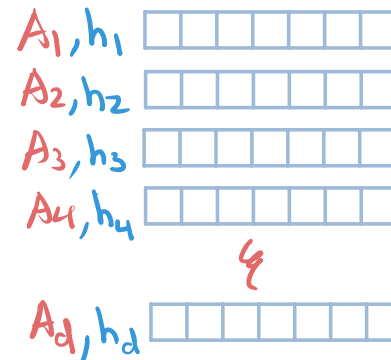
Markov's Inequality (special case)

Let $X \geq 0$ be nonnegative RV

$$\Pr[X \geq 2E[X]] \leq \frac{1}{2}$$

Algorithms

Frequency Estimation, Heavy Hitters (in streams)



estimate f_i w/
 $\min_{j=1, \dots, d} A_j[h_j(i)]$

Lecture 3

Probability

k-wise independence,
higher order moments
(generalizing universal hash
functions and Markov's inequality)

Algorithms

Hash tables w/
chaining +
linear probing

k-wise independent hash functions

$$h: [U] \rightarrow [m]$$

s.t. for any k keys $x_1, x_2, \dots, x_k \in [U]$,
 k values $y_1, y_2, \dots, y_k \in [m]$,

$$P[h(x_1)=y_1, h(x_2)=y_2, \dots, h(x_k)=y_k] = \frac{1}{m^k}$$

eg. $h(x) = a_0 + a_1x + a_2x^2 + \dots + a_{k-1}x^{k-1} \pmod{p}$

where $p \geq U$ prime,

$a_0, a_1, \dots, a_{k-1} \in [p]$ indep, uniformly random

Important lemma

let $X_1, \dots, X_n \in \{0, 1\}$ be 4-wise independent
and $\mu = E[X_1 + \dots + X_n]$. then for $\beta > 0$,

$$P[X_1 + \dots + X_n \geq \mu + \beta] \leq \frac{4 \max\{\mu, \mu^3\}}{\beta^4}$$

k-wise independent random variables

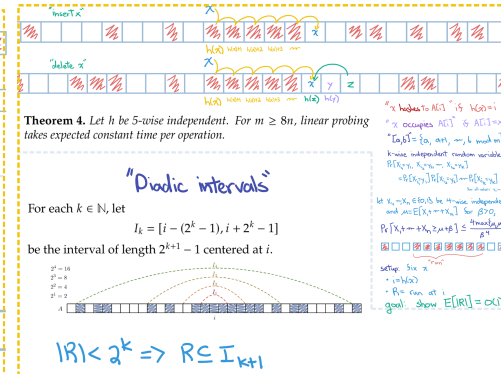
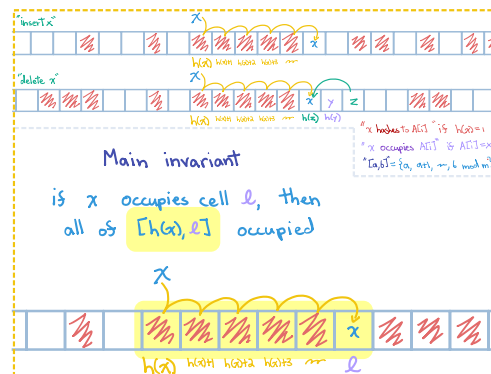
$$P[X_1=y_1, X_2=y_2, \dots, X_k=y_k]$$

$$= P[X_1=y_1] P[X_2=y_2] \dots P[X_k=y_k]$$

for all values $y_1, \dots, y_k$

Main invariant

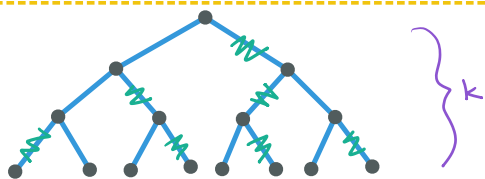
if x occupies cell i , then
all of $[h(x), i]$ occupied



Lecture 4

Probability

Galton-Watson processes



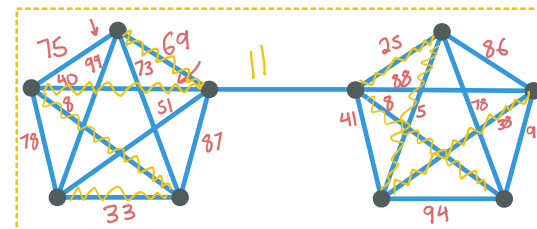
let T = complete binary tree of height k ,
w/ every edge deleted w/ prob $\leq 1/2$

Fact: $P[\text{some leaf connected to root}] \geq \frac{1}{k}$

"Galton-Watson process"

Algorithms

Random Contractions for min cut



Randomized contractions

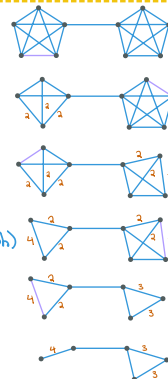
Input: $G=(V,E)$, weights $c \in \mathbb{R}_{\geq 0}^E$
(unweighted)

① while $|E| > 1$

 Ⓐ sample $e \in E$ in proportion to c

 Ⓑ "contract" e

② return set of edges (in original graph)
 contracted to remaining edge



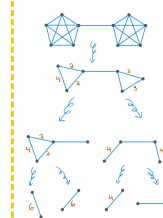
Amplification by branching

Instead of restarting at the beginning, restart partway through

High-level algorithm

① randomly contract until prob.
of ruining the min-cut is $\sim 1/2$

② Branch:
make two copies of current graph
recurse on both copies



Lecture 5

Probability

Multiplicative Chernoff Bound

Multiplicative Chernoff bound

let $X_1, \dots, X_n \in [0, 1]$ be independent

$$\mu = E[X_1 + \dots + X_n]$$

$$\epsilon \in (0, 1)$$

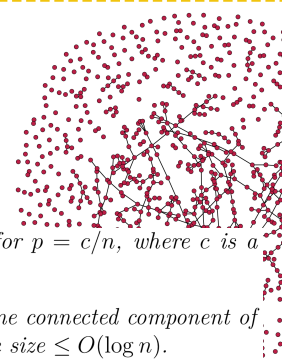
- $P[X_1 + \dots + X_n \geq (1 + \epsilon)\mu] \leq e^{-\epsilon^2 \mu / 3}$
- $P[X_1 + \dots + X_n \leq (1 - \epsilon)\mu] \leq e^{-\epsilon^2 \mu / 2}$

Algorithms (/ results)

Random graphs & the giant component

Erdős-Rényi graphs $G(n, p)$

- n vertices
- each edge sampled independently w/ probability p



Theorem 5.3. Consider a random graph $G \sim \mathcal{G}(n, p)$ for $p = c/n$, where c is a constant.

- If $c > 1$, then with high probability, there is exactly one connected component of G with $\Omega(n)$ vertices, and all other components have size $\leq O(\log n)$.
- For $c < 1$, then with high probability, all connected components of G has size $< O(\log n)$.

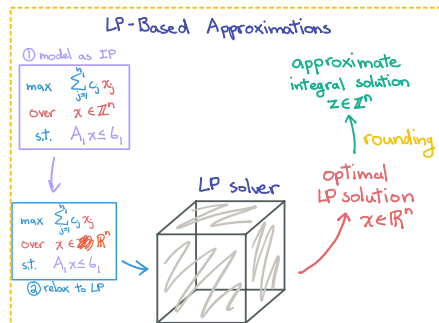
Lecture 6

Probability (& techniques)

Approximations via linear programming & randomized rounding

Linear programs (LPs)
Similar to IPs, except no integrality constraints

min/max $\sum_{j=1}^n c_j x_j$ linear objective
over $x \in \mathbb{R}^n$ linear integral variables
st. $A_1 x \leq b_1$ linear constraints
 $A_2 x = b_2$
 $A_3 x \geq b_3$



Continuous optimization vs Discrete optimization

randomization

Algorithms

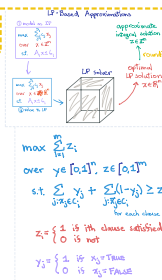
Max Sat Set Cover Covering Integer Programs

For each variable x_j , independently, randomly set $x_j = t$ with probability g_j , and $x_j = f$ otherwise.

$$E[\# \text{ sat. clauses}] \geq (1 - 1/e) \text{OPT}$$

k	oblivious	LP	average
1	1/2	$z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
2	3/4	$(3/4)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
3	7/8	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
4	15/16	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
5	31/32	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$
...	...	...	...
k	$1 - 1/2^k$	$(1 - 1/e)z_i \geq (3/4)z_i$	$\geq (3/4)z_i$

Can we get $\frac{3}{4}$ -APX?



Rounding LP solution
 $x \in [0, 1]^n \rightsquigarrow F \subseteq \{S_1, \dots, S_n\}$

For each set S_j , independently, take S_j with probability x_j .

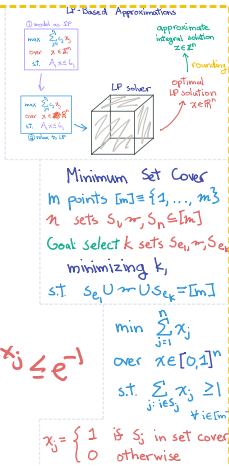
$$E[|F|] = \sum x_j \leq \text{OPT}$$

Fix a point i .

$$E[\# \text{ sets in } F \text{ covering } i] = \sum_{j \in S_i} x_j \geq z_i$$

$$P[F \text{ not cover } i] = \prod_{j \in S_i} (1 - x_j) \leq e^{-\sum_{j \in S_i} x_j} \leq e^{-z_i}$$

$$P[F \text{ cover } i] \geq 1 - e^{-z_i}$$



Lecture 7

Probability

Continuous Random Variables

Variance, Chebyshev's Variance Reduction

Median trick

Variance

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Chebyshev's inequality

$$P[|X - E[X]| \geq d] \leq \frac{\text{Var}(X)}{d^2}$$

"median trick", "median of means"

- ① take mean to reduce variance, get constant prob. of error
- ② take median to amplify success prob. exponentially

Algorithms

Distinct Elements



Goal: estimate # of distinct elements in the stream

Suppose you had an ideal hash function

$$h: [m] \rightarrow [0, 1]$$

and the following facts:

let $Y_1, \dots, Y_k \in [0, 1]$ independent and uniformly random

$$X = \min \{Y_1, \dots, Y_k\}$$

$$\text{then } E[X] = \frac{1}{k+1}$$

$$\textcircled{1} X \leftarrow 1$$

② for each element e

$$\textcircled{a} X \leftarrow \min \{X, h(e)\}$$

$$/* E[X] = \frac{1}{k+1}, k = (\text{\# distinct elements}) */$$

$$\textcircled{3} \text{ return } \frac{1}{X} - 1$$

$$\text{recall } E[X] = \frac{1}{k+1}$$

$$\text{goal: show } |X - E[X]| \leq \frac{\epsilon}{k+1}$$

w/ good prob.

$$P[|X - E[X]| \geq \frac{\epsilon}{k+1}]$$

$$\leq \text{Var}(X) \cdot \left(\frac{k+1}{\epsilon}\right)^2 \leftarrow \text{Chebyshev's inequality}$$

where

$$\text{Var}(X) = \frac{k}{(k+1)^2(k+2)} \approx \frac{1}{k^2}$$

Q: how to reduce $\text{Var}(X)$?

we can get $(1 \pm \epsilon)$ -APX

w/ 10% error prob. in $O(\frac{1}{\epsilon^2})$ space

what about high probability?

① make q copies $Y_1, \dots, Y_q$ of

② return median of $Y_1, \dots, Y_q$

Algo correct $\Leftrightarrow \geq \frac{1}{2}$ of Y_i 's correct

let $Z = \# \text{correct } Y_i$'s $E[Z] = \frac{q}{2}$

$$P[Z \leq (1 - \frac{\epsilon}{4})E[Z]] \leq e^{-\left(\frac{\epsilon}{4}\right)^2 E[Z]/2} \leq \frac{1}{n^{O(n)}} \text{ for } q = O(\log n)$$

ideal hash $h: [m] \rightarrow [0, 1]$

let $Y_1, \dots, Y_k \in [0, 1]$

indep. and unifs. rand.

$$X = \min \{Y_1, \dots, Y_k\}$$

$$\text{then } E[X] = \frac{1}{k+1}$$

$$\text{and } E[X^2] = \frac{2}{(k+1)(k+2)}$$

$$\text{Var}(X) = E[(X - E[X])^2]$$

$$X_1, \dots, X_k \text{ indep. copies of } X$$

$$\bar{X} = \frac{1}{k}(X_1 + \dots + X_k)$$

$$\text{Var}(\bar{X}) = \frac{1}{k} \text{Var}(X)$$

$$P[X_1 + \dots + X_k \leq (1 - \epsilon)k] \leq e^{-\epsilon^2 k/4}$$

Lecture 8

Probability

Gaussian random variables

Gaussian random variables

$$x \sim \mathcal{N}(\mu, \sigma^2)$$

mean variance

Def: the nicest, most pleasant, convenient random variable ever

Fact let $x \sim \mathcal{N}(\mu, \sigma^2)$, $d \in \mathbb{R}$.

$$\text{then } x+d \sim \mathcal{N}(\mu+d, \sigma^2)$$

$$dx \sim \mathcal{N}(d\mu, d^2\sigma^2)$$

Fact let $x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$$\text{then } x_1 + x_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Fact let $x \sim \mathcal{N}(0, 1)$, $t < \frac{1}{2}$

$$\text{then } E[e^{tx^2}] = \frac{1}{\sqrt{1-2t}}$$

Algorithms

Dimensionality reduction

Bag-of-words

literally any sentence is fine

any ↑ is literally sentence English
five encodes text in high dimensions

Lemma 8.8. Let $A \sim \mathcal{N}^{k \times d}$ be a random matrix where each coordinate A_{ij} is an independently drawn sample from $\mathcal{N}$. Let $\epsilon > 0$ be sufficiently small. For any vector x ,

$$\mathbb{P}\left[(1-\epsilon)\|x\|^2 \leq \frac{1}{k}\|Ax\|^2 \leq (1+\epsilon)\|x\|^2\right] \geq 1 - 2e^{-k/8}$$

$$k \left\{ \left[\frac{1}{k} \|Ax\|^2 \right] \leftarrow \alpha_i \right.$$

$$E[\|Ax\|^2] = \sum_{i=1}^k E[(Ax)_i^2] = k\|x\|^2$$

$$E[(\sum_{i=1}^k A_{ij} x_j)^2] = \|x\|^2$$

$$A_{ij} x_j \sim \mathcal{N}(0, x_j^2)$$

$$\sum_{j=1}^n A_{ij} x_j \sim \mathcal{N}(0, \|x\|^2)$$

$$(Ax) \sim \mathcal{N}^k(0, \|x\|^2)$$

$$\Pr[\|Ax\|^2 \geq k\|x\|^2 \cdot (1+\epsilon)] \leq e^{-\epsilon^2 k/8}$$

Dimensionality Reduction - high level

$O(\log n)$ dimensions suffice

where $n = \#$ points

$$\mathbb{R}^{\text{very large}} \xrightarrow{f(x)} \mathbb{R}^{O(\log n)}$$

One can map n points into $\mathbb{R}^{O(\log n/\epsilon^2)}$ while preserving all distances up to $(1 \pm \epsilon)$ -APX

Lecture 9

Probability

Locality-sensitive hashing

Fix $\theta \in [0, \pi]$. given q , either:

(a) output $p \in P$ w/ $\angle(p, q) \leq 4\theta$

(b) declare no $p \in P$ w/ $\angle(p, q) \leq \theta$

define $h: S^{d-1} \rightarrow \{-1, 1\}^k$ by

$$h_i(x) = \text{sign}(\langle g_i, x \rangle) \text{ w/ } g_i \sim \mathcal{N}^d$$

when querying y :

- if there is $x \in P$ w/ $\angle(x, y) \leq \theta$, then it collides w/ probability $\geq 1/\sqrt{n}$

- expect to collide w/ at most one point $x \in P$ w/ $\angle(x, y) > 2\theta$

Lemma let $\bullet x, y \in S^{d-1}$
 $\bullet g \sim \mathcal{N}^d$
then
 $P[\text{sign}(\langle g, x \rangle) \neq \text{sign}(\langle g, y \rangle)] = \frac{\angle(x, y)}{\pi}$

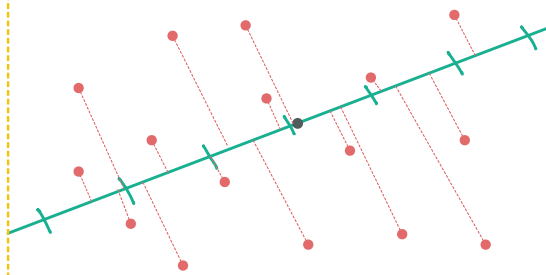


$k = \pi(\log(n)/2\theta)$

d-dimensional sphere

Random line embedding & Buckets

$$h(x) = \lfloor \langle g, x \rangle + d \rfloor \text{ w/ } g \sim \mathcal{N}^d, d \in [0, 1] \text{ unifs. rand.}$$

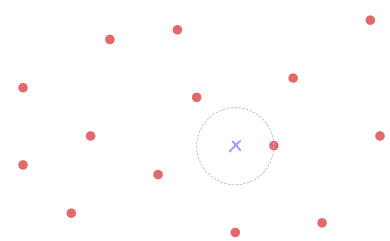


Algorithms

Approximate nearest neighbors

Nearest neighbor search

Input: n points $P \subset \mathbb{R}^d$



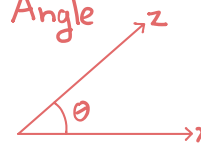
Query: given $q \in \mathbb{R}^d$, find closest point in P to q

Types of distances

- Euclidean distance

$$\|x - z\|_2 = \sqrt{\sum_i (x_i - z_i)^2}$$

- Angle



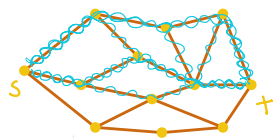
Lecture 10

Probability (and techniques)

LP duality

L_1 -metric embeddings

Max Flow
Min Cut



Path Packing LP

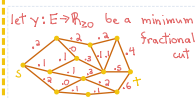
maximize $\sum_{p \in P} x_p$
over $x: P_{st} \rightarrow \mathbb{R}_{\geq 0}$
(all s-t paths)

s.t. $\sum_{p \in P} x_p \leq c_e$ for all edges e
"Ax ≤ c" (capacity of e)

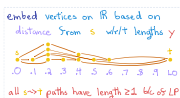
Path Covering LP
(i.e. cut)

minimize $\sum_{e \in E} c_e y_e$
over $y: E \rightarrow \mathbb{R}_{\geq 0}$
(total capacity of "fractional cut")

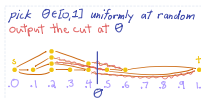
s.t. $\sum_{e \in P} y_e \geq 1$ for all $p \in P_{st}$
"A'y ≥ b" (continuous analog of s-t cut)



let $y: E \rightarrow \mathbb{R}_{\geq 0}$ be a minimum fractional cut



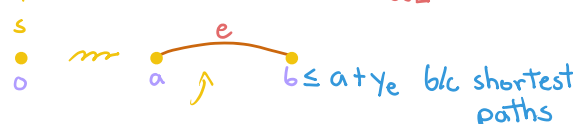
embed vertices on $\mathbb{R}$ based on distance from s with lengths y



pick $\theta \in [0, 1]$ uniformly at random output the cut at θ

$E[\text{capacity of cut}]$

$$\stackrel{\text{(linearity of expectation)}}{=} \sum_{e \in E} c_e \Pr[e \text{ cut}] \leq \sum_{e \in E} c_e y_e$$



minimum capacity of any fractional cut

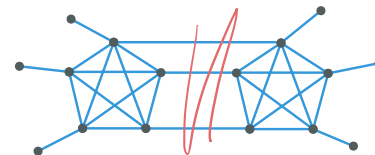
b/c shortest paths

Algorithms

Sparsest Cut

Special case: uniform sparsest cut
w/ uniform demands $b(u,v) = 1 \quad \forall u,v$

$$\text{"sparsity"} = \frac{c(\delta(s))}{|S| |T|}$$



useful primitive for graph partitioning, divide & conquer

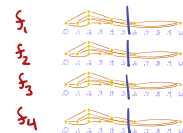
Q: how to round LP to sparse cut?

min $\sum_{e \in E} c_e d(u,v)$ over metrics $d: V \times V \rightarrow \mathbb{R}_{\geq 0}$
s.t. $\sum_{i=1}^k b_i d(s_i, t_i) \geq 1$

Suppose $d(u,v)$ was of the form

$$d(u,v) = \sum_{i=1}^k |f_i(u) - f_i(v)| \quad \text{for } f_i: V \rightarrow [0, 1]^k$$

(assume $\min f_i(a) = 0, \max f_i(a) = 1$)



" L_1 -metric"

Sparsest cut
Input: undirected graph $G=(V,E)$, edge capacities $c(e) > 0$ for $e \in E$, demands $b(u,v) > 0$ for $u,v \in V$.
Goal: compute SCF minimizing "sparsity" $\frac{c(\delta(s))}{|S| |T|}$ where S and T are edge sets.

Lecture 11

Probability (and techniques)

Algorithms

Other metric-based problems
(sometimes via correspondence between cuts $\leftrightarrow$ metrics)

Buy-at-bulk network design
k-median clustering
metric labeling
group Steiner tree
metrical task systems
vehicle routing
hierarchical placement
distributed k-server
covering Steiner tree
⋮



these problems are easy for tree metrics

Approximation by tree metrics

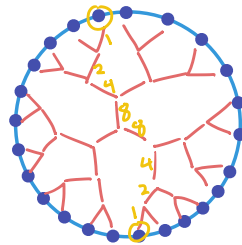
Input: undirected $G=(V,E)$ w/ edge weights $w:E \rightarrow \mathbb{R}_{\geq 0}$

(High-level)
Goal: compute tree T (containing V as nodes)
so that shortest paths on T are close to T

$$d_T(u,v) \geq d_G(u,v)$$

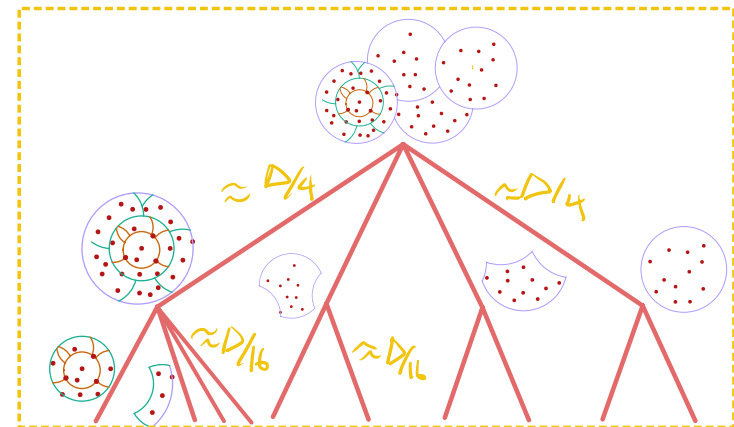
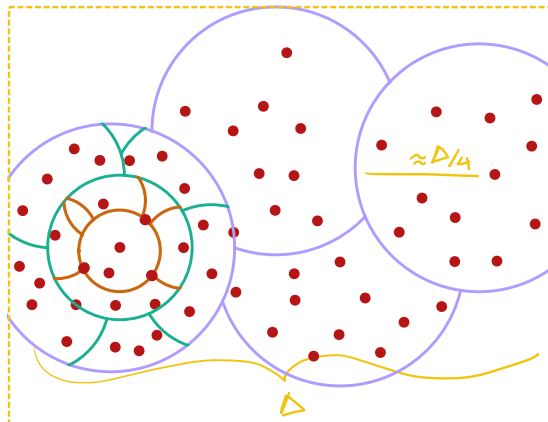
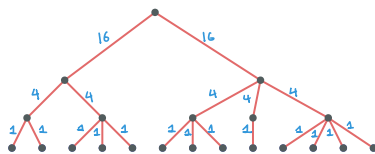
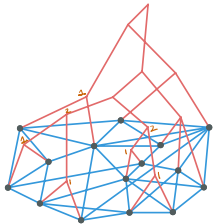
$$\text{For } e=f_u,v \quad \begin{matrix} \text{(distance in } T) \\ \text{(stretch } e) = \frac{d_T(u,v)}{d_G(u,v)} \\ \text{(distance in } G) \end{matrix}$$

ideally
(stretch e) = $\text{polylog}(n) \forall e$



Randomized Tree metrics

Theorem In randomized polynomial time,
randomized hierarchical dominating tree metric w/
 $\mathbb{E}[\text{stretch } e] \leq O(\log n) \quad \forall e \in E$



Lecture 12

Probability (and techniques)

Additive Chernoff Inequality

Double-sampling

VC dimension

Additive Chernoff bound

let $X_1, \dots, X_n \in [0, 1]$ be independent

$$\mu = E[X_1 + \dots + X_n] / n$$

$\epsilon \in (0, 1)$

- $P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right] \leq e^{-2\epsilon^2 n}$
- $P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$

Algorithms

ϵ -samples and ϵ -nets

ϵ -sample $|M_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition
 $k = O\left(\frac{\ln(g(|P|))}{\epsilon^2}\right) \Rightarrow \epsilon$ -sample $k = O\left(\frac{\ln(g(|P|))}{\epsilon^2}\right) \Rightarrow \epsilon$ -net (w/ constant probability)

ϵ -sample theorem [VC71]
 $k \geq O\left(\frac{\ln(g(2k))}{\epsilon^2}\right) \Rightarrow \epsilon$ -sample (w/ constant probability)

e.g. $g(n) = O(n^d)$
 $k \geq O\left(\frac{d \ln(d/\epsilon)}{\epsilon^2}\right) \Rightarrow \epsilon$ -sample
 ↑ independent of $|P|$ (!)

$P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right], P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$ $g(n) = \max\{|RAQ| : \text{asf, lalsn}\}$

$\mu(r) = \frac{|P \cap r|}{|P|}$
 $M_Q(r) = \frac{|Q \cap r|}{|Q|}$
 $RAQ = \{r \cap q : r \in R, q \in Q\}$

ϵ -sample $|M_Q(r) - \mu(r)| \leq \epsilon \quad \forall r \in R$

let Q sample k points w/ repetition
 $k = O\left(\frac{\ln(g(|P|))}{\epsilon^2}\right) \Rightarrow \epsilon$ -sample $k = O\left(\frac{\ln(g(|P|))}{\epsilon^2}\right) \Rightarrow \epsilon$ -net (w/ constant probability)

ϵ -net theorem [HW87]
 $k \geq O\left(\frac{\ln(g(2k))}{\epsilon^2}\right) \Rightarrow \epsilon$ -net (w/ constant probability)

e.g. $g(n) = O(n^d)$
 $k \geq O\left(\frac{d \ln(d/\epsilon)}{\epsilon^2}\right) \Rightarrow \epsilon$ -net
 ↑ independent of $|P|$ (!)

$P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right], P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$ $g(n) = \max\{|RAQ| : \text{asf, lalsn}\}$

$\mu(r) = \frac{|P \cap r|}{|P|}$
 $M_Q(r) = \frac{|Q \cap r|}{|Q|}$
 $RAQ = \{r \cap q : r \in R, q \in Q\}$

Lecture 13

Probability
(and techniques)

Algorithms
(models and applications)

Consistency model

PAC learning model

Generalization error

Occam's Razor

Generalization by growth

Lecture 14

Probability
(and techniques)

Algorithms
(models and applications)

Entropy

(Optimal)
Error-Correcting
Codes

Lecture 15

Probability (and techniques)

Lovász Local Lemma

Algorithms (models and applications)

Moser's resampling algorithm

Lecture 16

Probability (and techniques)

Algorithms (models and applications)

Random Walks

Stationary Distribution

Perron-Frobenius Theorem

Pagerank

Lecture 16

Probability (and techniques)

High-dimensional
random walks

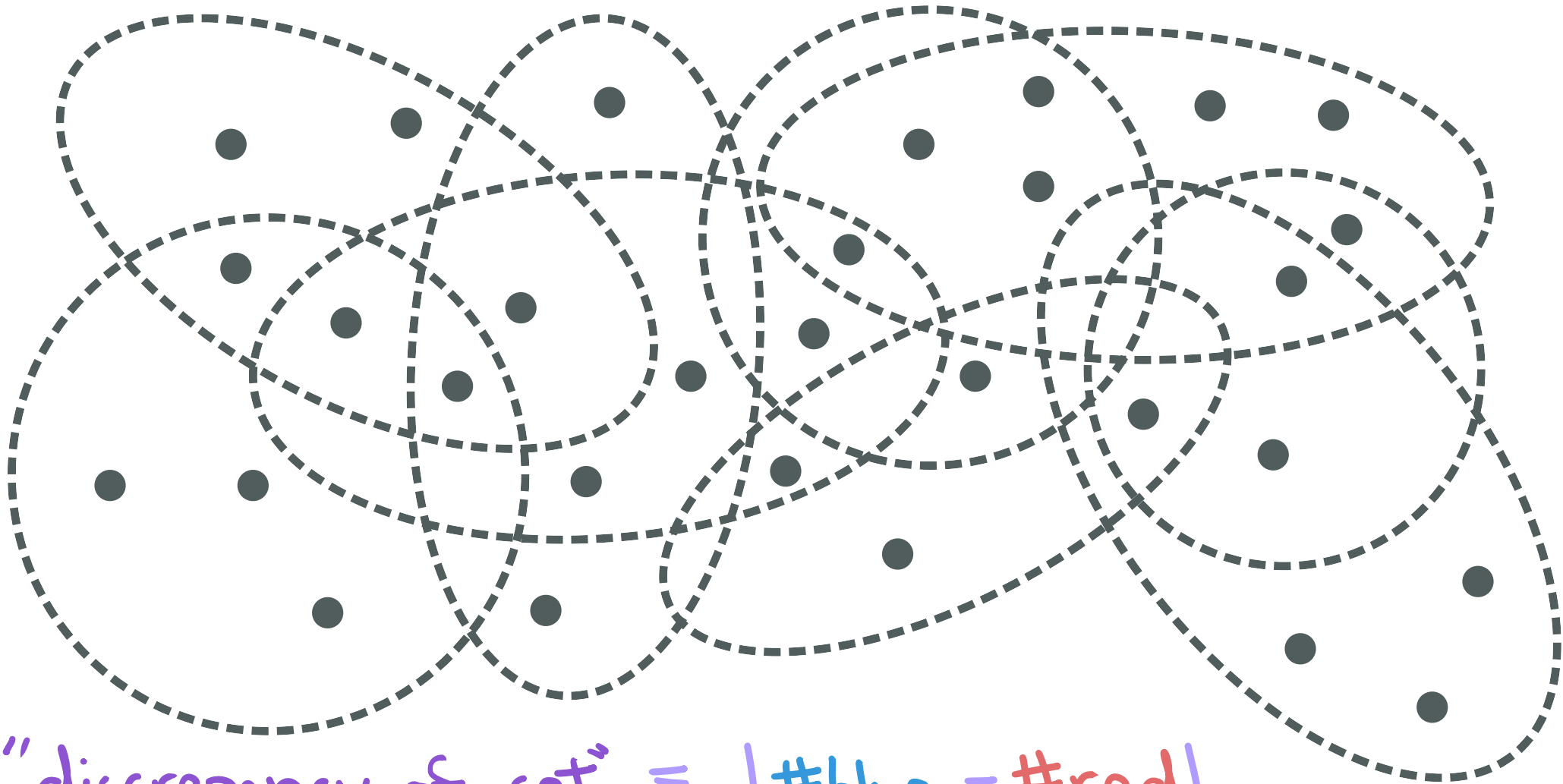
More fun w/ Gaussians

Algorithms (models and applications)

Discrepancy
minimization

Balanced coloring

color points red/blue s.t. all sets have same # of each

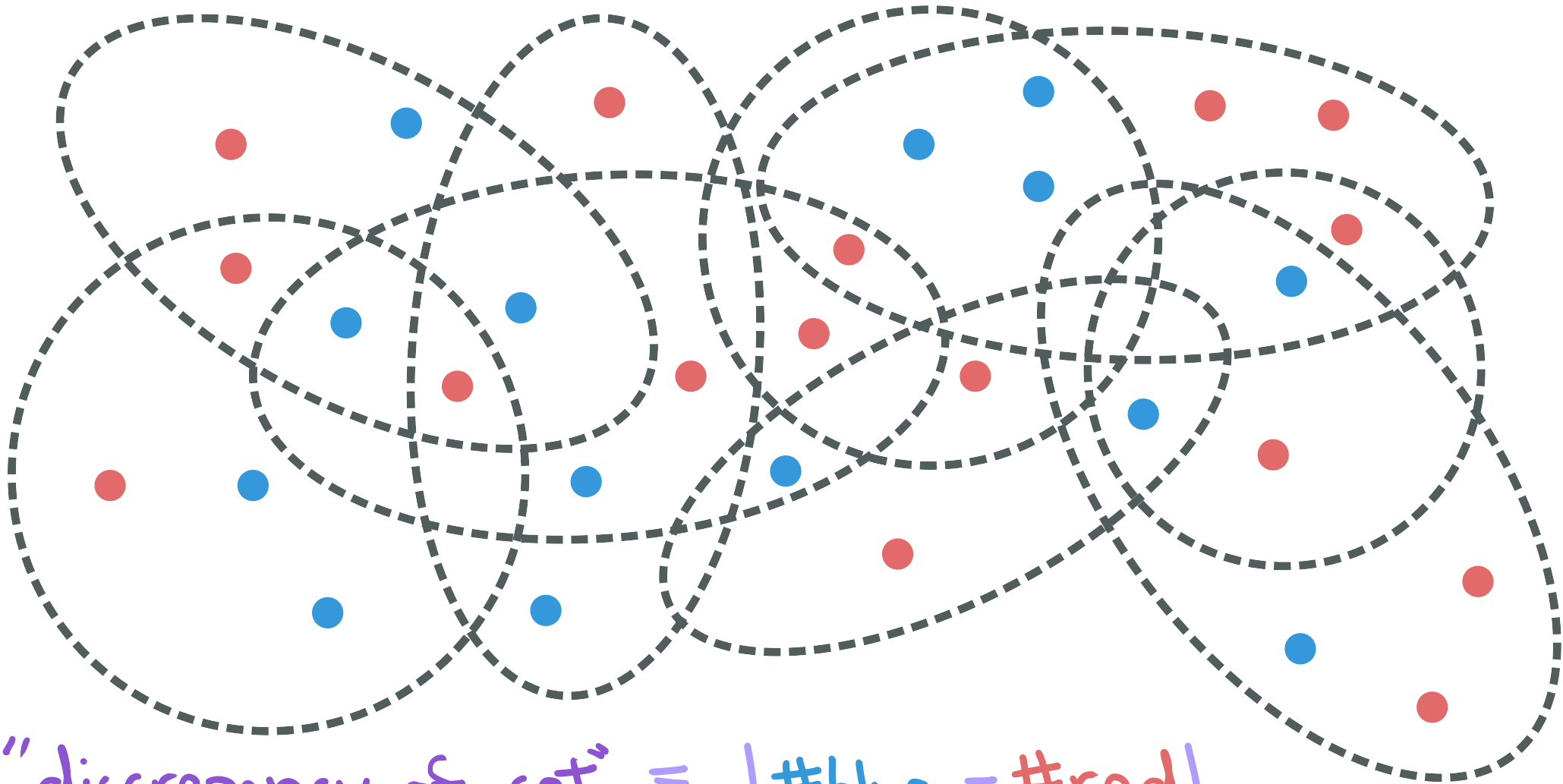


"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Balanced coloring

color points red/blue s.t. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Additive Chernoff bound

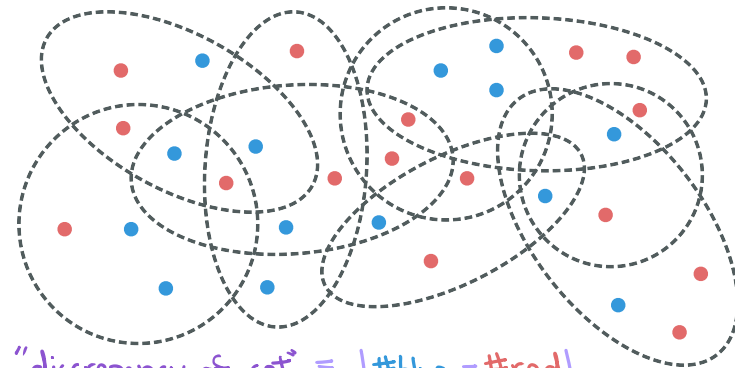
let $X_1, \dots, X_n \in [0, 1]$ be independent

$$\mu = E[X_1 + \dots + X_n] / n \quad \epsilon \in (0, 1)$$

- $P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right] \leq e^{-2\epsilon^2 n}$
- $P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$

Balanced coloring

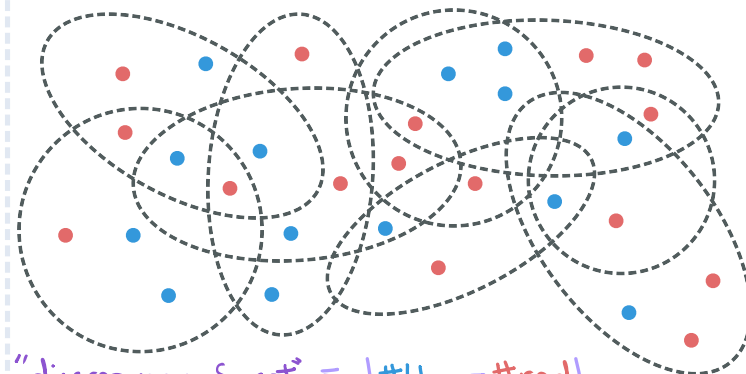
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\#blue - \#red|$
goal: minimize max discrepancy

assume n sets and n points
(uniformly)
max discrep. of random coloring?

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Additive Chernoff bound

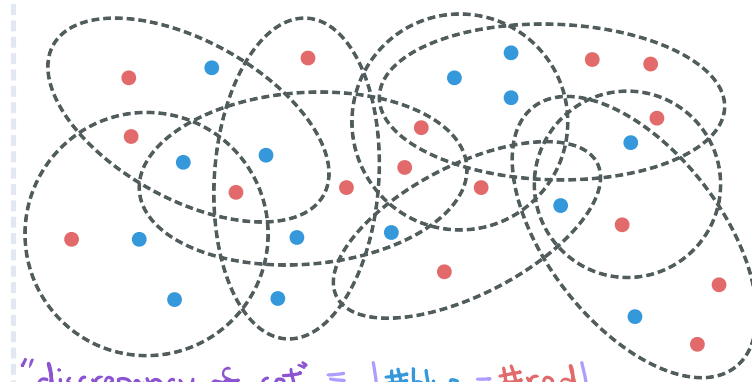
let $X_1, \dots, X_n \in [0, 1]$ be independent

$$\mu = E[X_1 + \dots + X_n] / n \quad \epsilon \in (0, 1)$$

- $P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right] \leq e^{-2\epsilon^2 n}$
- $P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$

assume n sets and n points
 max discrep. of ^(uniformly) random coloring?

Balanced coloring
 color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
 goal: minimize max discrepancy

$\sqrt{n \log n}$ by additive Chernoff,
 union bound.

Better?

[Spencer, 1985]: $O(\sqrt{n})$

more generally $O(\sqrt{n \log(m/n)})$ for m sets, n points

proof was existential in nature.

Algorithm?

Additive Chernoff bound

let $X_1, \dots, X_n \in [0, 1]$ be independent

$$\mu = E[X_1 + \dots + X_n] / n \quad \mu \in (0, 1)$$

- $P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right] \leq e^{-2\epsilon^2 n}$
- $P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$

Better?

[Spencer, 1985]: $O(\sqrt{n})$

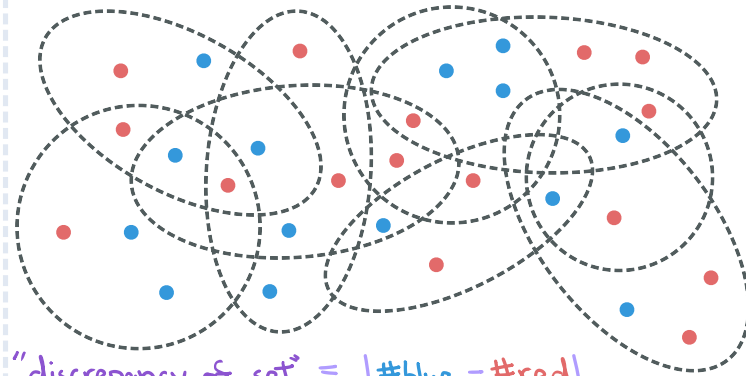
more generally $O(\sqrt{n \log(m/n)})$
for m sets, n points
(proof was existential in nature.)

Algorithm?

[Bansal 2010] $O(\sqrt{n \log(m/n)})$
(Based on rounding SDP)

[Lovett+Meka, 2012] $O(\sqrt{n \log(m/n)})$
high-dimensional random walk

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Additive Chernoff bound

let $X_1, \dots, X_n \in [0, 1]$ be independent

$$\mu = E[X_1 + \dots + X_n]/n \in (0, 1)$$

- $P\left[\frac{X_1 + \dots + X_n}{n} \geq \mu + \epsilon\right] \leq e^{-2\epsilon^2 n}$
- $P\left[\frac{X_1 + \dots + X_n}{n} \leq \mu - \epsilon\right] \leq e^{-2\epsilon^2 n}$

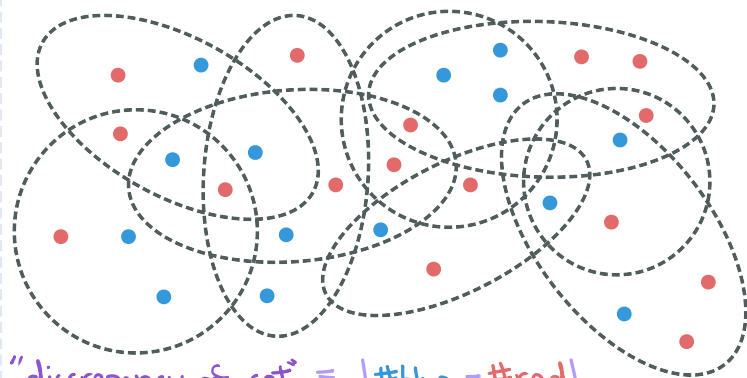
$O(\sqrt{n \log(m/n)})$ disc. [Lovett+Meka, 2012]
high-dimensional random walk

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{\chi \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$

where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} \chi_j$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

"Fractional coloring" $\chi \in [-1, 1]^n$

χ_i is "tight" if $|\chi_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

otherwise χ_i is "free"

"Fractional coloring" $x \in [-1, 1]^n$

x_i is "tight" if $|x_i| \geq 1 - \epsilon$
where $\epsilon = \frac{1}{n}$

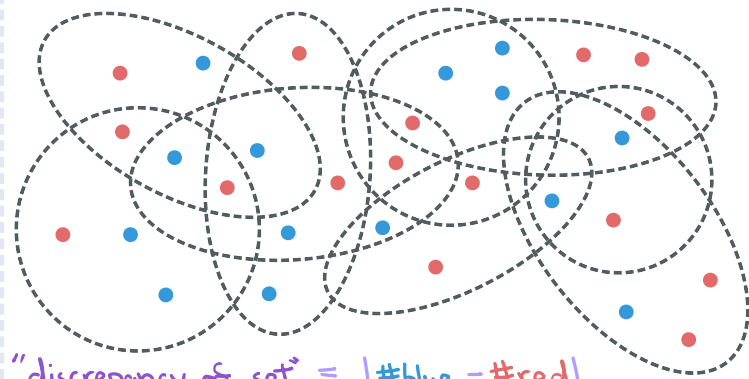
claim: if all x_i are tight, then
we can make x integral w/ $O(1)$ loss in discrepancy

proof:

just round to nearest integer

discrepancy increases by $\leq \epsilon n = 1$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Partial coloring lemma

(high-level)

rand. algo. computes $x \in [-1, 1]^n$ s.t.

$$(a) |x(S_i)| \leq O(\sqrt{n \log(m/n)}) \quad \forall S_i$$

(b) ^(at least) half of coordinates are tight

Partial $\rightarrow$ full tight coloring (high-level)

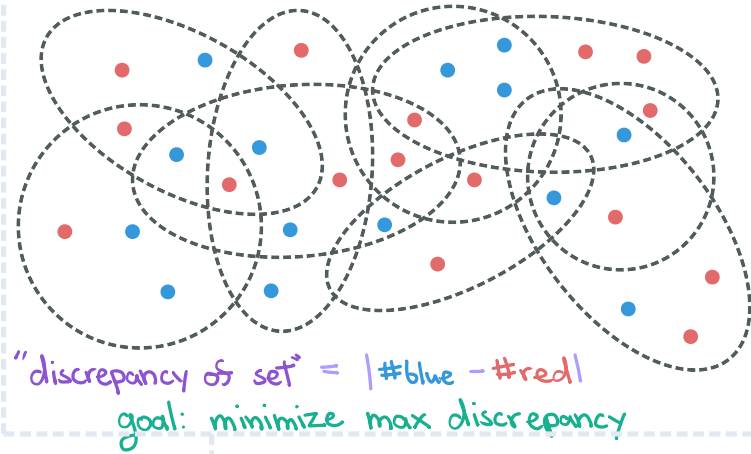
Repeatedly until all x_i tight

(a) partially color

(b) fix $n/2$ tight coordinates, remove them

Note: " $\sqrt{n \log(m/n)}$ " decreases geometrically

Balanced coloring
color points red/blue s.t. all sets have same # of each



Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, 1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

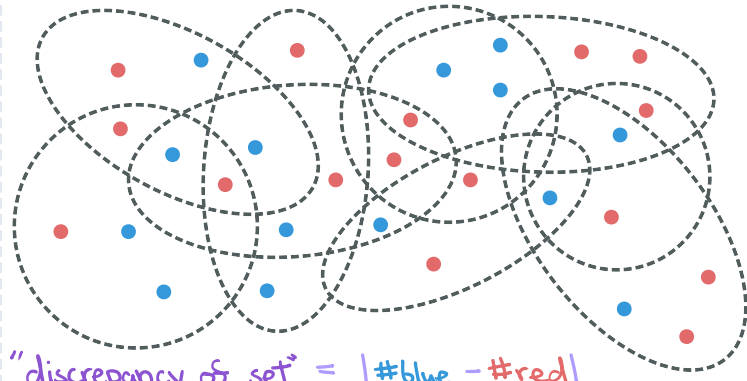
Partial coloring lemma - detailed

given $x_0 \in [-1, 1]^n$, rand. algo
computes $x \in [-1, 1]^n$ s.t. (w/ high prob.)

$$(a) |x_0(S_i) - x(S_i)| \leq O(\sqrt{n \log(m/n)}) \quad \forall S_i$$

(at least)
(b) half of x_i 's are tight

Balanced coloring
color points red/blue s.t. all sets have same # of each



goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, 1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

(x_0 used to adjust for fractional coloring from previous iterations)

Random walk algorithm (high-level)

$\lambda = c\sqrt{n \log(m/n)}$ for constant c

$$\mathcal{P} = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

take random walk

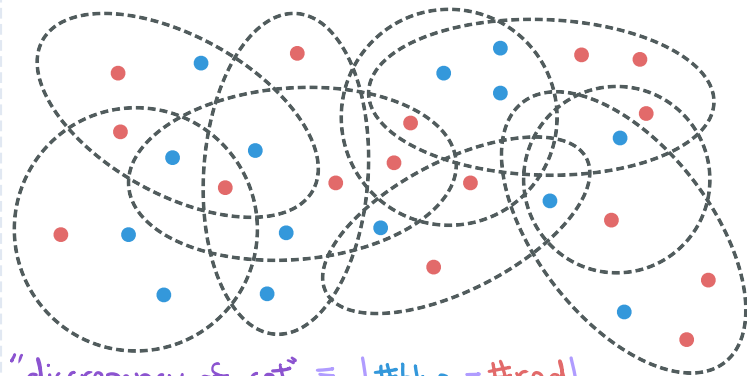
$$x^{(0)} = x_0, x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(T)} \in \mathcal{P}$$

where $\forall t$,

$x^{(t)} - x^{(t-1)}$ is a Gaussian vector from
a subspace of $\mathbb{R}^n$

chosen to avoid "tight sets"
& tight coordinates

Balanced coloring
color points red/blue st. all sets have same # of each



Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians.

① for $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② let $Y_1, \dots, Y_T \in \mathbb{R}$ where for each t ,

"conditionally Gaussian"

$$Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \text{ for } \sigma_t^2 \leq 1$$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

(for independent $Y_1, \dots, Y_T \sim \mathcal{N}(0, 1)$), ① $\Rightarrow$ ②)

Random walk algorithm (high-level)

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

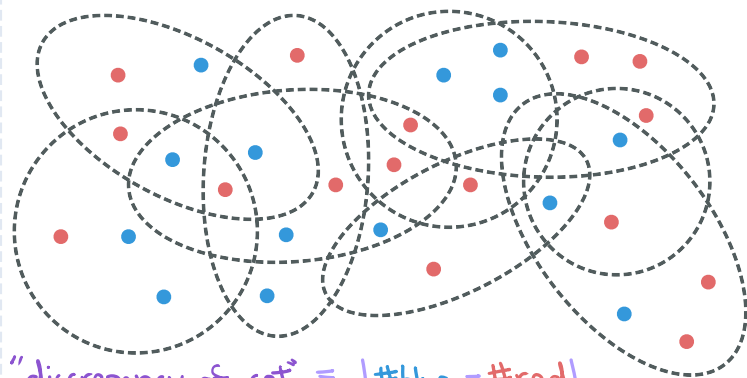
fix $x^{(t)} \in P$.

- $j \in [n]$ is "tight" if $|x_j^{(t)}| \geq 1 - \epsilon$
- s_i is "tight" if $|x_0(s_i) - x^{(t)}(s_i)| \geq \lambda - \epsilon$

"safe subspace":

$$Y^{(t)} = \left\{ y \in \mathbb{R}^n : \begin{array}{l} y(s_i) = 0 \ \forall \text{ tight } s_i, \\ y_j = 0 \ \forall \text{ tight } j \in [n] \end{array} \right\}$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) (at least) half of x_i 's are tight

Two lemmas about Gaussians

① for $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$
then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

Random walk algorithm (high-level)

$\lambda = c\sqrt{n \log(m/n)}$ for constant c

$$\mathcal{P} = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

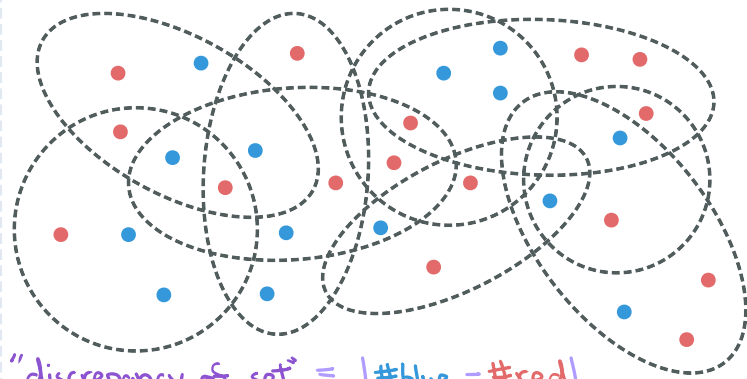
take random walk

$$x^{(0)} = x_0, x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(T)} \in \mathcal{P}$$

where $\forall t$,

$x^{(t)} - x^{(t-1)}$ is a Gaussian vector from
safe subspace $\Upsilon^{(t)}$

Balanced coloring
color points red/blue st. all sets have same # of each



Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) ^(at least) half of x_i 's are tight

"safe subspace":

$$Y^{(t)} = \left\{ y \in \mathbb{R}^n : \begin{array}{l} y(S_i) = 0 \quad \forall \text{ tight } S_i, \\ y_j = 0 \quad \forall \text{ tight } j \in [n] \end{array} \right\}$$

sampling a Gaussian vector from $Y^{(t)}$:

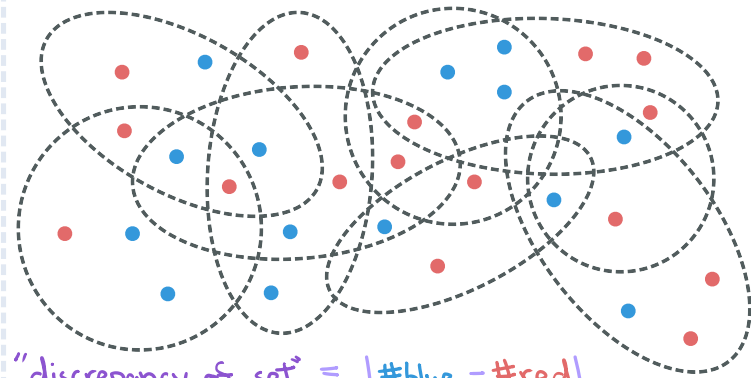
let $c_1, \dots, c_k$ be orthonormal basis of $Y^{(t)}$

let $U = [c_1 | c_2 | \dots | c_k] \in \mathbb{R}^{n \times k}$

sample $g \sim \mathcal{N}^k(0, 1)$

return $Ug \in \mathbb{R}^n$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(S_i) - x(S_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$\forall v \in \mathbb{R}^n$:

Lemma: $\|U^T v\| \leq \|v\|$

sampling a Gaussian vector from $Y^{(t)}$:
 $c_1, \dots, c_k$ orthonormal basis of $Y^{(t)}$
 $U = [c_1 | c_2 | \dots | c_k] \in \mathbb{R}^{n \times k}$
sample $g \sim \mathcal{N}^k(0, 1)$
return $Ug \in \mathbb{R}^n$

Proof 1: U^T projects onto subspace $Y^{(t)}$

(subspace projections contract)

proof 2: $\|\cdot\|$ rotation invariant

$$\Rightarrow U = \begin{bmatrix} I_k \\ 0 \end{bmatrix} = \begin{bmatrix} \overset{\text{1}}{\underset{\text{1}}{\overset{\text{1}}{\vdots}}}_1 & \dots & 0 \\ 0 & \dots & \underset{\text{1}}{\overset{\text{1}}{\vdots}}_{n-k} \end{bmatrix} \left. \vphantom{\begin{bmatrix} I_k \\ 0 \end{bmatrix}} \right\}^k \left. \vphantom{\begin{bmatrix} I_k \\ 0 \end{bmatrix}} \right\}_{n-k}$$

$\forall v \in \mathbb{R}^n$:

Lemma: $\|U^T v\| \leq \|v\|$

Lemma: $\langle v, U g \rangle \sim \mathcal{N}(0, \sigma^2)$ w/ $\sigma^2 \leq \|v\|^2$

proof:

$$\langle v, U g \rangle = \langle U^T v, g \rangle \sim \mathcal{N}(0, \|U^T v\|^2)$$

sampling a Gaussian vector from $\mathcal{Y}^{(t)}$:
 $c_1, \dots, c_k$ orthonormal basis of $\mathcal{Y}^{(t)}$
 $U = [c_1 | c_2 | \dots | c_k] \in \mathbb{R}^{n \times k}$
sample $g \sim \mathcal{N}^k(0, 1)$
return $U g \in \mathbb{R}^n$

(Sum of gaussians is gaussians,
b/c gaussians are the nicest ~)

Random walk algorithm (high-level)

$$\lambda = c\sqrt{n \log(m/n)} \text{ for constant } c$$

$$\mathcal{P} = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

take random walk

$$x^{(0)} = x_0, x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(T)} \in \mathcal{P}$$

where $\forall t$,

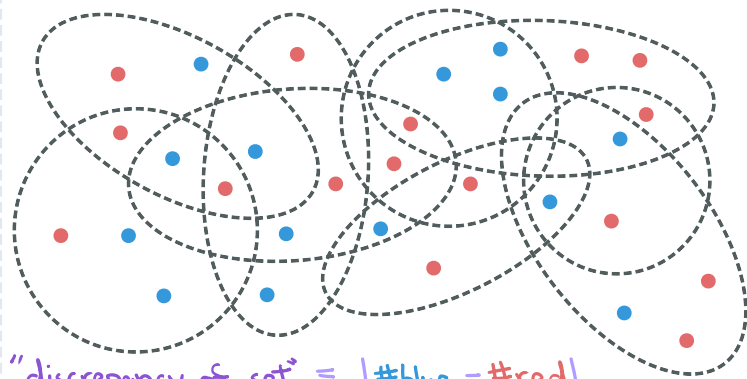
$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

where columns of $U^{(t)}$ are orthonormal

$$\text{basis of } Y^{(t-1)} = \left\{ y \in \mathbb{R}^n : \begin{array}{l} y(s_i) = 0 \ \forall \text{ tight } s_i, \\ y_j = 0 \ \forall \text{ tight } j \in [n] \end{array} \right\}$$

$$\delta = \varepsilon / \log^{O(1)}(m, n), \quad g \sim \mathcal{N}^k \text{ for } k = \dim(Y^{(t-1)})$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \varepsilon$ where $\varepsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

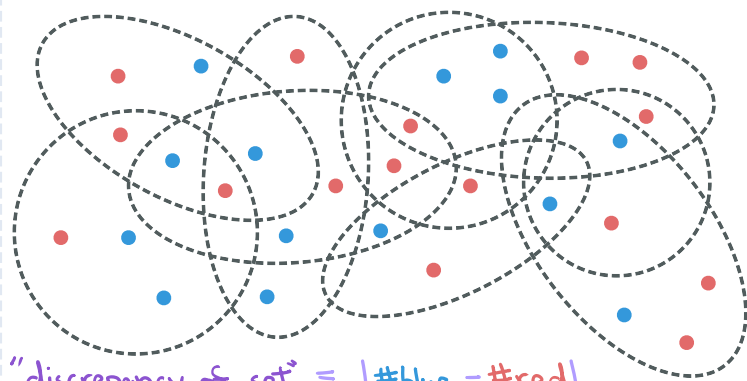
(b) ^(at least) half of x_i 's are tight

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \quad \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$
then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

claim: if $x^{(t-1)} \in P$, then $x^{(t)} \in P$ w/h/p
(w/ high probability)

tight variables/constraints are fixed

"loose" variables/constraints have ϵ -slack

$$\langle \mathbb{1}_S, \delta U g \rangle \sim \mathcal{N}(0, \delta^2 \|S\|^2) \quad \forall \text{ sets } S \quad (\text{incl. } S = \{i\})$$

$$\leq \epsilon^2 / \text{poly}(m, n)$$

$$|\langle \mathbb{1}_S, \delta U g \rangle| \leq \epsilon \quad \text{w/h/p}$$

$$\text{also implies } \|x^{(t)}\|^2 \leq n$$

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \quad \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

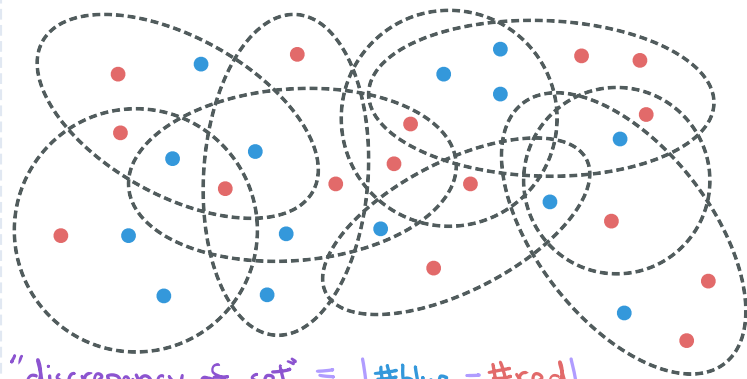
claim: if $x^{(t-1)} \in P$, then $x^{(t)} \in P$ w/h/p
(w/ high probability)

union bound $\Rightarrow$

all $x^{(1)}, \dots, x^{(T)} \in P$ w/h/p

$$\text{also: } E[\|x^{(t)}\|^2] \leq n \quad \forall t$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1 + \dots + Y_T \sim \mathcal{N}(0, \sigma_T^2)$ for $\sigma_T^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \quad \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

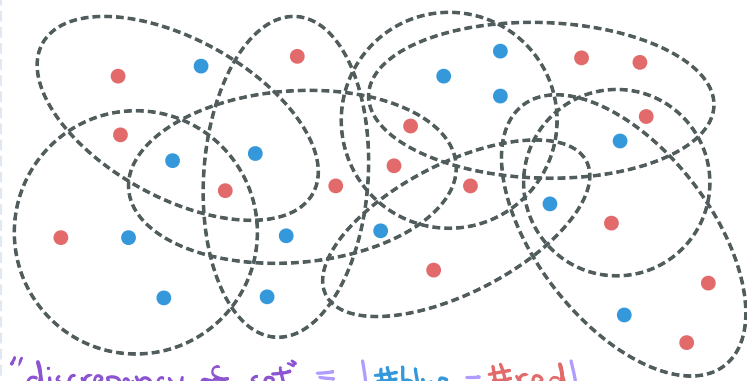
$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

$$\checkmark x^{(1)}, \dots, x^{(T)} \in P \text{ w/h/p, } E[\|x^{(t)}\|^2] \leq n$$

goal: show half of coordinates in $x^{(T)}$ are tight.

idea: argue $\|x^{(T)}\|^2 \geq n$ big otherwise

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) (at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$\chi^{(t)} = \chi^{(t-1)} + \delta U^{(t)} g^{(t)}$$

claim:

$$E[\|\chi^{(t)}\|^2 - \|\chi^{(t-1)}\|^2] \geq \delta^2 E[\dim(Y^{(t)})]$$

$\Rightarrow$

$$n \geq \|\chi^{(T)}\|^2 \geq \sum_t E[\|\chi^{(t)}\|^2 - \|\chi^{(t-1)}\|^2]$$

$$\stackrel{\text{claim}}{\geq} \delta^2 \sum_t E[\dim(Y^{(t)})] \stackrel{\text{why?}}{\geq} \delta^2 E[\dim(Y^{(T)})]$$

e.g. as long as

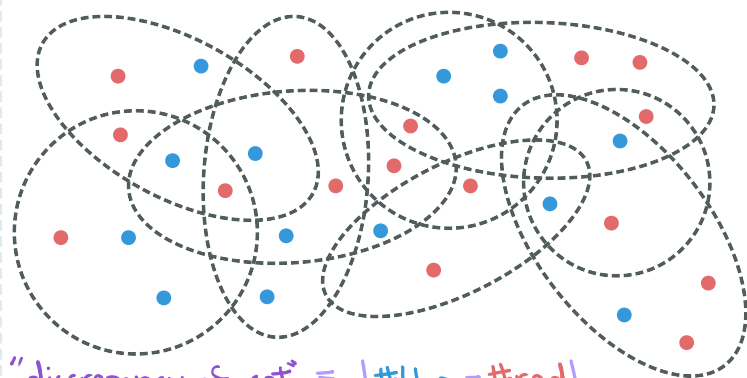
$\leq 0.99n$ constraints/variables are tight

$$\Rightarrow \dim(Y^{(t)}) \geq \Omega(n)$$

$$\Rightarrow \|\chi^{(t)}\|^2 \xrightarrow{\text{(integrality)}} n \text{ rapidly}$$

$$Y^{(t-1)} = \left\{ y \in \mathbb{R}^n : \begin{array}{l} y(s_i) = 0 \text{ } \forall \text{ tight } s_i, \\ y_j = 0 \text{ } \forall \text{ tight } j \in [n] \end{array} \right\}$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{\chi \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$
where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} \chi_j$

"Fractional coloring" $\chi \in [-1, 1]^n$

χ_i "tight": $|\chi_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$\chi_0 \in [-1, 1]^n \Rightarrow \chi \in [-1, 1]^n$ st.

(a) $|\chi_0(s_i) - \chi(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) ^(at least) half of χ_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1 | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$\chi^{(t)} = \chi^{(t-1)} + \delta U^{(t)} g^{(t)}$$

claim:

$$E[\|\chi^{(t)}\|^2 - \|\chi^{(t-1)}\|^2] \geq \delta^2 E[\dim(Y^{(t)})]$$

Proof:

up to
rotation

$$\chi^{(t)} - \chi^{(t-1)} = \gamma^{(t)} \sim \mathcal{N}^k(0, \delta^2)$$

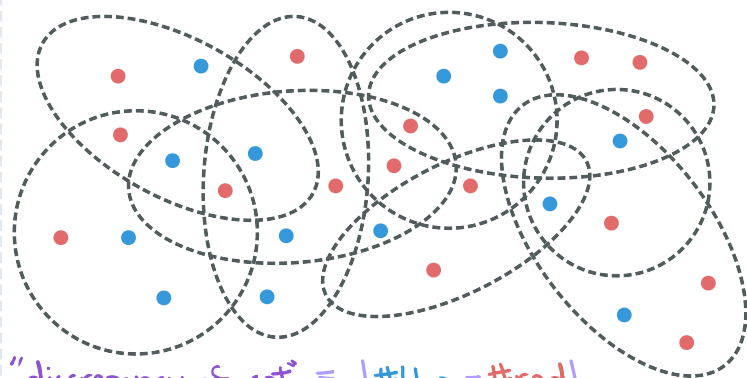
$$k = \dim(Y^{(t)})$$

$\|\cdot\|$ rotationally invariant

$$\Rightarrow E[\|\chi^{(t)} - \chi^{(t-1)}\|^2] = k\delta^2$$

sum of k Gaussians squared

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{\chi \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$
where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} \chi_j$

"Fractional coloring" $\chi \in [-1, 1]^n$

χ_i "tight": $|\chi_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$\chi_0 \in [-1, 1]^n \Rightarrow \chi \in [-1, 1]^n$ st.

(a) $|\chi_0(S_i) - \chi(S_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of χ_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$Y^{(t-1)} = \left\{ y \in \mathbb{R}^n : \begin{array}{l} y(S_i) = 0 \text{ } \forall \text{ tight } S_i, \\ y_j = 0 \text{ } \forall \text{ tight } j \in [n] \end{array} \right\}$$

$$\chi^{(t)} = \chi^{(t-1)} + \delta U^{(t)} g^{(t)}$$

claim: $E[\|\chi^{(T)}\|^2] \geq \delta^2 \sum_{t=0}^{T-1} E[\dim(Y^{(t)})]$
 (why?) $\geq \delta^2 T E[\dim(Y^{(T)})]$

Proof:

up to rotation

$$\chi^{(t)} - \chi^{(t-1)} = \gamma^{(t)} \sim \mathcal{N}^k(0, \delta^2)$$

$k = \dim(Y^{(t)})$

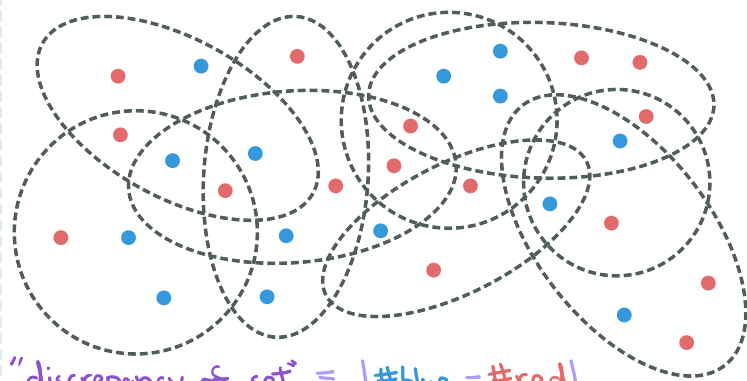
$\|\cdot\|$ rotationally invariant

$$\Rightarrow E[\|\chi^{(t)} - \chi^{(t-1)}\|^2] = k \delta^2$$

sum of k Gaussians squared

$$Y^{(t-1)} = \left\{ y \in \mathbb{R}^n : y_{(S_i)} = 0 \text{ } \forall \text{ tight } S_i, \right. \\ \left. y_j = 0 \text{ } \forall \text{ tight } j \in [n] \right\}$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(S_i) - x(S_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) (at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1 | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \quad \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

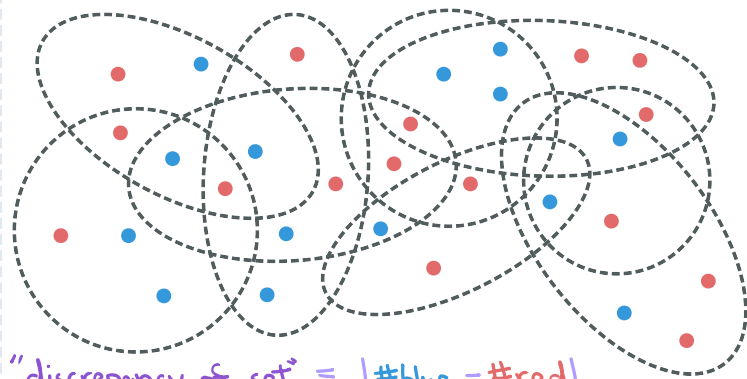
✓ all $x^{(1)}, \dots, x^{(T)} \in P$ w/h/p

$$✓ n \geq E[\|x^{(T)}\|^2] \geq \delta^2 T E[\dim(Y^{(T)})]$$

claim: after T iterations,

$$E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq \left(1 - \frac{1}{\delta^2 T}\right)n - E\left[\binom{\# \text{ tight}}{\text{sets } s_i}\right]$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, 1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) (at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$\checkmark \quad n \geq E[\|x^{(\tau)}\|^2] \geq \delta^2 T E[\dim(Y^{(\tau)})]$$

claim: after T iterations,

$$E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq \left(1 - \frac{1}{\delta^2 T}\right)n - E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right]$$

proof:

$$n \geq \delta^2 T E[\dim(Y^{(\tau)})]$$

$$= \delta^2 T \left(n - E\left[\binom{\# \text{ tight}}{j \in [n]}\right] - E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right] \right)$$

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

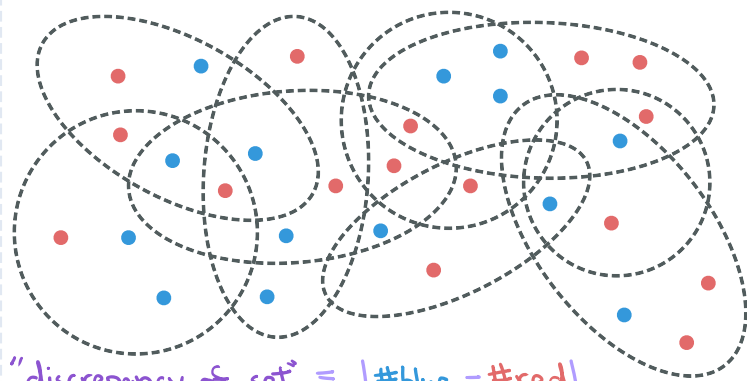
$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

✓ all $x^{(1)}, \dots, x^{(T)} \in P$ w/h/p

$$\text{✓ } E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq (1 - \frac{1}{\delta^2 T}) n - E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right]$$

goal: bound $E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right]$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \quad \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

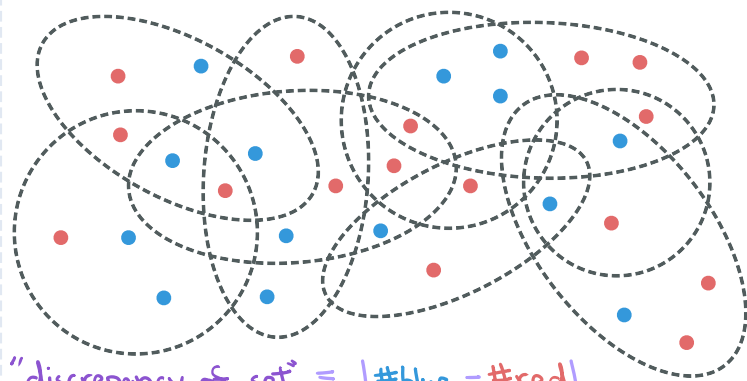
$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

✓ all $x^{(1)}, \dots, x^{(T)} \in P$ w/h/p

$$\text{✓ } E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq (1 - \frac{1}{\delta^2 T})n - E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right]$$

$$\text{claim: } E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right] \leq \frac{n}{4}$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$\chi^{(t)} = \chi^{(t-1)} + \delta U^{(t)} g^{(t)}$$

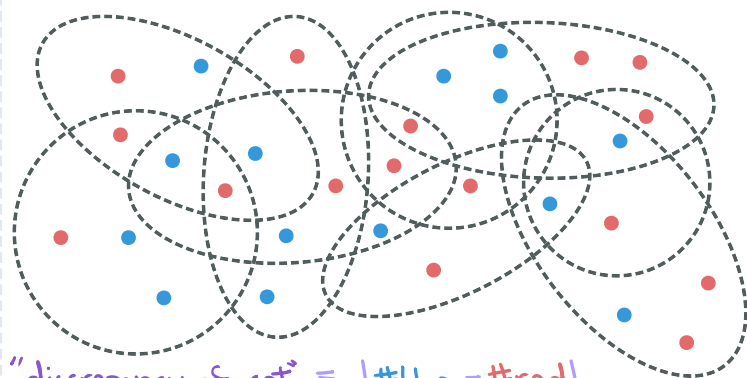
$$\text{claim: } E\left[\left(\# \text{ tight sets } S_i\right)\right] \leq \frac{n}{4}$$

ideas?

for fixed set S_i , what is

$$\chi^{(t)}(S_i) - \chi^{(t-1)}(S_i) ?$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{\chi \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$
where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} \chi_j$

"Fractional coloring" $\chi \in [-1, 1]^n$

χ_i "tight": $|\chi_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$\chi_0 \in [-1, 1]^n \Rightarrow \chi \in [-1, 1]^n$ st.

(a) $|\chi_0(S_i) - \chi(S_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of χ_i 's are tight

Two lemmas about Gaussians

① for $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

sampling a Gaussian vector from $Y^{(t)}$:

$c_1, \dots, c_k$ orthonormal basis of $Y^{(t)}$

$$U = [c_1 | c_2 | \dots | c_k] \in \mathbb{R}^{n \times k}$$

sample $g \sim \mathcal{N}^k(0, 1)$

return $Ug \in \mathbb{R}^n$

$\forall v \in \mathbb{R}^n$:

Lemma: $\|U^T v\| \leq \|v\|$

Lemma: $\langle v, Ug \rangle \sim \mathcal{N}(0, \sigma^2)$
w/ $\sigma^2 \leq \|v\|^2$

$$\chi^{(t)} = \chi^{(t-1)} + \delta U^{(t)} g^{(t)}$$

sampling a Gaussian vector from $Y^{(t)}$:

$c_1, \dots, c_k$ orthonormal basis of $Y^{(t)}$

$$U = [c_1 | c_2 | \dots | c_k] \in \mathbb{R}^{n \times k}$$

sample $g \sim \mathcal{N}^k(0, 1)$

return $Ug \in \mathbb{R}^n$

$$\forall v \in \mathbb{R}^n:$$

$$\text{Lemma: } \|U^T v\| \leq \|v\|$$

$$\text{Lemma: } \langle v, Ug \rangle \sim \mathcal{N}(0, \sigma^2) \text{ w/ } \sigma^2 \leq \|v\|^2$$

fix set S_i .

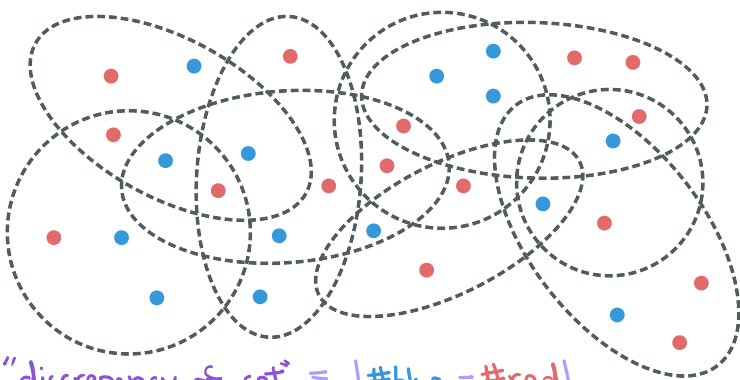
$$\text{let } Y_t = \chi^{(t)}(S_i) - \chi^{(t-1)}(S_i)$$

$$= \delta(U^{(t)} g^{(t)})(S_i)$$

$$\Rightarrow Y_t \sim \mathcal{N}(0, \sigma^2) \text{ w/ } \sigma^2 \leq \delta^2 |S_i| \leq \delta^2 n$$

Balanced coloring

color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$
 where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n \text{ st.}$$

$$(a) |x_0(S_i) - \chi(S_i)| \leq O(\sqrt{n \log(m/n)}) \quad \forall S_i$$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

$$(1) \text{ for } Y \sim \mathcal{N}(0, \sigma^2), \Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$$

$$(2) Y_1, \dots, Y_T \in \mathbb{R} \text{ w/ } Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \text{ for } \sigma_t^2 \leq 1$$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

let $Y_t = \chi^{(t)}(S_i) - \chi^{(t-1)}(S_i)$

✓ $Y_t \sim \mathcal{N}(0, \sigma^2)$ w/ $\sigma^2 \leq \delta^2 n$

$\Pr[S_i \text{ tight}] = \Pr[|Y_1 + \dots + Y_T| \geq \lambda - \epsilon]$

$\leq e^{-\frac{(\lambda - \epsilon)^2}{2T\delta^2 n}}$

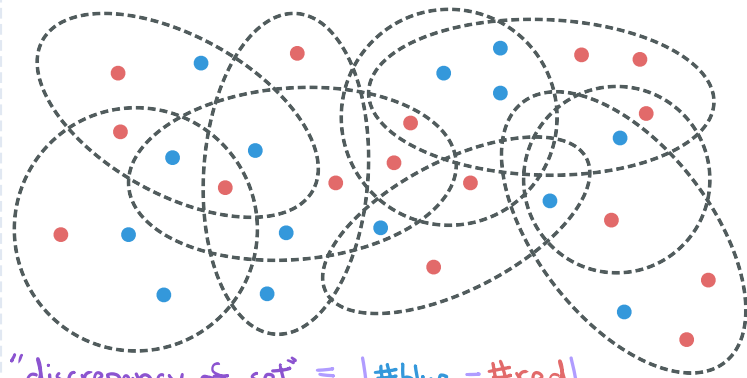
$\leq e^{-\Omega\left(\frac{n \log(m/n)}{n}\right)}$

$\leq \frac{n}{\delta m}$

$(\lambda = c\sqrt{n \log(m/n)} \text{ for } c > 0)$

$(T = O(1/\delta^2))$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{\chi \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$
where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} \chi_j$

"Fractional coloring" $\chi \in [-1, 1]^n$

χ_i "tight": $|\chi_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$\chi_0 \in [-1, 1]^n \Rightarrow \chi \in [-1, 1]^n$ st.

(a) $|\chi_0(S_i) - \chi(S_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of χ_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

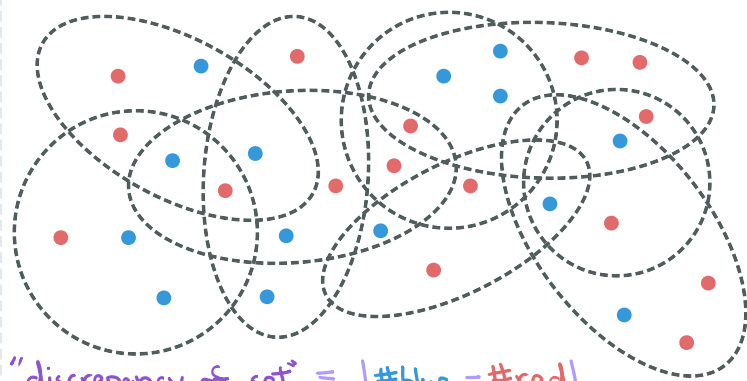
② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$
then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$\checkmark \Pr[S_i \text{ tight}] \leq \frac{n}{8m}$$

$$\Rightarrow E\left[\left(\# \text{ tight sets } S_i\right)\right] \leq \frac{n}{8}$$

Balanced coloring

color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{\chi \in \{-1, +1\}^n} \max_{S_i} |\chi(S_i)|$
 where $\chi(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} \chi_j$

"Fractional coloring" $\chi \in [-1, 1]^n$

χ_i "tight": $|\chi_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$\chi_0 \in [-1, 1]^n \Rightarrow \chi \in [-1, 1]^n$ st.

(a) $|\chi_0(S_i) - \chi(S_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall S_i$

(b) ^(at least) half of χ_i 's are tight

Two lemmas about Gaussians

① for $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1}, Y_{t+1}, \dots, Y_T \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

✓ all $x^{(1)}, \dots, x^{(T)} \in P$ w/h/p

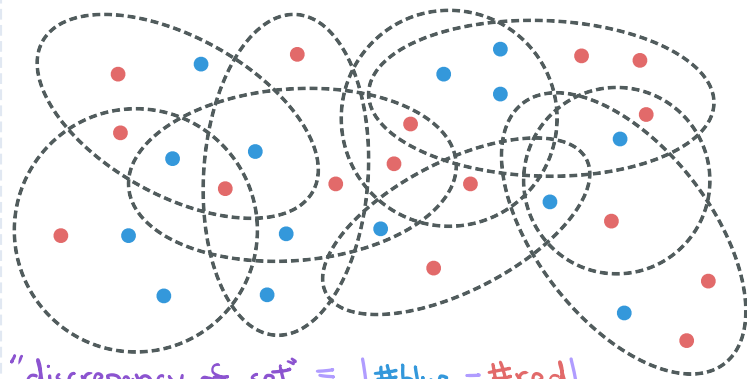
$$\text{✓ } E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq (1 - \frac{1}{\delta^2 T})n - E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right]$$

$$\text{✓ } E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right] \leq \frac{n}{4}$$

$$\Rightarrow E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq \frac{3n}{4}$$

$$\text{for } T = 8/\delta^2$$

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(s_i) - x(s_i)| \leq O(\sqrt{n \log(m/n)})$ $\forall s_i$

(b) ^(at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $\Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_1, \dots, Y_{t-1}, Y_{t+1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$
then $\Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

$$E\left[\left(\sum_{j \in [n]} \text{tight}_j\right)\right] \geq \frac{3n}{4} \Rightarrow \Pr\left[\left(\sum_{j \in [n]} \text{tight}_j\right) \geq \frac{n}{2}\right] \geq \frac{1}{2}$$

why?

$$P = \{x \in [-1, 1]^n : |x_0(s_i) - x(s_i)| \leq \lambda \ \forall s_i\}$$

$$x^{(0)} = x_0, x^{(1)}, \dots, x^{(T)} \in P \quad (T = O(1/\delta^2))$$

$$x^{(t)} = x^{(t-1)} + \delta U^{(t)} g^{(t)}$$

✓ all $x^{(1)}, \dots, x^{(T)} \in P$ w/h/p

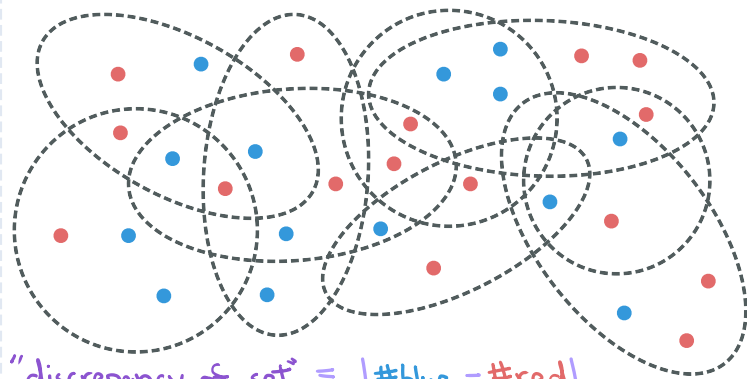
$$E\left[\binom{\# \text{ tight}}{j \in [n]}\right] \geq (1 - \frac{1}{\delta^2 T})n - E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right]$$

$$E\left[\binom{\# \text{ tight}}{\text{sets } S_i}\right] \leq \frac{n}{4}$$

$$Pr\left[\binom{\# \text{ tight}}{j \in [n]} \geq \frac{n}{2}\right] \geq \frac{1}{2}$$

which completes the proof!

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$

goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, +1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

Partial coloring lemma

$x_0 \in [-1, 1]^n \Rightarrow x \in [-1, 1]^n$ st.

(a) $|x_0(S_i) - x(S_i)| \leq O(\sqrt{n \log(m/n)}) \ \forall S_i$

(b) (at least) half of x_i 's are tight

Two lemmas about Gaussians

① For $Y \sim \mathcal{N}(0, \sigma^2)$, $Pr[|Y| \geq \lambda \sigma] \leq 2e^{-\lambda^2/2}$

② $Y_1, \dots, Y_T \in \mathbb{R}$ w/ $Y_t | Y_1, \dots, Y_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$ for $\sigma_t^2 \leq 1$

then $Pr[|Y_1 + \dots + Y_T| \geq \lambda \sqrt{T}] \leq 2e^{-\lambda^2/2}$

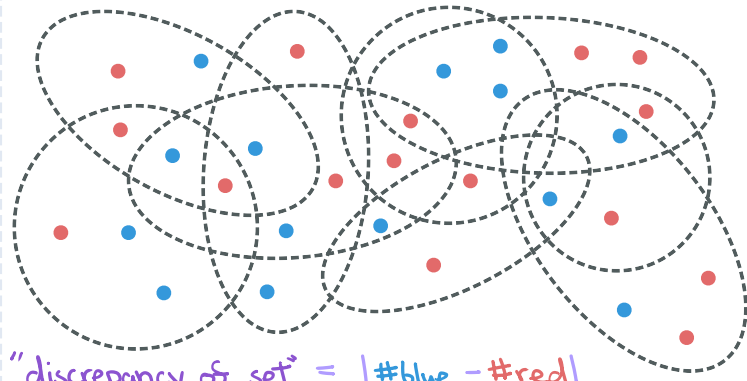
Partial coloring lemma - detailed

given $x_0 \in [-1, 1]^n$, rand. algo
computes $x \in [-1, 1]^n$ st. (w/ high prob)

$$(a) |x_0(S_i) - x(S_i)| \leq O(\sqrt{n \log(m/n)}) \quad \forall S_i$$

(b) ^(at least) half of x_i 's are tight

Balanced coloring
color points red/blue st. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$
Goal: $\min_{x \in \{-1, 1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$
 x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

(x_0 used to adjust for fractional coloring from previous iterations)

Partial coloring lemma

(high-level)

rand. algo. computes $x \in [-1, 1]^n$ s.t.

$$(a) |x(S_i)| \leq O(\sqrt{n \log(m/n)}) \quad \forall S_i$$

(b) ^(at least) half of coordinates are tight

Partial $\rightarrow$ full tight coloring (high-level)

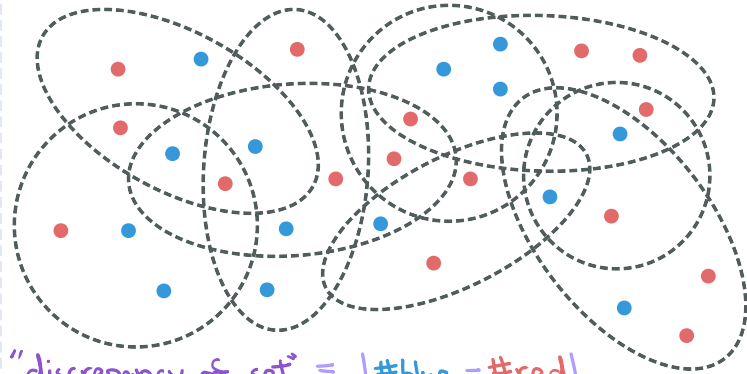
Repeatedly until all x_i tight

(a) partially color

(b) fix $n/2$ tight coordinates, remove them

Note: " $\sqrt{n \log(m/n)}$ " decreases geometrically

Balanced coloring
color points red/blue s.t. all sets have same # of each



"discrepancy of set" = $|\# \text{blue} - \# \text{red}|$
goal: minimize max discrepancy

Input: $S_1, \dots, S_m \subseteq [n]$

Goal: $\min_{x \in \{-1, 1\}^n} \max_{S_i} |x(S_i)|$
where $x(S_i) \stackrel{\text{def}}{=} \sum_{j \in S_i} x_j$

"Fractional coloring" $x \in [-1, 1]^n$

x_i "tight": $|x_i| \geq 1 - \epsilon$ where $\epsilon = \frac{1}{n}$

