Randomized Algos Lecture #3

Concentration of Measure.

- better "concentration" bounds.
  - Chernoff–Hoeffding Bounds.
- Applications
  - Balls and Bins
  - Low congestion Routing
  - Power of 2 choices.
  - Graph Sparsification

  → Next lecture?
  - (Exercise: Mean Estimation / Parameter Estimation).

  ... and tons of other examples.

Last time:

Markov: $\forall X \geq 0$: $Pr(X \geq \lambda) \leq \frac{EX}{\lambda}$ $\forall \lambda \geq 0$

X better have EX defined

Chebyshev: $\forall$ r.v. X $Pr(|X-\mu| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}$ $\forall \lambda \geq 0$

need second moment to be defined too.

Today: use "all moments".

Actually: Chebyshev is tight

$$X = \begin{cases} -\lambda & \text{wp } 1/2\lambda^2 \\ +\lambda & \text{— —} \\ 0 & \text{otherwise} \end{cases}$$

$E(X)=0$.
$Var(X) = EX^2 = 1$

$Pr(|X| \geq \lambda) = \frac{1}{\lambda^2}$.

So what can we do that's better?

Assume something stronger about the random variable
$\Rightarrow$ prove something stronger.

Today's running examples: sums of bounded independent r.v.s

$X_i \sim Ber(p)$. "Bernoulli (p) r.v"

$X = \sum_{i=1}^{n} X_i$

$Y_i = \begin{cases} -1 & \text{wp } 1/2 \\ 1 & \text{wp } 1/2 \end{cases}$ "Rademacher r.v."

$Y = \sum_{i=1}^{n} Y_i$

$\mu = EX = \sum EX_i = np$.
$\sigma^2 = Var(X) = np(1-p)$

$\mu = EX = \sum EX_i = 0$.
$\sigma_i^2 = Var(X_i) = EX^2 = 1$;
$\sigma^2 = \sum \sigma_i^2 = n$
"random walk on the integers"

Note: if $p = 1/2$, then $X_i = \frac{1+Y_i}{2}$

so get $EX = n\left(\frac{1+0}{2}\right) = n/2$

$Var(X) = n(1/2)^2 = n/4$.

Chebyshev says: $\Pr[\geq 750$ heads in $\underbrace{1000}_{n}$ flips has prob.$]$

$$= \Pr[Y \geq \underbrace{250}_{n/4}] \leq \frac{\mathrm{Var}(X)}{(n/4)^2} = \frac{16n}{n^2} = 0.016.$$

But it turns out: $\Pr(Y \geq 250) \leq 10^{-54}$. (!!)

How do we prove such strong bounds?

Before we answer this, one last digression.

Suppose sum a collection of independent identical r.v.s (iid r.v.s)

$X_1 X_2 \ldots X_n$ $\quad E X_i = \mu \quad V(X_i) = \sigma^2$

and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be their sample mean

then as $n \to \infty$ $\quad \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$

converges in distribution

standard normal "Gaussian" r.v.

$\boxed{\text{Central Limit thm}}$ $\quad$ " $\frac{\sum (X_i - \mu)}{\sqrt{n\sigma^2}}$.

So, loosely, the tails of the sum of n iid random variables with {zero mean, variance 1 say} should start looking like the tails of the gaussian.

i.e. $\sum_i X_i \approx N(0,1) \cdot \sqrt{n\sigma^2}$

$= N(0, n)$ ← unit variance

$$\Rightarrow \Pr\left(\sum_i X_i \geq \lambda\right) \lessapprox e^{-\lambda^2/n}.$$

This is indeed close to the truth!!

But the CLT is a limiting statement!

How do we give "finite n" claims?

Quantitative bounds?

Ans: Chernoff bounds. (also Cramér, Hoeffding, Bernstein, Rubin ...)

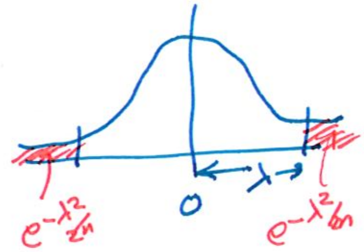
Thm: Concentration Bounds for Rademachers

$Y_i \sim$ iid Rademacher rvs. $\quad Y = \sum_{i=1}^{n} Y_i$

then $\Pr(Y \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2n}\right)$

$EY = 0$
$V(Y) = n.$

Note: Y is symmetric about 0, so $\Pr(Y \leq -\lambda) = \Pr(Y \geq \lambda)$.



Compare to Chebyshev, which says $\leq n/\lambda^2 = 1/\left(\frac{\lambda^2}{n}\right)$

So if $n = 1000$, $\lambda = 250 = n/4$ then get $\exp\left(-\frac{n^2/16}{2 \cdot n}\right) = \exp\left(-\frac{n}{32}\right) \leq \exp(-30)$.

vs. 0.016

Much better!!

— X —

In particular, says that if $\lambda = k\sqrt{n} = k\sigma$ (for this case)

then tail probability $\leq \exp(-k^2/2)$.

~~exponential tails~~ or "Gaussian" tails.

vs. Chebyshev's $\leq 1/k^2$ $\quad$ inverse "polynomial tails"

But needs more:

sum of "small" "independent" random variables.

→ later lectures / HWs

Each of these can be relaxed a bit, but still need strong assumptions to get such exponential / gaussian tails.

Before we prove this, give another tail bound.

Thm 2: Sum of Independent & Bounded RVs.

$X_1, X_2, \ldots, X_n$ indep. r.v.s. $\in [0,1]$ ← not necc int. valued.

st $EX_i = p_i$. ← not necc identical

$X = \sum_{i=1}^{n} X_i$ $\Rightarrow EX = \sum p_i = \mu$ (say).

then. ① $Pr(|X-\mu| \geq \varepsilon\mu) \leq 2.\exp(-\varepsilon^2\mu/3)$ for all $\varepsilon \in [0,1]$.

② $Pr(X-\mu \geq \varepsilon\mu) \leq \exp\left(-\frac{\varepsilon^2}{2+\varepsilon}\mu\right)$ for all $\varepsilon \geq 0$

Gaussian behavior for deviations $\leq \mu$

~~Wierder~~ behavior when $\varepsilon \geq 1$.
Poisson

if you just want $Pr(X-\mu \geq \varepsilon\mu)$ then drop the factor of 2 on RHS.

Examples:

(a) $X_i = Ber(1/2)$ $\forall i$ $\Rightarrow \mu = n/2$

$\Rightarrow Pr(|X-\mu| \geq \varepsilon n) \leq 2\exp(-\varepsilon^2 n/6)$ for $\varepsilon \leq 1/2$.

e.g. $Pr(|X-\mu| \geq K\sqrt{n}) \leq 2.\exp(-K^2/6)$ very similar to previous bound.

(b) $X_i = Ber(1/n)$ $\forall i$ $\Rightarrow \mu = n.1/n = 1$.

$Pr(X-1 \geq t) \leq \exp\left(-\frac{t^2}{2+t}\mu\right) \leq \exp(-t/3)$

!

so $Pr(X \geq 6\log n + 1) \leq 1/n^2$.

Easy, very little calculations!!

n balls and n bins

says that load on any fixed bin $\leq 6\log n + 1$ whp. $1 - 1/n^2$

(union bound) $\Rightarrow$ load on every bin $\leq (6\log n + 1)$ wp. $1 - (1/n^2 \cdot n) = 1 - 1/n$.

OK: did not get $O\left(\frac{\log n}{\log\log n}\right)$ ☹

Not fault of these bounds!

We did not give strongest bounds. (Later).

a bit ~~slightly~~ messier but not too bad.

in fact proof gives directly, we'll see.

For now: let's (a) see intuition / proof

(b) some more applications

—— x ——

(Almost)

Proof of Rademacher Bound (Other is very similar)

← choose t later!

← "monotone" operation, "rescaling" $\forall t \geq 0$.

$$\Pr(X \geq \lambda) = \Pr(e^{tX} \geq e^{t\lambda})$$

$$\leq \frac{\mathbb{E}[e^{tX}]}{e^{t\lambda}}$$ by Markov

$$= \frac{\mathbb{E}[e^{tX_1}]\,\mathbb{E}[e^{tX_2}] \ldots \mathbb{E}[e^{tX_n}]}{e^{t\lambda}}$$ independence.

OK. $\mathbb{E}[e^{tX_i}] = \frac{1}{2}\cdot e^{t} + \frac{1}{2}\cdot e^{-t} = \frac{e^{t}+e^{-t}}{2}$

let's use the following approx : $\boxed{e^t \leq 1+t+t^2}$ for $|t| \leq \frac{1}{2}$.

$\Rightarrow \mathbb{E}[e^{tX_i}] \overset{(*)}{\leq} (1+t^2)$. Now substitute back....

$$\Pr(X \geq \lambda) \leq \frac{(1+t^2)^n}{e^{t\lambda}}$$ approx as above, plus identical...

$$\leq \frac{e^{t^2 n}}{e^{t\lambda}}$$ $\boxed{1+x \leq e^x}$ $\forall x$

$$= e^{-\underbrace{(t\lambda - t^2 n)}}$$

↑ choose t so that ~~want~~ this ~~to~~ be large. quadratic. so take derivatives wrt t

$\lambda = 2tn \Rightarrow t = \frac{\lambda}{2n} \leq \frac{1}{2}$ for all $\lambda \leq n$.

factor 2 weaker in exponent, but so easy.... ☺

$$\boxed{\leq e^{-\lambda^2/4n}}$$ ← for $t = \lambda/2n$.

Note: to get $e^{-t^2/2n}$ (or even better) don't use $\mathbb{E}[e^{tX_i}] \le 1+t^2$

but keep it as $\frac{e^t+e^{-t}}{2} = \cosh(t)$

and optimize later.

— X —

Great. But what happened? What's the intuition.

For that, let's consider the general case:—

X_i indep r.vs. $\in [0,1]$. $X = \sum_i X_i$

$EX_i = p_i \Rightarrow EX =: \mu = \sum_i p_i$

$\Pr(X \geq \mu+\lambda)$ ~~$\mathbb{E}[e^{tX}]$~~ ~~$\Pr[e^{tX}$~~

$= \Pr(g(X) \geq g(\mu+\lambda))$

for monotone function $g(z)$.

$g(z) = z^2$ gives Chebyshev

$= e^{tz}$ gives Chernoff

$\le \dfrac{\mathbb{E}[g(X)]}{g(\mu+\lambda)}$

to make this small, want $\mathbb{E}[g(X)]$ to be small

$g(\mu+\lambda)$ large.

Say $g(z) = e^{tz}$...

$= \dfrac{\mathbb{E}[e^{tX}]}{e^{t\mu+t\lambda}}$.

Suppose miraculously $\mathbb{E}[e^{tX}] \approx e^{t\mathbb{E}[X]} = e^{t\mu}$

then we'd get $\frac{1}{e^{t\lambda}}$ and can make this go to zero

by choosing $t \to \infty$.

But why is this going to be true? ☺

OK: when is $\mathbb{E}[e^{tX}] \approx e^{t\mathbb{E}[X]}$?

sps tz is tiny then $e^{tz} \approx 1+tz$

thankfully $X = X_1 + \cdots + X_n$ (sum of indep rvs)

so $\mathbb{E}[e^{tX}] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}]$

$\boxed{g(a+b) = e^{t(a+b)} = e^{ta} \cdot e^{tb} = g(a)g(b) \text{ so the functional form of } g(\cdot) \text{ helps here!}}$

and each $X_i \in [0,1]$ say — small

so if t is small then

$$\mathbb{E}[e^{tX_i}] \cong \mathbb{E}[1 + tX_i] = 1 + t\,\mathbb{E}[X_i]$$
$$\leq e^{t\mathbb{E}[X_i]}$$

So: want t small so that $\mathbb{E}[e^{tX}]$ is close to $e^{t\mu}$ in the denom. (approx cancelled by)

want t large to get $e^{t\lambda}$ large in the denom.

Balance the two terms! thats the game here. (and all the intuition I have).

— X —

For the rest of the proof of the general case, see the notes I've provided.

Same idea, but more little tricks to control the approximations

— X —

I'd suggest going back and reading over the Rademacher proof and seeing why $t = \frac{\lambda}{2n}$ makes sense ....

Applications of Chernoff Bounds

(A) Mean Estimation (Polling).

Take n ~~copies of~~ indep. samples from distribution. to estimate Y/N. or 1/0

$X_1, X_2 \dots X_n$ $\quad X = \sum_i X_i$ $\quad$ (say p fraction are Y, or 1)

each Ber(p). $\quad \bar{X} = \frac{1}{n} \sum_i X_i = X/n$

$EX = \mu = np$.

$$\Pr(|X - \mu| \geq \overbrace{n\varepsilon}^{\lambda})$$

$$= \Pr(|\bar{X} - p| \geq \varepsilon)$$

$$\leq \exp\left(-\frac{\lambda^2}{2\mu + \lambda}\right)$$

(convenient restatement of the $\left(\frac{\varepsilon^2}{2+\varepsilon}\mu\right)$ bound)

$$= \exp\left(-\frac{\varepsilon^2 n^2}{(2p+\varepsilon)n}\right)$$

$$\leq \exp\left(-\frac{\varepsilon^2 n}{3}\right) \leq \delta$$

if $n = \frac{3}{\varepsilon^2} \log 1/\delta$.

So: if you take $\frac{3}{\varepsilon^2} \log 1/\delta$ samples and use the sample mean, then it is within $\underbrace{\pm\varepsilon \text{ of}}$ correct bias p $\underbrace{\text{wp } 1-\delta}$

nearly accurate

high confidence, set to 95% say.

confidence interval $[p-\varepsilon, p+\varepsilon]$

ε = margin of error.

(congestion min #1)

(B) Routing with Small Congestion

Given graph $G = (V, E)$.

"Terminal" Pairs $(s_i, t_i) \in V \times V$. $i = 1..k$

Want to find paths P_i from s_i to t_i (one path per pair)

Such that # of paths using any edge is minimized.

$\max_{e \in E} \{\#\text{ paths using edge } e\}$:= def as "congestion" of the routing.

Thm: this problem is NP-hard (if exact algo in polytime $\Rightarrow$ $P = NP$). See Karp's original paper.

Thm: Sps the optimal congestion = OPT

then $\exists$ a polytime algo with congestion $\leq OPT \cdot O(\log m)$.

m = #edges.

— x —

Idea of algorithm:

(1) Find optimal fractional routing (in poly-time) deterministic.

(2) Randomly "round" to integer paths while keeping low congestion.

— x —

Step 1: write Integer Linear Program. Let $\mathcal{P}_i$ = paths from s_i to t_i

ILP =

$$\min \cancel{\sum_{P \in \mathcal{P}_i} x_P} \; Z$$

$$\text{st} \sum_{P \in \mathcal{P}_i} x_P = 1 \quad \forall i.$$

$$\sum_{i=1}^{k} \sum_{\substack{P \in \mathcal{P}_i \\ e \in P}} x_P \leq Z \quad \forall e$$

$$1 \leq Z$$

$$x_P \in \{0,1\}$$

Fact 1: ILP = exactly the optimal solution

Def: LP: replace $x_P \in \{0,1\}$ by $x_P \in [0,1]$

Fact 2: $LP \leq ILP$

and moreover LP can be solved in polynomial time (congestion min) #2

Great: Onto Step (2)

(a) Solve LP to get solution $\{x^*_p\}_{p \in \cup_i \mathcal{P}_i}$

(b) $\forall i$, pick one (randomly) path from $\mathcal{P}_i$ where $P \in \mathcal{P}_i$ picked w.p. x_p.

Fact 3: define $z_{e,i} = \sum_{P \in \mathcal{P}_i}' x_p$.

then edge e congested by path for pair s_i-t_i wp $z_{e,i}$

Now: Let $X_{e,i} = \mathbb{1}$(path for pair s_i-t_i uses e)

$$\mathbb{E} X_{e,i} = z_{e,i}$$

$X_{e,i}$ and $X_{e,j}$ are indep for $i \neq j$

$X_e = \sum_i' X_{e,i}$ = congestion of edge e, $\mathbb{E} X_e = \underbrace{\sum_i' z_{e,i}}_{z_e} \leq z$

$\Rightarrow$ by Chernoff bound, ~~$\Pr[X_e \geq \mathbb{E}X_e \ldots] \leq$~~

$$\Pr[X_e \geq z_e + cz \log m] \leq \exp\left(-\frac{(cz\log m)^2}{2z_e + cz\log m}\right)$$

$$\leq \exp\left(-\frac{cz\log m}{3}\right)$$

$$\leq \frac{1}{\text{poly}(m)}$$ since $z \geq 1$ can set c to large constant.

$\Rightarrow \Pr[\exists \text{ edge with } X_e \geq cz\log m + z]$

$$\leq \frac{m}{\text{poly}(m)}.$$

$\Rightarrow$ max congestion $\leq O(\text{OPT} \cdot \log m)$ whp. ∎