[SCRIBE] · Dubhashi-Panconesi · McDiarmid survey · Mitz-Upf · Joag-Dev—Proschan. ①

Lecture 3: Chernoff/Tail/Large Deviation Bounds

[[Massive naming confusion. CS people call them "Chernoff" bds after a '52 paper. Or sometimes Hoeffding, who did sthg strongerish in '40s. Though Bernstein '20s was even strongerish.]]

Let H = # heads in n [[fair]] flips.

rec: Berry-Esseen $\Rightarrow$ $\Pr[H \geq \frac{n}{2} + \frac{\sqrt{n}}{2} \cdot t] \approx \overline{\Phi}(t)$

$\sim \frac{\phi(t)}{t} \leq \phi(t)$ [[$t \geq 1$]]

$\leq e^{-t^2/2}$.

Let $t = 10\sqrt{\ln n}$. $= e^{-50 \ln n} = \frac{1}{n^{50}}$.

So $\Pr[H \geq \frac{n}{2} + 5\sqrt{n \ln n}] \leq \frac{1}{n^{50}}$? Wrong: $\boxed{\approx O(\frac{1}{\sqrt{n}})}$

[[But actually, this claim is more-or-less (in fact, indeed) correct. But CLT doesn't give it; CLT is only good for "small devs" from mean. Need "Chernoff Bds" for "large" devs]]

Bounding r.v.s: more info = better bounds

① Only know mean

Say r.v. X is ≥ 0, $X \not\equiv 0$.

Markov Ineq. [[Major workhorse when you know nothing.]]

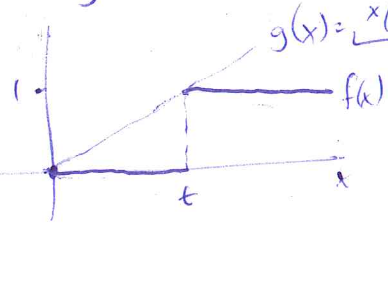
$$\Pr[X \geq t \cdot E[X]] \leq \frac{1}{t} \qquad \forall t > 0 \ [[\geq 1]]$$

pf 1: "words": WOLOG $E[X] = 1$. [[Why? Div X by its mean. mean 0 case triv.]]

$\Pr[X \geq t] \leq 1/t$?

[[Well, mean is 1. How could X be $\geq t$ with prob $> \frac{1}{t}$? That would already make X's mean $> \frac{1}{t} \cdot t = 1$.]]

pf 2: "pictures"



$\Pr[X \geq t] = E[f(X)]$

$\leq E[g(X)]$

$= E[X/t]$

$= 1/t$. □

② Similar fact: Say $0 \le X \le 1$ always, $E[X] = \varepsilon$. Then
$\Pr[X \ge \varepsilon/2] \ge \varepsilon/2$. ($\ge$)

"Markov-type"
"averaging arg."

Words pf: [[Sps $\Pr[X \ge \varepsilon/2] < \varepsilon/2$. Biggest $E(X)$ could be is if whenever it's $\ge \frac{\varepsilon}{2}$ it's 1, and rest of time it's $\varepsilon/2$. But that would still make its expec. $< \frac{\varepsilon}{2} \cdot 1 + 1 \cdot \frac{\varepsilon}{2} < \varepsilon$.]]

---

② Know abt. mean & variance. ($\sqrt{Var}$)

"AKA Second-Moment Method"

Chebyshev Ineq.: Let $E[X] = \mu$, stddev$[X] = \sigma \ne 0$.

Then $\forall t > 0$, $\Pr[|X - \mu| \ge t \cdot \sigma] \le \frac{1}{t^2}$.

[[words interp.]]

pf: WOLOG $\mu = 0$ [[why?]], $\sigma = 1$ [[why?]]
$\Leftrightarrow E[X^2] = 1$.

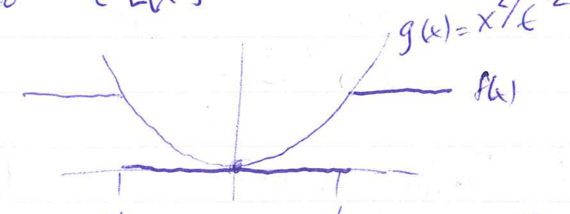
Need $\Pr[|X| \ge t] \le \frac{1}{t^2}$.

(same event)

Pf 1: Markov: $= \Pr[X^2 \ge t^2] \le \frac{1}{t^2}$. □

≥ 0 $\qquad$ $= t^2 \cdot E[X^2]$

2: Pics: $E[f(X)] \le E[g(X)]$



□

e cor: Say $X \ge 0$ always. Then $\Pr[X = 0] \le \frac{\sigma^2}{\mu^2} = \frac{Var[X]}{E[X]^2} \le \frac{E[X^2]}{E[X]^2}$

---

③ Common scenario: $X = X_1 + \cdots + X_n$, X_is indep., or "kinda indep."

E.g. coin flips: $H = X_1 + \cdots + X_n$
$\left\{ \begin{array}{l} 1 \text{ w.p. } 1/2 \\ 0 \text{ w.p. } 1/2 \end{array} \right\}$

⟦this will occupy us for a while⟧

⟦yuck!⟧

$$\Pr\Big[H \ge \underbrace{\tfrac{n}{2}}_{\mu} + \underbrace{\tfrac{\sqrt{n}}{2}}_{\sigma}\cdot 10\sqrt{\ln n}\Big] \overset{?}{\le} (\text{Markov}) \overset{ex}{\Rightarrow} 1-\Theta\Big(\frac{\sqrt{\ln n}}{\sqrt{n}}\Big).$$

☹

(Chebyshev?)

$$\overset{\text{symm.}}{=} \tfrac{1}{2}\Pr\big[|H-\mu| \ge 10\sqrt{\ln n}\cdot\sigma\big] \le \frac{1}{100\ln n}$$

⟦$\to 0$ at least. Still, truth is like $\frac{1}{n^{50}}$.⟧

☹

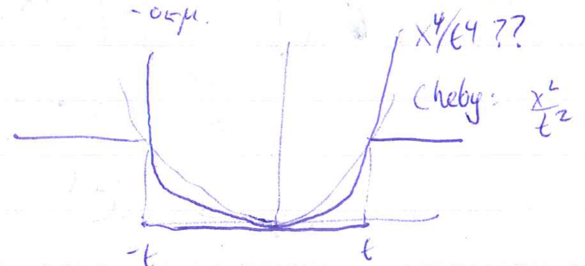
(*) ☺ Chebyshev doesn't need independence, just "pairwise indep." ⟦← much commoner in practice⟧.

$$\begin{aligned}\mathrm{Var}[X_1+\dots+X_n] &= E[(X_1+\dots+X_n)^2] - E[X_1+\dots+X_n]^2 \\ &= \sum_i E[X_i^2] + \sum_{i\ne j} E[X_iX_j] - \dots \\ &= \sum_{i\ne j} E[X_i]E[X_j] \quad \text{if pairwise indep} \\ &= \text{What } \mathrm{Var}[X_1+\dots+X_n] \text{ would be, if fully indep.}\end{aligned}$$

⟦Back to coins. How can we beat $\frac{1}{100\ln n}$?⟧

Switch to $X = X_1+\dots+X_n$, X_i indep $\begin{cases}+1 & \text{w.p. } 1/2\\ -1 & \text{w.p. } 1/2.\end{cases}$

$$\Pr\big[|X| \ge 10\sqrt{\ln n}\cdot \underbrace{\sqrt{n}}_{\text{stddev}[X]}\big]$$



"4th moment method": ⟦works w/ 4-wise indep. r.v.s⟧

⟦Markov ver.⟧:
$$\Pr[|X| \ge t\sqrt{n}] = \Pr[X^4 \ge t^4 n^2] \le \frac{E[X^4]}{t^4 n^2}.$$

$E[X^4]$? $(\sum X_i)^4 = \sum_i X_i^4 + X_i^3 X_j$ terms $+ X_i^2 X_j^2$ terms $+ X_i^2 X_j X_k$ $+ X_i X_j X_k X_\ell$

In expec: $n \cdot E[X_i^4] = n$. $\quad 0 \quad 1 \cdot \underbrace{\binom{n}{2} \cdot \binom{4}{2}} \quad 0 \quad 0.$ [[how many?]]

$= 3n^2 - 3n$

$= 3n^2 - 2n \le 3n^2$.

$\therefore \Pr[|X| \ge t\sqrt{n}] \le \frac{3n^2}{t^4 n^2} = 3/t^4$. $\quad = \frac{3}{100 \ln^2 n}$:)

for $t = 10\sqrt{\ln n}$.

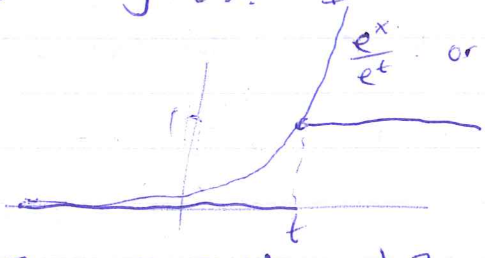
Keep going? $E[X^{2s}] \le C_s \cdot n^s$

grows with s, kind of pain.

$\therefore \Pr[|X| \ge t\sqrt{n}] \le \frac{C_s}{t^{2s}}$. Optimize over s...

[[Pain. More flexible way....?]]

"Chernoff method": $\quad \frac{e^x}{e^t}$ or $\frac{e^{\lambda x}}{e^{\lambda t}}$.



Markov: $\Pr[X \ge u] = \Pr[e^{\lambda X} \ge e^{\lambda u}]$ for any $\lambda > 0$.

$\overset{[\text{Markov}]}{\le} \frac{E[e^{\lambda X}]}{e^{\lambda u}}$ ← [[a fcn of λ, called M.G.F.]]

[[a number]] → $e^{\lambda u}$

easy to compute if $X = X_1 + \dots + X_n$ indep.

Plan: Compute/bound it, then optimize λ.

$$E[e^{\lambda X}] = E[\exp(\lambda X_1 + \cdots + \lambda X_n)]$$
$$= E[\exp(\lambda X_1)\cdot \cdots \cdot \exp(\lambda X_n)]$$
$$= E[\exp(\lambda X_1)]\cdot \cdots \cdot E[\exp(\lambda X_n)] \quad \text{(by indep.)}$$
$$= E[\exp(\lambda X_1)]^n$$
$$= \left[\tfrac{1}{2}e^{\lambda} + \tfrac{1}{2}e^{-\lambda}\right]^n$$

$1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots \qquad 1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \cdots$

$$= \left[1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \cdots\right]^n$$
$$\le \left[e^{\lambda^2/2}\right]^n \qquad 1 + \frac{\lambda^2}{2} + \frac{\lambda^4}{2^2\cdot 2!} + \frac{\lambda^6}{2^3\cdot 3!} + \cdots$$
$$= e^{\frac{\lambda^2 n}{2}}$$

So $\Pr[X \ge u] \le \dfrac{e^{\frac{\lambda^2}{2}\cdot n}}{e^{\lambda u}} = \exp\left(\frac{n}{2}\lambda^2 - u\lambda\right)$

minimize over λ

$\lambda := \frac{u}{n}$

$= e^{-\frac{u^2}{2n}}$.

So for $u = 10\sqrt{\ln n}\cdot\sqrt{n}$, $\le \frac{1}{n^{50}}$:)

[[Specific to sum of iid 50/50's]]

Next: treat $X_i = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{w.p. } 1-p_i \end{cases}$. Must bound $E[\exp(\lambda X_i)]$, choose λ well.

Moderately annoying calcs . . .

$\Longrightarrow$

6

Let $X = X_1 + \cdots + X_n$, X_i's indep.
Let $\mu = E[X] = \Sigma E[X_i]$

"Hoeffding": Say $a_i \le X_i \le b_i$ always.
Then $\Pr[X \ge \mu + t]$, $\Pr[X \le \mu - t]$ $\le \exp\left(-\frac{2t^2}{\Sigma (b_i - a_i)^2}\right)$ $\forall t > 0$

[Jibes w/ our result for $X = \pm 1$]

[better if μ is surprisingly small]

"Chernoff": Say $0 \le X_i \le 1$ $\forall i$.

$\forall \varepsilon > 0$,
- $\Pr[X \le (1-\varepsilon)\mu_L] \le \exp\left(-\frac{\varepsilon^2}{2}\mu_L\right)$
- $\Pr[X \ge (1+\varepsilon)\mu_H] \le \exp\left(-\frac{\varepsilon^2}{2+\varepsilon}\mu_H\right)$ [!]

[can't delete]

$\le \exp\left(-\frac{\varepsilon^2}{3}\mu_H\right)$ if $\varepsilon \le 1$.

[Memorize this one like a poem. Look up the rest]

Fact: If $\mu_L \le \mu \le \mu_H \ldots$

[Give "sampling thm" 1 mult.?]

[Bonus results] Neg. Dep. r.v.'s: do first

def: $X_1, \ldots, X_n$ are negatively associated (NA) if:
$$E[f(X_i : i \in A)\, g(X_i : i \in B)] \le E[f(X_i : i \in A)]\, E[g(X_i : i \in B)]$$

$\forall$ nondecr f, g, $\forall$ disjoint $A, B \subseteq [n]$.

$\Rightarrow \operatorname{Cov}[X_i, X_j] \le 0$

$\not\Leftarrow$

ex: $\Rightarrow$ $E[e^{\lambda X_1 + \cdots + \lambda X_n}] \le E[e^{\lambda X_1}] \cdot \cdots \cdot E[e^{\lambda X_n}]$

$\Rightarrow$ Hoeffding/Chernoff holds

Facts: • indep $\Rightarrow$ N.A.

• closed under subsets, indep unions

• " " applying nondecr fcns to disjoint subsets; e.g.: $X_1+X_2+X_3,\ X_4+\cdots+X_{10},\ X_{11}+X_{12}, \cdots$

eg: [see Joag-Dev – Proschan '83]

① Put n pts randomly on unit circle, $X_1, \ldots, X_n$ = arc lengths

② Throw N balls into n bins [[diff probs for each bin " " ball ok]]
$X_1, \ldots, X_n$ occupancies

③ Let $x_1, \ldots, x_n$ be fixed #s, $X_1, \ldots, X_n$ a rand permutation. $\Rightarrow$ sampling from finite population WITHOUT replacement.

All N.A. → Intuitive meaning: knowing that some values are "large" makes it more likely other values are "small".

Martingales: Let $X_1, \ldots, X_n$ be any discrete r.v.'s. [not nec indep]
Let $f: \mathbb{R}^n \to \mathbb{R}$
Let $Y_i = E[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_{i-1}]$, so
$Y_0 = E[f(X)]$, (constant r.v.) $Y_n = f(X_1, \ldots, X_n)$

e.g. $f(X) = \sum X_i c_i$, $|X_i| \le c_i$

Then $\mu := Y_0, Y_1, \ldots, Y_n$ is a "martingale" (w.r.t. $X_1, \ldots, X_n$), meaning $E[Y_i \mid X_1, \ldots, X_{i-1}] = Y_{i-1}$ $\forall i \in [n]$.

"Azuma" / "Method of Bdd Diffs"
thm: Suppose $X_1, \ldots, X_n$ indep, f sats $|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \le c_i$ always

Then $\left.\begin{array}{l}\Pr[f(X) \ge \mu + t] \\ \Pr[f(X) \le \mu - t]\end{array}\right\} \le \exp\left(-\dfrac{2t^2}{\sum c_i^2}\right)$.