Lecture 3: Chernoff/Tail/Large Deviation Bounds

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

Let $H = \# \text{heads in n \text{ fair flips}}$.

Rec: Berry-Esseen $\Rightarrow$ $\Pr[H \geq n/2 + 5\sqrt{n} \epsilon] \sim \Theta(\epsilon^3)$

$\sim O(\epsilon^3) \leq O(1)$ if $\epsilon < 1/2$

Let $\epsilon = 10 \sqrt{n/n^4} = e^{-50n} \approx \frac{1}{n^{50}} \Rightarrow e^{-e^{2n}}$

So $\Pr[H \geq n/2 + 5\sqrt{n} \epsilon] \leq \frac{1}{n^{50}}$? Wrong.

$\Rightarrow O(\epsilon^3)$

But actually, this claim is more-or-less (in fact, indeed) correct.

But CLT doesn't give it; CLT is only good for "small dev" from mean. Need "Chernoff bounds" for "large dev" cases.

Bounding rvs: more info = better bounds

1. Only know mean
2. Say r.v. $X$ is $\geq 0$, $X \neq 0$.

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

$\Pr[X \geq \epsilon E[X]] \leq \frac{1}{\epsilon}$ for $\epsilon > 0$ ($\geq 1$)

PF 1: "words": WLOG $E[X] = 1$. [Why? Div X by its mean.]

$\Pr[X \geq \epsilon] \leq \frac{1}{\epsilon}$?

Well, mean is 1. How could $X$ be $\geq \epsilon$ with $\Pr[X \geq \epsilon] \leq \frac{1}{\epsilon}$?

That would already make $X$'s mean $> \frac{1}{\epsilon} \epsilon = 1$,]

PF 2: "pictures"

$y(x) = \frac{f(x)}{t}$

$\Pr[X \geq \epsilon] = \int_{\epsilon}^{\infty} f(x) \frac{x}{t} dx \\
\leq \int_{\epsilon}^{\infty} f(x) \frac{x}{\epsilon} dx \\
\leq \epsilon \int_{\epsilon}^{\infty} f(x) dx = \epsilon \epsilon$.
2. Know $\mu$, mean & variance.

Chebyshev Ineq: Let $E[X] = \mu$, std dev $[X] = \sigma > 0$.

Then $\forall \varepsilon > 0$, $\Pr \left[ |X - \mu| > \varepsilon \right] \leq \frac{1}{\varepsilon^2}$.

Proof:

Without loss of generality, let $\mu = 0$ and $\sigma = 1$.

$$\Rightarrow E[X^2] = 1.$$

Need $\Pr \left[ |X| > \varepsilon \right] \leq \frac{1}{\varepsilon^2}$.

Proof 1: Markov:

$$\Pr \left[ X^2 > \varepsilon^2 \right] \leq \frac{1}{\varepsilon^2},$$

where $g(x) = x^2/\sigma^2$.

Ex. cor: Say $X > 0$ always, then $\Pr \left[ X > \varepsilon \right] \leq \frac{1}{\varepsilon^2}$.

Common scenario: $X = X_1 + \ldots + X_n$, $X_i$ is independent, or "kinda independent."

E.g. coin flips: $H = X_1 + \ldots + X_n$.

$$\Pr \left[ H \geq \frac{n}{2} \right] \geq \frac{1}{\varepsilon^2}$$
This will occupy us for a while.

\[ \Pr \left[ H \geq \frac{n}{2} + \frac{\varepsilon}{2} \cdot 10\sqrt{\ln n} \right] \leq \begin{cases} \varepsilon \Pr \left[ \frac{H}{n} \geq \frac{1}{2} + \frac{\varepsilon}{2} \cdot \sqrt{\frac{10}{n}} \right] & \text{(Markov)} \\ \varepsilon \cdot \Pr \left[ \frac{H}{n} \geq \frac{1}{2} + \frac{\varepsilon}{2} \cdot \sqrt{\frac{10}{n}} \right] & \text{(Chebyshev)} \end{cases} \]

Chebyshev doesn't need independence, just "pairwise indep," much common in practice.

\[ \operatorname{Var}[X_1 + \ldots + X_n] = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] - \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]^2 = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j] - \sum_{i=1}^{n} \mathbb{E}[X_i] \mathbb{E}[X_i] \quad \text{if pairwise indep} \]

= What \( \operatorname{Var}[X_1 + \ldots + X_n] \) would be, if fully indep.

Back to coins. How can we beat \( \frac{1}{2} \) coin?

Switch to \( X = X_1 + \ldots + X_n \), \( X_i \) indep \( \{1, -1\} \) w.p \( \frac{1}{2} \)

\[ \Pr \left[ |X| \geq 10\sqrt{\ln n} \cdot \sqrt{n} \right] \]

\[ \Pr \left[ \frac{|X|}{\sqrt{n}} \geq \frac{10\sqrt{\ln n}}{\sqrt{n}} \right] \]

Chebyshev: \( \frac{\varepsilon^2}{2} \)

"4th moment method":

[works w/ 4-wise indep rvs]

\[ \Pr\left[ X \geq C \right] = \Pr\left[ X^4 \geq C^4 \right] \leq \mathbb{E}[X^4] \cdot \frac{C^4}{\varepsilon^4} \]
$E[X^4] = \sum X_i^4 + X_i X_j$ terms $+ X_i X_j X_k$ terms $+ X_i X_j X_k X_l$ terms

In expect: $n E[X_i^4] = n \sum_i X_i^4 \leq C_4 \cdot (\frac{n}{2})^4 = \frac{3n^4}{16} \leq \frac{3}{(\log n)^{\frac{1}{2}}}$

$= 3n^2 - 2n \leq 3n^2$

$\Pr[|X| > Cn] \leq \frac{3n^2}{C^4 n^8} = \frac{3}{C^4 n^6}$

Keep going: $E[X^8] \leq C_8 \cdot n^8$

$\Pr[|X| > Cn] \leq C_8 \cdot n^{10}$

"Chernoff method":

Markov: $\Pr(X > t \mu) = \Pr(e^{tX} > e^{t\mu})$ for any $t > 0$

$\Pr(e^{tX}) \leq E[e^{tX}] \leq \frac{e^{t\mu}}{e^{t \mu}}$

$a$-number $\Rightarrow e^{\lambda u}$

easy to compute if $X = X_1 + \cdots + X_n$ independent

Plan: Compute bound it, then optimize $\lambda$. 
\[ E[e^{AX}] = E[\exp(AX_1 + \cdots + AX_n)] \\
= E[\exp(AX_1) \cdots \exp(AX_n)] \\
= E[\exp(AX_1)] \cdots E[\exp(AX_n)] \quad \text{(by independence)} \\
= [\frac{1}{2} e^x + \frac{1}{2} e^{-x}]^n \\
= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)^n \\
\leq \left(e^{\frac{x^2}{2}}\right)^n \\
= e^{\frac{x^2}{2} n} \\
So \quad \Pr[X \leq u] \leq e^{\frac{\frac{1}{2} n}{n} u^n} = e^{\exp(\frac{n}{2} u^2 - u)} \\
\text{minimize over } l \\
\overset{\text{choose } l = \frac{u}{n}}{=} e^{-\frac{u^2}{2n}} \\
So \text{ for } u = 10 \sqrt{\ln n}, \quad \leq \frac{1}{n^{50}} \quad \text{??} \]

\[ \text{Specific to sum of iid 50/50's} \]

Next: treat \( X_i = \{1 \text{ w.p. } p; \quad 0 \text{ w.p. } 1-p\} \). Must bound \( E[\exp(AX_i)] \)

Moderately annoying cases...

\[ \Rightarrow \]
Let $X = X_1 + \ldots + X_n$, $X_i$'s indep
Let $\mu = E[X] = \sum E[X_i]$

"Hoeffding" Say $a_i \leq X_i \leq b_i$, always.
Then

$$
\frac{\Pr[X > \mu + \epsilon]}{\Pr[X < \mu - \epsilon]} \leq \exp\left(-\frac{2 \epsilon^2}{\sum (a_i - b_i)^2}\right)
$$

[Uses w/ our result for $X = 1$]

- Better if $\mu$ is surprisingly small

"Chernoff" Repeating, say $0 \leq X_i \leq 1$.

For $\epsilon > 0$,

$$
\Pr[X \leq (1-\epsilon)\mu] \leq \exp\left(-\frac{\epsilon^2}{2} \mu\right)
$$

$$
\Pr[X \geq (1+\epsilon)\mu] \leq \exp\left(-\frac{\epsilon^2}{2+\epsilon} \mu\right)
$$

- $\leq \exp\left(-\frac{\epsilon^2}{2} \mu\right)$ if $\epsilon \leq 1$

Fact: If $\mu_L \leq \mu \leq \mu_H$

Bonus results 1. Neg. Dep. r.v.'s. - do first.

Def. $X_1, \ldots, X_n$ are negatively associated (NA) if:

$$
E[f(X_i; i \in A)g(X_i; i \in B)] \leq E[f(X_i; i \in A)]E[g(X_i; i \in B)]
$$

for nondec $f,g$, all disjoint $A, B \subseteq [n]$

ex: $E[e^{\lambda X_1 + \ldots + \lambda X_n}] \leq E[e^{\lambda X_1}] \cdots E[e^{\lambda X_n}]$

$\Rightarrow$ Hoeffding/Chernoff holds
Facts: \( \text{indep} \implies \text{N.A.} \)

- closed under subsets, \( \text{indep} \)
- applying nondecr. fons to disjoint subset; e.g.: \( X_1 + X_2 + X_3, X_4 + \ldots + X_6, X_{11} + X_{12}, \ldots \)

eg: [see Jouy-Dev-Prescher 88]

1. Put \( n \) pts randomly on unit circle, \( X_1, \ldots, X_n \) = arc lengths
2. Throw \( N \) balls into \( n \) bins \( \{ \text{diff prob for each bin} \} \)
   \( X_1, \ldots, X_n \) occcupancies
3. Let \( x_1, \ldots, x_n \) be fixed \& \( X_1, \ldots, X_n \) a random permutation; \( \implies \text{sampling from finite population} \)
   \( \text{without replacement} \)

All \( \text{N.A.} \implies \text{Intuitive meaning: knowing that some values are "large" makes it more likely other values are "small"} \)

Martingales: Let \( X_1, \ldots, X_n \) be \( \text{indep} \) r.v.s. \( \text{discrete} \) \( \text{not nec indep} \)

- Let \( f: \mathbb{R}^n \to \mathbb{R} \)
- Let \( Y_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_{i-1}] \), so \( Y_n = \mathbb{E}[f(X_1, \ldots, X_n)] = (\text{constant n.v.}) \)

Then \( \mu = Y_0, Y_1, \ldots, Y_n \) is a "martingale" w.r.t. \( X_1, \ldots, X_n \)

"Method of Focal Diff" (Azuma):

- Suppose \( X_1, \ldots, X_n \) \( \text{indep} \), \( f \) sats \( f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n) \leq c_i \)
- Then \( \Pr[f(X) \geq \mu + t] \leq \exp \left( -\frac{2t^2}{\sum c_i^2} \right) \)
- \( \Pr[f(X) \leq \mu - t] \leq \exp \left( -\frac{2t^2}{\sum c_i^2} \right) \)