LECTURE 4

CONCENTRATION INEQUALITIES.

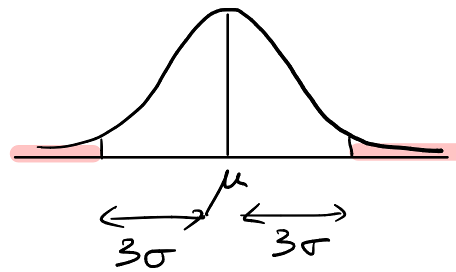
- "$X \approx \mathbb{E}X$ with high probability"

  closer to 1 than you think.

- Ex: normal distribution. $X \sim N(\mu, \sigma^2)$ satisfies

  $|X-\mu| \le 3\sigma$ with prob. 0.9973

  (see 68-95-99.7 rule)



A general tail bound:

Prop (Gaussian tails) $g \sim N(0,1)$ satisfies

$$\mathbb{P}\{g \ge t\} \le \frac{1}{t\sqrt{2\pi}} e^{-t^2/2} \qquad \forall t \ge 0$$

decays fast in t.

Proof (Recall: $\mathbb{P}\{X \in A\} = \int_A p(x)\,dx$ density of X)

$$\mathbb{P}\{g \ge t\} = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

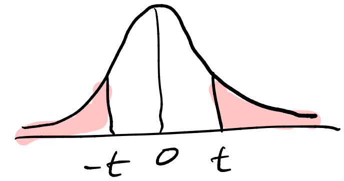
Not computable analytically.

To estimate, change variables $x = t + y \Rightarrow \frac{x^2}{2} = \frac{t^2}{2} + ty + \frac{y^2}{2}$

$$\mathbb{P}\{g \ge t\} = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} e^{-ty} \underbrace{e^{-y^2/2}}_{\le 1} dy$$

$$\le \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \underbrace{\int_0^\infty e^{-ty}\,dy}_{= 1/t \text{ (DIY)}}.$$

QED.



- By symmetry,

$$\mathbb{P}\{|g| \geq t\} = 2 \cdot \mathbb{P}\{g \geq t\} \leq \frac{1}{t}\sqrt{\frac{2}{\pi}}\, e^{-t^2/2} \qquad (*)$$

- More generally, if $X \sim N(\mu, \sigma^2) \Rightarrow X = \mu + \sigma g$

$$\Rightarrow \mathbb{P}\{|X-\mu| \geq t\sigma\} = \mathbb{P}\{|g| \geq t\} \leq e^{-t^2/2} \qquad \forall\, t \geq 1.$$

- <u>Ex</u>: $t=3$: $\quad \mathbb{P}\{|X-\mu| \leq 3\sigma\} \geq 1 - \frac{1}{3}\sqrt{\frac{2}{\pi}}\, e^{-3^2/2} \geq 0.9970$

almost as good as the exact number 0.9973 on p. 3

- <u>Remark</u>: a simpler form of (*) often suffices.
  Since $\sqrt{2/\pi} \leq 1$, we have:

$$\boxed{\mathbb{P}\{|g| \geq t\} \leq e^{-t^2/2} \qquad \forall\, t \geq 1}$$

"Gaussian tail bound"

- For general distributions?

- **Prop** (Markov's inequality) $\forall$ non-negative r.v. X:

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}X}{t} \qquad \forall t > 0.$$

Proof $\forall x \in \mathbb{R}$ can be decomposed as

$$x = x \cdot \mathbb{1}_{\{x \geq t\}} + x \cdot \mathbb{1}_{\{x < t\}}$$

($\mathbb{1}_A$ is the indicator $\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if not} \end{cases}$)

Apply this for X and take expectations on both sides:

$$\mathbb{E}X = \mathbb{E}[\underbrace{X \mathbb{1}_{\{X \geq t\}}}_{\geq\, t \mathbb{1}_{\{X \geq t\}}}] + \underbrace{\mathbb{E}[X \mathbb{1}_{\{X < t\}}]}_{\geq\, 0}$$

$$\geq t\, \mathbb{E}\, \mathbb{1}_{\{X \geq t\}} = t \cdot \mathbb{P}\{X \geq t\}.$$

Divide both sides by $t \Rightarrow$ QED

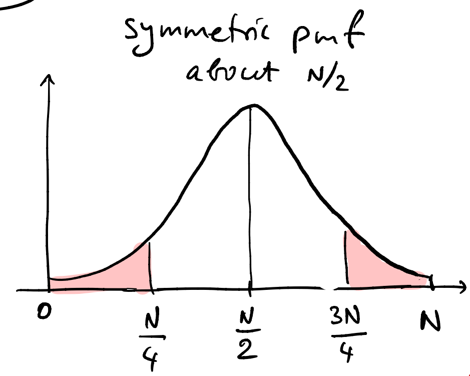
- **Prop** (Chebyshev's inequality) $\forall$ r.v. X with mean μ, variance σ^2:

$$\mathbb{P}\{|X-\mu| \geq t\} \leq \frac{\sigma^2}{t^2} \qquad \forall t > 0$$

Proof $\mathbb{P}\{|X-\mu| \geq t\} = \mathbb{P}\{(X-\mu)^2 \geq t^2\} \leq \frac{\mathbb{E}(X-\mu)^2}{t^2}$ (Markov for $(X-\mu)^2$)

$$= \sigma^2/t^2. \qquad \text{QED}$$

- **QUESTION** Toss a fair coin N times. $P\{\text{at least } \frac{3}{4}N \text{ heads}\} = ?$



- **Solution 1**, based on Chebyshev:

$$S_N = \#\text{heads} \sim \text{Binom}(N, \tfrac{1}{2})$$

$$\mathbb{E} S_N = \frac{N}{2}, \quad \text{Var}(S_N) = \frac{N}{4}$$

Chebyshev $\Rightarrow$

$$P\{S_N \geq \tfrac{3}{4}N\} \underset{\substack{\uparrow \\ \text{symmetry}}}{=} \frac{1}{2} P\left\{ \left|S_N - \frac{N}{2}\right| \geq \frac{N}{4} \right\} \leq \frac{1}{2} \cdot \frac{N/4}{(N/4)^2} = \boxed{\frac{2}{N}} = 0.025 \quad \text{if } N = 80$$

- **Solution 2**, based on CLT:

$$\frac{S_N - \mathbb{E}S_N}{\sqrt{\text{Var}(S_N)}} \to N(0,1) \quad \text{as } N \to \infty$$

$$\Rightarrow P\{S_N \geq \tfrac{3}{4}N\} = P\left\{ \frac{S_N - N/2}{\sqrt{N/4}} \geq \sqrt{\frac{N}{4}} \right\}$$

$$\approx P\{g \geq \underbrace{\sqrt{\tfrac{N}{4}}}_{=:t}\} \quad \text{where } g \sim N(0,1)$$

$$\leq e^{-t^2/2} = \boxed{e^{-N/8}} \quad \text{(Gaussian tail, p. 2)}$$

≈ 0.000045 if $N = 80$

MUCH BETTER!

- But Sol. 2 has a gap: the error term in CLT.

What is it? $1/\sqrt{N}$:)

• QUANTITATIVE CLT:

THM (Berry – Esseen) Let X_i be iid r.v.'s with mean 0, var. 1,

$$\Rightarrow \left| P\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \geq t \right\} - P\{g \geq t\} \right| \leq \frac{\rho}{\sqrt{N}}$$

where $g \sim N(0,1)$ and $\rho = E|X_1|^3$.

• THAT'S SAD : taking this error into account in Sol. 2 yields probability

$$\underbrace{\frac{1}{\sqrt{N}}}_{\uparrow \text{ BIG}} + e^{-N/8}$$

Not better than Sol. 1 based on Chebyshev.

• Can we improve $\frac{1}{\sqrt{N}}$ in CLT ?

NO : $P\{\text{exactly } \frac{N}{2} \text{ heads}\} = P\{S_N = \frac{N}{2}\} = 2^{-N}\binom{N}{N/2} \asymp \frac{1}{\sqrt{N}}$

while $P\{g = 0\} = 0$

$\Downarrow$

error $\frac{1}{\sqrt{N}}$ is unavoidable.

WHAT SHOULD WE DO ?