LECTURE 6

THM (Hoeffding's inequality) Let $X_1, \dots, X_N$ be iid symmetric Bernoulli r.v.'s i.e. $P\{X_i = 1\} = P\{X_i = -1\} = 1/2$. Then

$$P\Big\{ \underbrace{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \geq t}_{} \Big\} \leq \exp\left(-\frac{t^2}{2}\right) \quad \forall t \geq 0.$$

Proof: "The MGF method" (moment generating function).

- Let $\lambda \geq 0$ be a parameter TBD later ("to be determined").
- Multiply both sides of the inequality by $\lambda\sqrt{N}$ and exponentiate:

$$P\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \geq t \right\} = P\left\{ \exp\left(\lambda \sum_{i=1}^{N} X_i\right) \geq \exp(t\lambda\sqrt{N}) \right\}$$

$$\overset{\text{Markov's ineq.}}{\leq} \frac{\mathbb{E}\exp\left(\lambda \sum_{i=1}^{N} X_i\right)}{\exp(\lambda t \sqrt{N})} = e^{-t\lambda\sqrt{N}}\, \mathbb{E} \prod_{i=1}^{N} e^{\lambda X_i}$$

(independence)

$$= e^{-t\lambda\sqrt{N}} \prod_{i=1}^{N} \underbrace{\mathbb{E} e^{\lambda X_i}}_{}$$

all terms are the same (identical distr.)

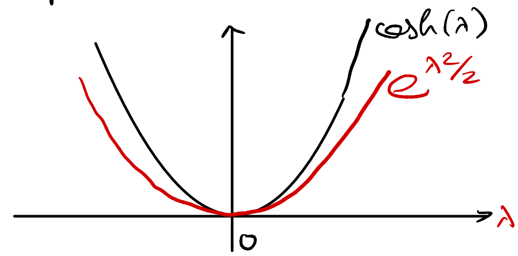
$$= e^{-t\lambda\sqrt{N}} \left(\mathbb{E} e^{\lambda X}\right)^N \leq \quad \text{where } X \sim \text{sym Bernoulli}$$

- Since X takes values 1 and -1 with probabilities $\frac{1}{2}$ each,

$$\mathbb{E} e^{\lambda X} = \frac{1}{2} e^{\lambda} + \frac{1}{2} e^{-\lambda}$$

"Moment generating function" (MGF)

$$= \cosh(\lambda) \underset{\text{Taylor}}{\overset{\text{DIY}}{\leq}} e^{\lambda^2/2}$$



$$\leq e^{-t\lambda\sqrt{N}} \left(e^{\lambda^2/2}\right)^N = \exp\Big(-t\underbrace{\lambda\sqrt{N}}_{=\mu} + \frac{\lambda^2 N}{2}\Big) = \exp\left(-t\mu + \frac{\mu^2}{2}\right).$$

- Minimize in μ $\Rightarrow$ for $t = \mu$:

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

Ex
(a) In N flips of a fair coin, $\mathbb{P}\{\text{at least } \frac{3}{4}N \text{ heads}\} \le \exp(-\frac{N}{8})$
(b) In N flips of biased coins that come up heads with prob $\le 1/4$,
$\mathbb{P}\{\text{at least } \frac{N}{2} \text{ heads}\} \le \exp(-cN)$.
(DIY)

APPLICATION: MEAN ESTIMATION

- **Problem:** $\boxed{\text{Estimate the mean } \mu \text{ of a distribution from a sample } X_1, \dots, X_n, \text{ which is drawn independently from this distribution.}}$

- Classical estimator: the sample mean $\boxed{\hat{\mu} := \frac{1}{N} \sum_{i=1}^{N} X_i}$

- $\mathbb{E}\hat{\mu} = \mu$, i.e. unbiased. Mean-squared error:

$$\mathbb{E}(\hat{\mu}-\mu)^2 = \mathrm{Var}(\hat{\mu}) = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{Var}(X_i) = \frac{\sigma^2}{N}.$$

$\Rightarrow$ the error is $\boxed{\frac{\sigma}{\sqrt{N}}}$ on average

- Confidence interval? Chebyshev's inequality:

$$\mathbb{P}\left\{ |\hat{\mu}-\mu| \ge \frac{t\sigma}{\sqrt{N}} \right\} \le \frac{\sigma^2/N}{(t\sigma/\sqrt{N})^2} = \boxed{\frac{1}{t^2}}$$

Meh.

- For Gaussian distribution, we get much better confidence:

if $X_i \sim N(\mu, \sigma^2) \Rightarrow \hat{\mu} \sim N(\mu, \frac{\sigma^2}{N})$.

Gaussian tail $\Rightarrow$ $\mathbb{P}\left\{ |\hat{\mu}-\mu| \ge \frac{t\sigma}{\sqrt{N}} \right\} \le \boxed{\exp(-t^2/2)}$

- Can we get the same for general distr.?

SURPRISINGLY, YES!

Ex: prove same result for uniform distr.

$\underline{\underline{\text{THM}}}$ $\exists$ estimator $\tilde{\mu} = \tilde{\mu}(X_1, \dots, X_N)$ that satisfies

$$\mathbb{P}\left\{ |\tilde{\mu} - \mu| \geq \frac{t\sigma}{\sqrt{N}} \right\} \leq 2e^{-ct^2} \qquad \forall t \geq 0$$

if X_i's are drawn from distr. with mean μ, var. σ.

- With only finite variance assumption, Gaussian power is surprising!

Proof $\tilde{\mu}$ = "Median-of-Means" (MoM):

- Partition the sample into K blocks of equal size M $\boxed{N = MK}$

$$\underbrace{X_1\, X_2 \cdots X_M}_{B_1},\ \underbrace{X_{M+1} \cdots X_{2M}}_{B_2},\ \cdots\cdots\ \underbrace{\cdots\cdots X_N}_{B_K}$$

- Take means of each block; then take the median.

(*) $\hat{\mu}_j := \frac{1}{M} \sum_{i \in B_j} X_i$; $\tilde{\mu} := \mathrm{Med}(\hat{\mu}_1, \dots, \hat{\mu}_K)$

(i.e. at least $\frac{K}{2}$ of $\hat{\mu}_j$ are $\leq \tilde{\mu}$ and at least $\frac{K}{2}$ are $\geq \tilde{\mu}$)

- Accuracy of each $\hat{\mu}_j$? Mean μ, variance σ^2/M $\Rightarrow$ by Chebyshev

$$\mathbb{P}\left\{ \hat{\mu}_j \geq \mu + \frac{t\sigma}{\sqrt{N}} \right\} \leq \frac{\sigma^2/M}{(t\sigma/\sqrt{N})^2} = \frac{N}{t^2 M} = \frac{K}{t^2} = \frac{1}{4} \text{ if we set } \boxed{K := t^2/4}$$

- By def. of the mean,

$$\mathbb{P}\left\{ \tilde{\mu} \geq \mu + \frac{t\sigma}{\sqrt{N}} \right\} \leq \mathbb{P}\left\{ \text{at least } \frac{K}{2} \text{ of } \hat{\mu}_j\text{'s are} \geq \frac{t\sigma}{\sqrt{N}} \right\}$$

$$= \mathbb{P}\left\{ \text{at least } \frac{K}{2} \text{ of } \frac{1}{4}\text{-biased coins come up heads} \right\}$$

$$\leq \exp(-cK) = \exp(-ct^2/4).$$

Hoeffding's ineq (Ex (b) p. 2)

QED.