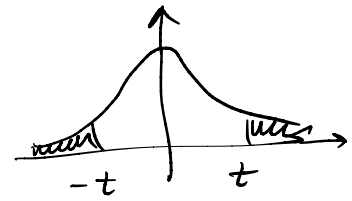
LECTURE 7

More general versions of Hoeffding's inequality:



① Two sided Bound:

$$P\left\{\left|\frac{1}{\sqrt{N}}\sum_1^N X_i\right| \geq t\right\} = \underbrace{P\left\{\frac{1}{\sqrt{N}}\sum_1^N X_i > t\right\}}_{\text{Hoeffding} \leq e^{-t^2/2}} + \underbrace{P\left\{\frac{1}{\sqrt{N}}\sum_1^N X_i < -t\right\}}_{\substack{\| \\ P\left\{\frac{1}{\sqrt{N}}\sum_1^N \underbrace{(-X_i)}_{\text{Sym Bernoulli, too}} > t\right\} \\ \text{Hoeffding} \leq e^{-t^2/2}}}$$

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

② THM (Hoeffding for Bounded r.v's) Let $X_1, \dots, X_n$ be independent r.v's such that $X_i \in [a_i, b_i]$ $\forall i$. Then $S_N = \sum_{i=1}^N X_i$ satisfies

$$P\{S_N - ES_N \geq t\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^N (b_i - a_i)^2}\right)$$

- Proof: the same MGF method.

- Remark This contains Hoeffding we proved earlier for $\sum_{i=1}^N \frac{Z_i}{\sqrt{N}}$ where $Z_i \sim$ Sym Bernoulli. Indeed, here $a_i = -1/\sqrt{N}$, $b_i = 1/\sqrt{N}$ $\Rightarrow \sum_{i=1}^N (b_i - a_i)^2 = 4$

$$\Rightarrow RHS = \exp\left(-\frac{2t^2}{4}\right) = \exp(-t^2/2).$$ ☺

- WEAK SPOT OF HOEFFDING'S INEQUALITY:

it only cares about extreme spread of X_i $(a_i \mapsto b_i)$ rather than average spread (like standard deviation)

- For example: if $X_i \sim Ber(p)$ and $p \to 0$, concentration should get better $(X_i \to \text{deterministic})$ but H.I. does not get better $(a_i = 0,\ b_i = 1\ \forall p)$.

- What does $\sum_{i=1}^{N} X_i$ look like in this case, say, if $p = \frac{\mu}{N}$ and $N \to \infty$? CLT is NOT applicable.

ANS: Poisson

- Poisson Limit Thm Let $X_1^{(N)}, \dots, X_2^{(N)} \sim Ber\left(\frac{\mu}{N}\right)$ be indep. $\forall N$.

Then $\sum_{i=1}^{N} X_i^{(N)} \longrightarrow Pois(\mu)$ in distribution

RECALL

$X \sim Pois(\mu)$ if takes values $0, 1, 2, \dots$ with prob's

$$\boxed{\mathbb{P}\{X = k\} = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \dots}$$

$\mathbb{E}X = Var(X) = \mu$.

- Ex (Poisson Tails) $\mathbb{P}\{X \geq t\} = e^{-\mu} \sum_{k \geq t} \frac{\mu^k}{k!} \underset{\text{Stirling}}{\leq} \boxed{e^{-\mu}\left(\frac{e\mu}{t}\right)^t}$ $\forall t \geq 0$

- Hence we should expect of $S_N = \sum_{i=1}^{N} X_i^{(N)}$ the same behavior, i.e. that

$$\mathbb{P}\{S_N \geq t\} \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t \qquad (*)$$

- Does NOT follow from Poisson Limit Thm, due to the approximation error. Nevertheless, $(*)$ is true:

THM (Chernoff's inequality) Let $X_i \sim Ber(p_i)$ be iid. r.v's.

Then $S_N = \sum_{i=1}^{N} X_i$ has mean $\mathbb{E}S_N = \sum_{i=1}^{N} p_i =: \mu$, and

$$\mathbb{P}\{S_N \geq t\} \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t \qquad \forall\, t \geq \mu.$$

Remark In particular, C.I. holds $\forall$ Binomial distr. with mean μ.

Proof use the same MGF method as the proof of Hoeffding:

- $\mathbb{P}\{S_N \geq t\} = \mathbb{P}\{e^{\lambda S_N} \geq e^{\lambda t}\} \leq e^{-\lambda t}\mathbb{E}e^{\lambda S} = e^{-\lambda t}\,\mathbb{E}\prod_{i=1}^{N} e^{\lambda X_i}$

$$\leq e^{-\lambda t}\prod_{i=1}^{N}\mathbb{E}e^{\lambda X_i} \quad (\leq)$$

- Each $e^{\lambda X_i}$ takes value $e^{\lambda \cdot 1}$ with prob. p_i, value $e^{\lambda \cdot 0} = 1$ with prob $1-p_i$

$$\Rightarrow \mathbb{E}e^{\lambda X_i} = e^{\lambda}p_i + (1-p_i) = 1 + (e^{\lambda}-1)p_i \leq \exp(e^{\lambda}-1)p_i$$

use ineq. $1 + x \leq e^{x}$

- $(\leq)\; e^{-\lambda t}\exp\Big[(e^{\lambda}-1)\underbrace{\sum_{i=1}^{n} p_i}_{=\mu}\Big] = \exp\big[-\lambda t + (e^{\lambda}-1)\mu\big] \quad (\leq)$

- Minimize in λ

$\Rightarrow$ for $\lambda = \ln(t/\mu)$ $\qquad (\leq)\; e^{-\mu}\left(\frac{e\mu}{t}\right)^t.$ $\qquad$ QED.

(note that $\lambda \geq 0$ by assumption $t \geq \mu$)