LECTURE 7

Recall

(stated in lec. 5, proved in HW 4)

THM (General Hoeffding) 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

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

- WEAKNESS: Hoeffding's Inequality only cares about extreme spread of X_i (a_i — b_i) rather than average spread (like standard deviation)

- For example: if $X_i \sim \mathrm{Ber}(p)$ and $p \to 0$, concentration should get better ($X_i \to$ 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 $X_1^{(N)}, \dots, X_2^{(N)} \sim \mathrm{Ber}\left(\frac{\mu}{N}\right)$ be indep. $\forall N$. Then $\sum_{i=1}^{N} X_i^{(N)} \longrightarrow \mathrm{Pois}(\mu)$ in distribution

RECALL

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

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

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

- Poisson tails, heuristically: $\forall t \geq \mu$:

$$\mathbb{P}\{X \geq t\} = e^{-\mu} \sum_{k=t}^{\infty} \frac{\mu^k}{k!} \sim e^{-\mu} \frac{\mu^t}{t!} \sim e^{-\mu} \left(\frac{e\mu}{t}\right)^t \quad (*)$$

decays exponentially fast $t! \sim (t/e)^t$ (Stirling)

HW 5 will make this argument rigorous

- Hence we should expect of $S_N = \sum_{i=1}^{N} X_i^{(N)}$ a similar behavior:

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

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

THM (Chernoff's inequality) Let $X_i \sim \mathrm{Ber}(p_i)$ 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

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

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

Proof use the same MGF method as in 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 $1-p_i$

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

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

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

- Minimize in λ
  $\Rightarrow$ for $\lambda = \ln(t/\mu)$ $\quad \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t$. QED.

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

REMARKS

① Chernoff is ~optimal: if $S_N \sim \mathrm{Binom}(N,\mu)$ then

"Reverse Chernoff ineq": $\boxed{\mathbb{P}\{S_N \geq t\} \geq (\mu/t)^t}$ $\forall\, t \geq \mu$ (HW)

② Lower tails are similar: $\boxed{\mathbb{P}\{S_N \leq t\} \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t \quad \forall\, 0<t<\mu}$

(HW) ← hint: $= \mathbb{P}\{-S_N \geq -t\}$, proceed similarly.

② Large deviations: when t is large, the "Poisson tail" $\sim t^{-t} = e^{-t\log t}$ is heavier than Gaussian $e^{-t^2/2}$.

③ Small deviations: when $t \approx \mu$, say $t = (1+\delta)\mu$, (δ ← small)

$$\cdot\ e^{-\mu}\left(\frac{e\mu}{t}\right)^t = e^{-\mu}\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} = e^{\delta\mu}\left(\frac{1}{1+\delta}\right)^{(1+\delta)\mu}$$

$$= \exp\Big[-\mu\,\underbrace{\big((1+\delta)\log(1+\delta) - \delta\big)}\Big] \leq \exp(-\delta^2\mu/3)$$

GAUSSIAN TAIL! ☺

Taylor: $\frac{\delta^2}{2} - \frac{\delta^3}{2\cdot 3} + \frac{\delta^4}{3\cdot 4} - \frac{\delta^5}{4\cdot 5} + \dots \geq \frac{\delta^2}{3}$ if $\delta \in [0,1]$

(move $\delta^3/3$ to the L.H.S ⇒ series with terms → 0 & alternating signs)

• Combine upper & lower tails

⇓

Cor (Chernoff's ineq: small deviations)

$$\boxed{\mathbb{P}\{|S_N - \mu| \geq \delta\mu\} \leq 2\exp(-\delta^2\mu/3) \quad \forall\, \delta \in [0,1]}$$

• Example: $S_N = \sum_{i=1}^{N} X_i$ where $X_i \sim Ber(p)$ $\Rightarrow$ $S_N \sim Binom(N, p)$

$\mu = \mathbb{E} S_N = Np$; $\sigma^2 = Var(S_N) = Np(1-p) \approx Np = \mu$ $\Rightarrow \sigma \approx \sqrt{\mu}$.

Cor $\Rightarrow$

> $\forall$ Binomial r.v. S_N with mean μ:
> $$\mathbb{P}\{ |S_N - \mu| \geq \underbrace{t\sqrt{\mu}}_{= \delta\mu} \} \leq \exp(-t^2/3) \quad \forall\, 0 \leq t \leq \sqrt{\mu}$$

$\delta\mu \Rightarrow \delta = t/\sqrt{\mu}$

Gaussian tail ☺

SUMMARY of $S_N \sim$ Binomial with mean μ: