Lecture 3:

If $X_1, X_2, \dots X_n$ i.i.d. with $\mathbb{E}[X_i] = \mu$ and variance σ^2

WLLN $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mu$ as $n \to \infty$ in probability

$\forall \varepsilon > 0 \quad P(|\bar{X}_n - \mu| \geq \varepsilon) \to 0$ as $n \to \infty$

CLT $W_n = \dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow[n\to\infty]{} N(0,1)$. in distribution. $Var(\bar{X}_n) = \frac{1}{n}\sigma^2$

$\forall w \in \mathbb{R}. \quad F_{W_n}(w) = P(W_n \leq w) \longrightarrow P(\underset{\substack{\uparrow \\ N(0,1)}}{W} \leq w)$

Question: Toss a fair coin n times. What is the probability that we get at least $\frac{3n}{4}$ heads?

$$X_i = \begin{cases} 1 & \text{if } i\text{'th flip is H} \\ 0 & \text{if } i\text{'th flip is T} \end{cases}$$

$X_1, X_2, X_3, \dots X_n \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ independent

$$P\left(\sum_{i=1}^{n} X_i \geq \frac{3n}{4}\right) = P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \frac{3}{4}\right)$$

$$= P\left(\bar{X}_n - \frac{1}{2} \geq \frac{1}{4}\right)$$

$$\leq P\left(\left|\bar{X}_n - \frac{1}{2}\right| \geq \frac{1}{4}\right)$$

$Var(\bar{X}_n) = \frac{1}{n}\sigma^2 = \frac{1}{4n}$. $\qquad \leq \dfrac{Var(\bar{X}_n)}{(1/4)^2} = \dfrac{4}{n}$.

CLT: $P\left(\bar{X}_n - \frac{1}{2} \geq \frac{1}{4}\right) = P\left(\underbrace{\frac{\bar{X}_n - 1/2}{\sqrt{1/4n}}}_{W_n} \geq \frac{1/4}{\sqrt{1/4n}}\right)$

$$\approx P\left(\underset{\downarrow\ \mathcal{N}(0,1)}{W} \geq \sqrt{\frac{n}{4}}\right)$$

$$Q(\alpha) = \int_\alpha^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \qquad = \int_{\sqrt{n/4}}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$\downarrow$

Q-function

$$= Q\left(\sqrt{\frac{n}{4}}\right) \leq \frac{1}{\sqrt{2\pi}} e^{-n/8}$$

Small technical result: $Q(\alpha) \leq \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2} \quad \forall \alpha > 0$

if $\alpha > 1$, $Q(\alpha) \leq \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2} \quad \forall \alpha > 0$

E.g. $n = 100 \qquad \frac{4}{n} = 0.04 \qquad Q\left(\sqrt{\frac{n}{4}}\right) \approx 2.8 \cdot 10^{-7}$

Concentration inequalities provide exponentially decaying tail bounds on $\bar{X}_n$, or equivalently $S_n = X_1 + X_2 + \dots + X_n$ and more generally on $f(X_1, X_2, \dots X_n)$ when f is a "nice" function

Hoeffding's Inequality: Let $X_1, X_2, \dots X_n$ be a sequence of i.i.d r.v's, $\mathbb{E}[X_i] = \mu$, and $a \leq X_i \leq b$. Then for any $t \geq 0$, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$

$$P\left(|\bar{X}_n - \mu| \geq t\right) \leq 2e^{-2nt^2/(b-a)^2}$$

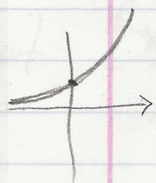
Apply to the previous example:

$$P\left(\left|\overline{X}_n - \frac{1}{2}\right| \geq \frac{1}{4}\right) \leq 2e^{-2n\left(\frac{1}{4}\right)^2} = 2e^{-\frac{n}{8}}$$

same decay rate as CLT!

Proof: $P(\overline{X}_n - \mu \geq t) = P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq t\right)$

for any $\lambda > 0$: $= P\left(e^{\lambda\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)} \geq e^{\lambda t}\right)$



$e^{\lambda x}$ is monotonically increasing when $\lambda \geq 0$.

$$\leq \frac{\mathbb{E}\left[e^{\lambda\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)}\right]}{e^{\lambda t}}$$

Markov's Inequality

$e^{a+b} = e^a \cdot e^b$.

$$= e^{-\lambda t}\,\mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda \frac{1}{n}(X_i - \mu)}\right]$$

$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$

if X and Y are independent.

$$= e^{-\lambda t}\prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda \frac{1}{n}(X_i - \mu)}\right]$$

$$= e^{-\lambda t}\left(\mathbb{E}\left[e^{\lambda \frac{1}{n}(X_i - \mu)}\right]\right)^n$$

moment generating function of $(X_i - \mu)$

Moment Generating Function of a r.v X

is defined as

$$M_X(s) = \mathbb{E}[e^{sX}]$$

$$= \begin{cases} \sum_x p_X(x)\,e^{sx} & \text{if } X \text{ is discrete} \\ \int e^{sx} f_X(x)\,dx & \text{if } X \text{ is contin} \end{cases}$$

Remark: There is one-to-one correspondence bet. the moment generating function for a r.v and its probability distribution fuction (pdf or pmf)

Moment generating property:

$$\frac{d}{ds} M_X(s) = \frac{d}{ds}\left(\sum_x e^{sx} p_X(x)\right)$$

$$= \sum_x x e^{sx} p_X(x)$$

$$M_X'(s)\Big|_{s=0} = \frac{d}{ds} M_X(s)\Big|_{s=0} = \sum_x x\, p_X(x) = \mathbb{E}[X] \quad \text{first mome}$$

$$M_X''(s)\Big|_{s=0} = \frac{d}{ds^2} M_X(s)\Big|_{s=0} = \frac{d}{ds}\left(\sum_x x e^{sx} p_X(x)\right)\Big|_{s=0}$$

$$= \sum_x x^2 e^{sx} p_X(x)\Big|_{s=0}$$

$$= \sum_x x^2 p_X(x) = \mathbb{E}[X^2] \quad \text{second moment of } X$$

n'th derivative of $M_X(s)$

$$M_X^n(s)\Big|_{s=0} = \mathbb{E}[X^n] \quad \text{n'th moment of } X$$

Property: $M_{X+Y}(s) = \mathbb{E}\left[e^{s(X+Y)}\right]$

$$= \mathbb{E}\left[e^{sX} e^{sY}\right]$$

if X and Y are independent

$$= \mathbb{E}\left[e^{sX}\right]\mathbb{E}\left[e^{sY}\right]$$

$$= M_X(s)\, M_Y(s)$$

If X and Y are independent, $M_{X+Y}(s) = M_X(s) M_Y(s)$