HDP Session 1

1. **Concentration**: Eg. X_i iid X, $\mathbb{E}X = \mu$, $Var(X) < \infty$

- $\bar{X}_n = \dfrac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p} \mathbb{E}X = \mathbb{E}\bar{X}_n = \mu$

Asymptotic

Frequentist Approach to Prob.

Q: At what rate?

- Markov's Ineq: $X \geq 0$ a.s., then $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}X}{a}$, $\mathbb{E}X < \infty$ $\forall a > 0$
- Chebychev's Ineq: $\mathbb{E}X^2 < \infty$, then

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{Var(X)}{a^2}. \quad \forall a > 0$$

☆ $\rightarrow$ ✓ $\mathbb{P}(|\bar{X}_n - \mathbb{E}\bar{X}_n| \geq \varepsilon) \leq \dfrac{Var(\bar{X}_n)}{\varepsilon^2} = \dfrac{Var(X)}{n\varepsilon^2} \sim O\left(\frac{1}{n}\right)$. $\rightarrow e^{-n\varepsilon^2}$

Non-asymptotic.

1) Can the rate of conv. be improved? — For some special classes of rvs.

Subgaussian Subexponential

- $\bar{X}_n \xrightarrow{p} \mathbb{E}\bar{X}_n$; $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

2) For what other function $f_n : \mathbb{R}^n \to \mathbb{R}$ satisfies

$$f_n(X_1, \dots, X_n) \approx \mathbb{E} f_n(X_1, \dots, X_n)$$

☆ Conc. ineq.

- $\mathbb{P}\left(|f_n(X_1, \dots, X_n) - \underbrace{\mathbb{E} f_n(X_1, \dots, X_n)}| \geq \varepsilon\right) \leq \dfrac{Var(f_n(X_1, \dots, X_n))}{\varepsilon^2}$

eg. $f_n(X_1, \dots, X_n) = \dfrac{X_1^2 + \dots + X_n^2}{n}$, X_i iid $N(0,1)$

$$f_n(X_1, \dots, X_n) \xrightarrow[p]{} \mathbb{E}(f_n(X_1, \dots, X_n))$$

$\longrightarrow$ Martingale methods

Indeed, showing conc. is sometimes the easy part.

$\mathbb{E} f(X_1, \dots, X_n) \longrightarrow$ harder. ✓

2. Supremum : • $\{X_t \mid t \in T\} \longrightarrow$ Some collection of rvs.

$$\mathbb{P}\left(\sup_{t\in T} X_t \leq x\right), \quad \mathbb{E} \sup_{t\in T} X_t$$

(*) Exc: $X_1, X_2, \ldots$ iid $N(0,1)$. Estimate $\mathbb{E}\left(\sup_{i=1}^{n} X_i\right)$.

Thm. (Fisher – Tippet - Gnedencko) X_i iid X with some good properties.
Then $\exists$ sequences $\{a_n\}, \{b_n\}$ s.t.

Extreme Value Theory.

$$\frac{\max_{i=1}^{n} X_i - a_n}{b_n} \xrightarrow{d} G$$

Gumbel distribution

• $F_G(x) = ce^{-e^{-x}}$.

$Y_n \xrightarrow{d} Y$

$F_{Y_n} \longrightarrow F_Y$
ptwise at all cont. pts of F_Y.

What is the non-asymptotic version of this ?

3) Universality : • CLT: X_i iid X, $\mathbb{E}X = \mu$, $\operatorname{Var} X = \sigma^2$

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} N(0,\sigma^2)$$

- Fourier Analytic Method
- Method of Moments $\rightsquigarrow$ Lindeberg Trick *

---

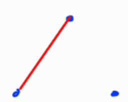
§. Moment Methods – FMM & SMM

Events $\overset{?}{\longleftrightarrow}$ Random Variable ; Ω

$$A \in \Omega \rightsquigarrow \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{o.w.} \end{cases}$$

Def$^{\underline{n}}$: (Erdös -Renyi RGs): $G(n,p)$ is a graph of n vertices $\{1,2,\ldots,n\} = [n]$,
s.t.

$\{i,j\} \in E(G(n,p))$ w.p. p indep of all other edges

$G(3,p)$ 

eg. Consider $G(n,p)$, and let $A_i = \{ i^{th} \text{ vertex is isolated} \}$.

Let $Y_{i,j} = \mathbb{1}_{\{\{i,j\} \in E(G(n,p))\}}$

$$\underline{\mathbb{P}(A_i)} = \mathbb{P}(Y_{i,1}, Y_{i,2}, \dots, Y_{i,n} = 0)$$

$$= \mathbb{P}(Y_{i,1} + \dots + Y_{i,n} = 0) \rightsquigarrow \text{Will be expressed in terms of a conc. ineq.}$$

$\mathbb{E}\, Y_{i,1} = p$

$$\leq \mathbb{P}\left(\left| \sum_{j \neq n} Y_{i,j} - (n-1)p \right| \geq (n-1)p \right)$$

$$\leq \dots$$

☆ First Moment method: $X \geq 0$ a.s. and $\mathbb{E} X < \infty$, X – integer valued

Then

$$\mathbb{E} X = \int_0^\infty \mathbb{P}(X \geq t)\, dt = \int_0^a \mathbb{P}(X \geq t)\, dt + \int_a^\infty \mathbb{P}(X \geq t)\, dt$$

$$\geq \int_0^a \mathbb{P}(X \geq t)\, dt \geq a \cdot \mathbb{P}(X \geq a)$$

$$\therefore \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E} X}{a} \qquad \text{[Markov's Ineq.]}$$

$$\boxed{\underbrace{\mathbb{P}(X > 0)} = \mathbb{P}(X \geq 1) \leq \mathbb{E} X} \longrightarrow \text{(I)}$$

☆ Second Moment Method: $X \geq 0$ a.s. and $\mathbb{E} X^2 < \infty$, then

$$(\mathbb{E} \underline{X})^2 = \left(\mathbb{E}(X \underline{\mathbb{1}}_{X>0}) \right)^2 \leq (\mathbb{E} X^2) \cdot \mathbb{E}(\mathbb{1}_{X>0})$$

$$= \mathbb{E} X^2 \cdot \mathbb{P}(X > 0)$$

$$\mathbb{P}(X > 0) \geq \frac{(\mathbb{E} X)^2}{\mathbb{E}(X^2)}$$

$\mathbb{N} = \{1, 2, \dots\}$
$\mathbb{N}_0 = \{0, 1, 2, \dots\}$

☆ $\longrightarrow$ $$\boxed{\frac{(\mathbb{E} X)^2}{\mathbb{E}(X^2)} \leq \mathbb{P}(X > 0) \leq \mathbb{E} X \quad \text{for } X - \mathbb{N}_0 \text{ valued r.v.}}$$

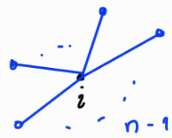
Eg 1.: $G(n,p)$: $\mathbb{P}(A_i)$, A_i = i^{th} vertex is isolated

$$\rightarrow \mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \mathbb{P}\big(G(n,p) \text{ has an isolated vertex}\big)$$

$$= \underline{\mathbb{P}(N>0)}$$

- N = # isolated vertices.

- $N = \sum_{i=1}^{n} N_i$, $N_i = \mathbb{1}[i^{th} \text{ vertex is isolated}]$



$$N_i \sim \text{Ber}\big((1-p)^{n-1}\big) \quad \forall i$$

$$\mathbb{E}N = \sum_{i=1}^{n} \mathbb{E}N_i = n(1-p)^{n-1} \quad ; \quad \boxed{(1-x) \leq e^{-x} \quad \forall x \in \mathbb{R}}$$

$$\leq e^{\log n} \cdot e^{-p(n-1)}$$

$$= \exp\big(\underbrace{\log n - (n-1)p}\big)$$

If $p \geq c\frac{\log n}{n}$ for some $c>1$, then $\mathbb{E}N \longrightarrow 0$

$$\mathbb{P}(N>0) \leq \mathbb{E}N \longrightarrow 0$$

$\therefore$ For $p > c\frac{\log n}{n}$ for some $c>0$, we have $\mathbb{P}(G(n,p) \text{ isol vert}) = o(1)$.

- 2^{nd} moment method: N_i's are not independent

$N_i = \mathbb{1}[i \text{ is isolated}]$

$$\mathbb{E}(N^2) = \mathbb{E}\Big(\Big(\sum_{i=1}^{n} N_i\Big)\Big(\sum_{j=1}^{n} N_j\Big)\Big)$$

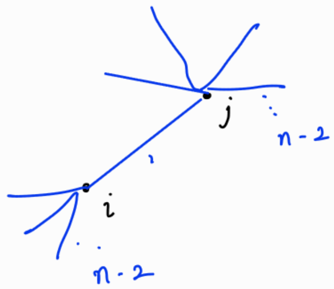
$$= \mathbb{E}\Big(\sum_{i=1}^{n} N_i^2\Big) + 2\,\mathbb{E}\Big(\sum_{i<j} N_i N_j\Big)$$

$$= \underbrace{\sum_{i=1}^{n} \mathbb{E}N_i}_{n(1-p)^{n-1}} + \underbrace{2\sum_{i<j} \mathbb{E}(N_i N_j)}_{n(n-1)(1-p)^{2n-3}}.$$

$$\mathbb{E}(N_iN_j) = \mathbb{E}\big(\mathbb{1}[i \text{ is isolated}] \cdot \mathbb{1}[j \text{ is isolated}]\big)$$

$$= \mathbb{E}\big(\mathbb{1}[i,j \text{ are isolated}]\big)$$

$$= (1-p)^{2n-3}$$



$$\mathbb{P}(N>0) \geq \frac{(\mathbb{E}N)^2}{\mathbb{E}(N^2)} = \frac{\left(n(1-p)^{n-1}\right)^2}{n(n-1)(1-p)^{2n-3} + \left(n(1-p)^{n-1}\right)}$$

$$\geq \frac{n^2(1-p)^{2n-2}}{n^2(1-p)^{2n-3} + n(1-p)^{n-1}}$$

$$\frac{1}{\mathbb{P}(N>0)} \leq \frac{1}{1-p} + \frac{1}{n(1-p)^{n-1}} \quad ; \longrightarrow ②$$

★ <u>Result</u>:

$$n(1-p)^{n-1} \longrightarrow \begin{cases} 0 & \text{if } \; p > C\frac{\log n}{n} \text{ for } C>1 \\ \infty & \text{if } \; p < c\frac{\log n}{n} \text{ for } c<1 \end{cases}$$

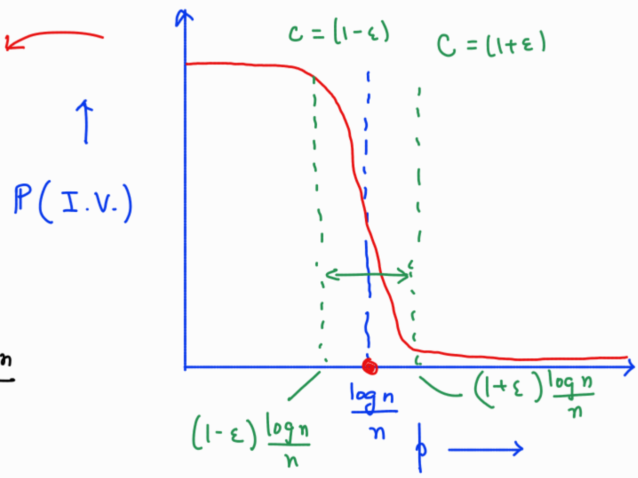
$\therefore$ For $p < c\frac{\log n}{n}$, we have

$$\frac{1}{\mathbb{P}(N>0)} \leq \underbrace{1+0}_{+o(1)} \Rightarrow \mathbb{P}(N>0) \geq 1 - o(1)$$

$$\therefore \lim_{n\to\infty} \mathbb{P}(N>0) = 1$$

$\square$

Phase Transition



☆ We have show that

$\{\exists$ an isolated vertex in $G(n,p)\}$

undergoes a phase transition at $p = \frac{\log n}{n}$

<u>Eg 2.</u> $\mathbb{P}\big(G(n,p)$ is connected$\big) \geq \mathbb{P}\big(G(n,p)$ has no isol. vertices$\big)$

$\therefore$ If $p \ll \frac{\log n}{n}$, then $\mathbb{P}\big(G(n,p)$ is connected$\big) = o(1)$

$p \gg \frac{\log n}{n}$, then $\mathbb{P}\big(G(n,p)$ is conn.$\big) = 1 - o(1)$.

Let Y_k be the # of connected components of size k, $k = 1, 2, \dots, n/2$.

Then,

$$\mathbb{P}(G(n,p) \text{ is } \underline{\text{not}} \text{ conn.}) = \underbrace{\mathbb{P}\left(\sum_{k=1}^{n/2} Y_k > 0\right)} \leq \sum_{k=1}^{n/2} \mathbb{E} Y_k.$$

- $\mathbb{E} Y_k \leq \binom{n}{k} \underline{(1-p)^{k(n-k)}}$.

$$\mathbb{E} Y_k = \mathbb{E}\left(\sum_{A \in \binom{[n]}{k}} Y_k^A\right)$$

subsets of n of size k

$$Y_k^A = \mathbb{1}[G_A \text{ is a conn. comp}]$$

$$\therefore\ \mathbb{P}(G(n,p) \text{ is disconn.}) \leq \sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)}$$

$$\leq \sum_{k=1}^{n/2} n^k (1-p)^{k(n-k)}$$

$k \leq \frac{n}{2} \implies n-k \geq \frac{n}{2}$

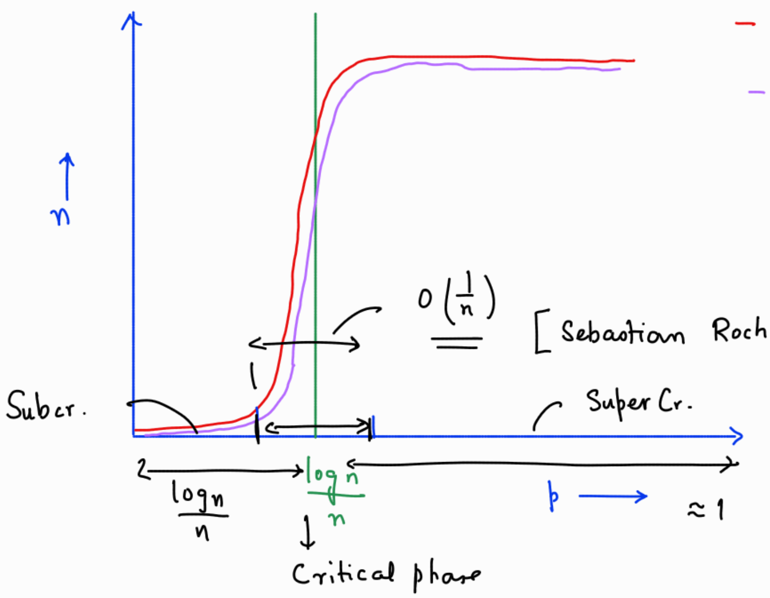
$$\leq \sum_{k=1}^{n/2} n^k \left((1-p)^{n/2}\right)^k \longrightarrow 0$$

as $n \to \infty$ for $p > C \frac{\log n}{n}$

$1-p \leq e^{-p}$

$$\therefore\ \mathbb{P}(G(n,p) \text{ is conn}) = 1 - \underline{o(1)} \text{ for } \underline{p > C \frac{\log n}{n}}.$$

☆



— $\mathbb{P}(G(n,p) \text{ has no isol. vertices})$

$\cup$

— $\mathbb{P}(G(n,p) \text{ is connected})$

As $n \uparrow \infty$, these prob. become indistinguishable

[Sebastian Roch: Example 2...]

- $|\text{SuperCr}|, |\text{Sub Cr}| \gg |\text{Crit}|$

Limit of Graph (V,E)

⊛ **Thm:** $\underbrace{p = \frac{\log n + s}{n}}$, then

$$\mathbb{P}(\#\text{ isol vertices}) = \frac{1}{1+e^{s}} + o(1)$$

$$\mathbb{P}\left(\sum_{k=2}^{n/2} Y_k > 0\right) = \underline{o(1)}$$