Lecture 18

Erdös-Renyi

Recall: $G \sim G(n,p)$ means that G is a graph on n (labeled) vertices, where each edge appears indep. w/ prob p.

& are diff. graphs

Monotone graph properties have sharp thresholds:

$$P\{G \sim G(n,p) \text{ has property } P\} \approx \begin{cases} 0 & \text{if } p \ll t(n) \\ 1 & \text{if } p \gg t(n) \end{cases}$$

Example: $P = \{G \text{ has } \geq 1 \text{ edge}\}$

monotone property

Claim: $t(n) = \frac{1}{n^2}$

↳ why?

Let X = # edges in G, $p_n = \frac{c}{n^2}$

$$P\{G \text{ has } 0 \text{ edges}\} = P\{X = 0\}$$
$$= (1-p_n)^{\binom{n}{2}} \quad \text{why?}$$
$$\to e^{-c/2} \text{ as } n \to \infty$$
$$\approx \begin{cases} 1 & \text{for } c \ll 1 \\ 0 & \text{for } c \gg 1 \end{cases}$$

Goal: For $P = \{G \text{ is connected}\}$, show $t(n) = \frac{\log n}{n}$

↳ Thm (Erdös-Rényi)

Let $p_n = \frac{\lambda \log n}{n}$, λ fixed

if $\lambda > 1$, then $P\{G \sim G(n,p_n) \text{ is connected}\} \to 1$ as $n \to \infty$

$\lambda < 1$, then $P\{G \sim G(n,p_n) \text{ is connected}\} \to 0$ as $n \to \infty$

Lemma: If X is a r.v. then

$$P(X=0) \leq \frac{\text{Var}(X)}{(\mathbb{E}[X])^2}$$

Pf: $$\text{Var}(X) = \mathbb{E}[X - \mathbb{E}[X]]^2 = \overbrace{P(X=0)\,\mathbb{E}[X - \mathbb{E}[X]^2 \mid X=0]}^{\mathbb{E}[X]^2}$$
$$+ \underbrace{P(X \neq 0)\,\mathbb{E}[X - \mathbb{E}[X]^2 \mid X \neq 0]}_{\geq 0}$$

$$\geq P(X=0)(\mathbb{E}[X])^2$$

Case $\lambda < 1$:

Will show: $P\{G \text{ has an isolated vertex}\} \to 1$

Note: $\{G \text{ has an isolated vertex}\} \subset \{G \text{ is disconnected}\}$ — subset of

$$X = \sum_{i=1}^{n} I_i = \# \text{ isolated vertices where } I_i = \begin{cases} 0 & \text{if } i \text{ not isolated} \\ 1 & \text{if } i \text{ isolated} \end{cases}$$

$$I_i \sim \text{Bern}(q) \qquad q := \underbrace{(1-p_n)}_{\text{prob of an edge not being placed}}{}^{\overbrace{n-1}^{(n-1)\text{ edges must not be placed}}}$$

$$\mathbb{E}[X] = n\mathbb{E}[I_1] = nq$$

$$\text{Var}(X) = \sum_i \text{Var}\, I_i + \sum_{i \neq j} \text{Cov}(I_i, I_j)$$

$$= nq(1-q) + n(n-1)\underbrace{\text{Cov}(I_1, I_2)}_{\text{due to the symmetric nature of the graphs}}$$

$$= \mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]^2$$

$$= nq(1-q) + n(n-1)(\underbrace{\mathbb{E}[I_1 I_2]}_{P\{\text{vertices } 1 \,\&\, 2 \text{ are isolated}\}} - \mathbb{E}[I_1]^2)$$

$$= \frac{q^2}{1-p_n} - q^2$$

$$= nq(1-q) + n(n-1)\left(\frac{p_n q^2}{1-p_n}\right)$$

? Why aren't we considering $\mathbb{E}[I_1]^2$

$$P\{G \text{ has no isolated vertices}\} = P\{X=0\} \leq \frac{nq(1-q) + n(n-1)\frac{p_n q^2}{1-p_n}}{n^2 q^2}$$

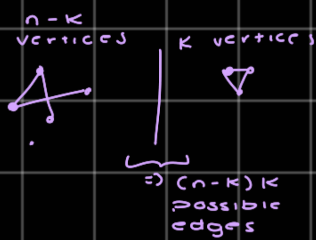
$$\leq \underbrace{\frac{(1-q)}{nq}}_{\to 0 \text{ bc } nq \to \infty} + \underbrace{\frac{p_n}{1-p_n}}_{\to 0} \text{ as } n \to \infty$$

bc $\log(nq) \sim \log n^{(1-\lambda)}$

$\therefore P\{G \text{ has an isolated vertex}\} \to 1$

Case 2: $\lambda > 1$:

$$P\{G \text{ disconnected}\} = P\left\{\bigcup_{k=1}^{n/2} \{\text{set of } k \text{ vertices seperated from the rest of } G\}\right\}$$



union bound

$$\leq \sum_{k=1}^{n/2} P\{\exists \text{ set of } k \text{ vertices seperated from the rest of } G\}$$

$$\leq \sum_{k=1}^{n/2} \binom{n}{k} P\{\text{vertices } \{1, k\} \text{ can be seperated from rest of } G\}$$

$$= \sum_{k=1}^{n/2} \binom{n}{k}(1-p_n)^{k(n-k)} \xrightarrow[\text{lots of tedious work}]{} 0$$

⤳ Now, onto module 3 !

Inference: want to infer X from knowing Y

· Basic setup:

model

X ⟶ $P_{Y|X}$ ⟶ Y

state of nature (often "parameter" or "hypothesis")

observational data

likelihoods $P_{Y|X}(\cdot|x)$ $f_{Y|X}(\cdot|x)$

X may or may not be a r.v.

↳ If it is, then distribution of X is called a "prior" (Bayesian inference)

↳ Example of inference when no prior is available

- Maximum Likelihood Estimation:

$$\mathrm{MLE}(X \mid Y=y) := \underset{x}{\arg\max}\ P_{Y|X}(y|x)$$ (break ties arbitrarily)

→ ex: model is $Y \sim \mathcal{N}(x, \sigma^2)$

$\mathrm{MLE}(X \mid Y=y) = Y$