March 31 2015

LECTURE #7

Tipping Point / Phase Transition / 0-1 Laws.

- Describe a phenomenon, where an event occurs or does not occur a.s. [almost surely]
- With a small change in the value of a critical parameter, the event of interest occurs either ALMOST NEVER (a.n. prob = 0) to ALMOST SURELY (a.s. prob = 1)
  → the transition in probability occurring very quickly.

GAME OF "FRIENDING"

- Imagine sending a friend request randomly to $(n-1)$ other individuals in a network (with total of n individuals).

◇ Key Assumptions:

(a) If the recipient is already a friend, he simply ignores the request.

(b) Otherwise, he receives your request for the first time and accepts you as a friend.

(c) Under no circumstances, does he ignore, decline or unfriend you.

TIPPING POINT

After $\Theta(n \ln n)$ requests, one will have a.s. befriended all the other $(n-1)$ individuals.

Coupon Collector's Problem

"COLLECT-ALL-COUPONS-AND-WIN" Contest.

Problem Statement

- There are n distinct coupons.
- Coupons can be collected with replacement

  $x_1, x_2, \ldots, x_n \sim \text{Bernoulli}(1/n)$ iid r.v.s.

  $(x_i = 1) \equiv$ Event that you obtain the i^{th} coupon.
- What is the probability that more than t ~~samples~~ trials are needed to collect all n coupons?

$\equiv$ How many coupons are expected to be drawn with replacement before each of the n coupons has been drawn at least once?

Example: $\left.\begin{array}{l}\text{Let } n = 52.\\ \text{Then } t = 225\end{array}\right\} t = \Theta(n \ln n)$

If you draw a card randomly (with replacement) from a full deck, then after $t = 225$ draws you would have seen every card at least once almost surely.

t_i = Time to collect i^{th} coupon after collecting $(i-1)^{th}$ coupon.

$\therefore$ t_i's are independent.

$$t = \sum_{i=1}^{n} t_i \equiv \text{Time to collect all coupons}$$

$$p_i = Pr[\text{ Collect a new coupon after } (i-1)^{th}]$$
$$= \frac{n-i+1}{n}$$

$$t_i \sim \text{Geometric}(1/p_i)$$
$$Pr[t_i = k] = (1-p_i)^k p_i$$

$$E(t_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$$

$$Var(t_i) = \frac{1}{p_i}\left(\frac{1}{p_i} - 1\right) = \frac{1-p_i}{p_i^2} = \frac{(i-1)n}{(n-i+1)^2}$$

$$E[t] = E[\Sigma t_i] \quad \text{By Linearity of Expect}^n$$
$$= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1}$$
$$= n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) = n H_n$$
$$= n \int^{n} \frac{1}{x}\, dx + \gamma n + 1/2 + o(n)$$
$$= n \ln n + O(n)$$

$\gamma \approx 0.577$ = Euler's const.

$Var(t) = Var\left(\sum t_i\right)$ $\quad t_i$'s are indep.

$$\le \frac{n^2}{n^2} + \frac{n^2}{(n-1)^2} + \cdots + \frac{n^2}{1} = \frac{\pi^2}{6} n^2$$

$$\sigma(t) = \frac{\pi n}{\sqrt{6}}$$

$$\Pr\left(|t - nH_n| \ge c \cdot n\right) = \Pr\left(|t - nH_n| \ge \left(c\frac{\sqrt{6}}{\pi}\right)\sigma\right)$$

$$\le \frac{\pi^2}{6c^2}$$

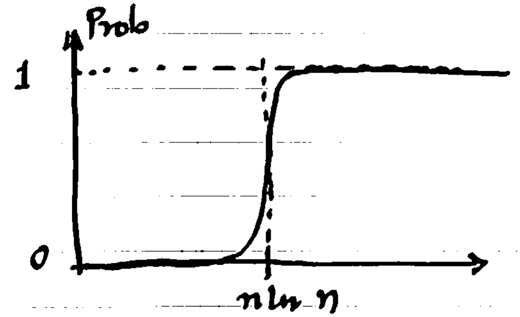
$$\Pr\left(|t - nH_n| \ge 10 \cdot n\right) \le \frac{\pi^2}{600} \approx \frac{1}{60}$$

$t < (1-\varepsilon) nH_n \Rightarrow$ You will not have all the coupons a.s.

$t > (1+\varepsilon) nH_n \Rightarrow$ You will have all the coupons a.s.

Generalization: t_k = First time k copies of each coupons are collected.

$$t_k \approx n \ln n + (k-1) n \ln \ln n + \Theta(n)$$



Random Graphs.

$G(n,p)$ - Model

$$G = (V, E) \sim G(n,p)$$

$$|V| = n$$

$$\langle |E| \rangle = \binom{n}{2} p$$

$$\langle d \rangle = (n-1)p$$

$$\Pr[d(v) = k] = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$\approx \frac{\lambda^k}{k!} e^{-\lambda} \qquad \lambda = \langle d \rangle = (n-1)p = \text{const.}$$

$$d(v) \sim \text{Poisson}(\lambda)$$

$$\mu[d(v)] = \lambda = (n-1)p$$

$$\sigma^2[d(v)] = \lambda = (n-1)p$$

THRESHOLD FUNCTIONS FOR CONNECTIVITY

(Erdös - Rényi 1961)

for the Erdös - Rényi model $G(n,p)$ is

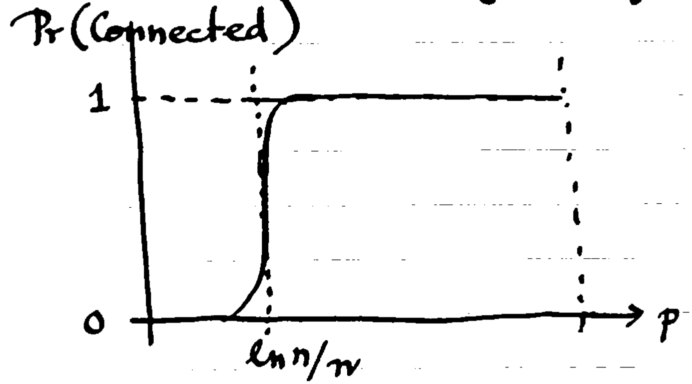
$$t(n) = \frac{\ln n}{n}$$

Small p: $p < \frac{(1-\epsilon)\ln n}{n} \Rightarrow$ Graph is almost surely DISCONNECTED

Large p: $p > \frac{(1+\epsilon)\ln n}{n} \Rightarrow$ Graph is almost surely CONNECTED.

For a graph $G=(V,E) \sim G(n, \lambda \frac{\ln n}{n})$

$$\Pr[G = \text{Connected}] = \begin{cases} 0 & \text{if } \lambda < 1 \\ 1 & \text{if } \lambda > 1 \end{cases}$$



Indicator Variable:

$$\mathbb{1}_i = \begin{cases} 1, & \text{if node } i \text{ is isolated;} \\ 0, & \text{o.w.} \end{cases}$$

$\mathbb{1}_i \sim \text{Bernoulli}(\pi)$

$$\pi = \Pr[\mathbb{1}_i = 1] = E[\mathbb{1}_i] = (1-p)^{n-1}$$

$$= (1-p)^{\frac{1}{p}(n-1)p} = e^{-(n-1)\lambda \frac{\ln n}{n}} \approx e^{-\lambda \ln n} \approx n^{-\lambda}.$$

$\diamond$ $\boxed{\mathbb{1}_i \sim \text{Bernoulli}(n^{-\lambda})}$

$X = \sum \mathbb{1}_i$ = Total # isolated vertices.

$$E[X] = \sum E[\mathbb{1}_i] = n \cdot n^{-\lambda} = n^{1-\lambda} \to \begin{cases} \infty & \text{if } \lambda < 1 \\ 0 & \text{if } \lambda > 1. \end{cases}$$

NOTE BOOK

Sharpening the Intuition.

- Prove: $\lambda < 1 \Rightarrow Pr[X=0] = 0$.
- Note : $\lambda > 1 \Rightarrow Pr[X=0] > 0$

$\not\Rightarrow$ Graph is connected!

(1) $E[X] = n^{1-\lambda}$

$$Var[X] = \sum_i Var(\mathbb{1}_i) + \sum_i \sum_{j \neq i} cov(\mathbb{1}_i \mathbb{1}_j)$$

$$= n \underbrace{Var(\mathbb{1}_i)}_{\pi(1-\pi)} + n(n-1) \underbrace{cov(\mathbb{1}_i \mathbb{1}_{i+1})}_{\pi^2\left(\frac{p}{1-p}\right)} \quad \leftarrow \text{Prove!}$$

$$Var[X] = n\pi - n\pi^2 + (n^2-n)\pi^2\left(\frac{p}{1-p}\right)$$

$$\approx n\pi + n^2\pi^2 p$$

$$\approx n^{1-\lambda} \approx E[X] \quad (\text{if } \lambda < 1)$$

$$Var[X] = \int (x - E[x])^2 \, p(x)\, dx$$

$$\geq (0 - E[X])^2 \; Pr[X=0].$$

$$\therefore \; Pr[X=0] \leq \frac{Var[X]}{E[X]^2} \approx \frac{1}{E[X]} \to 0$$

NOTE BOOK

"Graph is Disconnected"

$$\Rightarrow \exists\, V' \subset V \qquad \text{Disconnect } [V', V \setminus V']$$
$$|V'| = k \le \frac{n}{2}$$

$$\text{Prob} \le (1-p)^{k(n-k)}$$
$$\approx (1-p)^{\frac{1}{p}\, k(n-k)p}$$
$$= e^{-\lambda\, k(n-k) \frac{\ln n}{n}}.$$

$$\text{Prob} \le \binom{n}{k} (1-p)^{k(n-k)}$$
$$= \frac{n!}{k!\,(n-k)!}\, e^{-\lambda k(n-k) \frac{\ln n}{n}}$$
$$< \frac{\sqrt{2\pi n}\,(n/e)^n}{\pi n\,(n/2e)^{n/2}\,(n/2e)^{n/2}}\, e^{-\lambda \frac{n^2}{4} \frac{\ln n}{n}} \qquad \left[k = \frac{n}{2}\right]$$
$$< \sqrt{\frac{2}{\pi n}} \cdot 2^n \cdot 2^{-\lambda n \lg n/4}$$

If $\lambda > 1$, then

$$\Pr[\text{Graph is disconnected}]$$
$$\le \sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)}$$
$$< \frac{n}{2} \sqrt{\frac{2}{\pi n}}\; 2^{-[n \lg n/4 - n]}$$
$$= \sqrt{\frac{n}{2\pi}}\; 2^{-[n \lg n/4 - n]}$$
$$= \frac{1}{\sqrt{2\pi}}\; 2^{-[n \lg n/4 - n - \lg n/2]} \xrightarrow[\text{as } n \to \infty]{} 0. \quad \blacksquare$$