Lecture #6

March 12 2013

Coupon - Collector Problem:

Given n coupons, how many coupons are expected to be drawn with replacement, before each coupon has been drawn at least once.

$$T = \Theta(n \lg n)$$

$$Pr\left(|T - nH_n| \geq c \cdot n \right)$$

$$= Pr\left(|T - nH_n| \geq \left(c \frac{\sqrt{6}}{\pi} \right) \frac{\pi}{\sqrt{6}} n \right) \leq \frac{\pi^2}{6c^2}$$

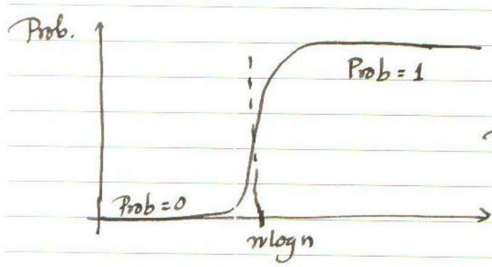
If $T < (1-\varepsilon)\, nH_n$, you are a.n. getting all the coupons.

If $T > (1+\varepsilon)\, nH_n$, you are a.s. getting all the coupons.

GENERALIZATION

T_k = First time k copies of each coupons are collected

$$E(T_k) = n \log n + (k-1)\, n \log\log n + \Theta(n) \quad \text{as } n \to \infty$$



0-1 Laws
Phase Transition.

Random Graph.

Two Models.

$G(n, M)$ Models. $|V| = n$, $|E| = M$.

$G(n, p)$ Models $|V| = n$

$\forall_{u,v \in V}$ $Pr[(u,v) \in E] = p$.
i.i.d.

Density of the graph.

$$\hat{p} = \frac{M}{\binom{n}{2}} \qquad \langle \text{density} \rangle = \frac{E(M)}{\binom{n}{2}} = \frac{p\binom{n}{2}}{\binom{n}{2}} = p.$$

At $\hat{p} = \frac{1}{2}$ all graphs on n vertices are chosen with equal probability

$Pr[G(n, M)]$

$$(\hat{p})^M (1-\hat{p})^{\binom{n}{2} - M} = \frac{1}{2^M} \cdot \frac{1}{2^{\binom{n}{2} - M}}$$

$$= 2^{-\binom{n}{2}} \longrightarrow \text{Ind. of } M.$$

$d(v) \sim Bin(n-1, p)$ $\mu_d = (n-1)p$ $\sigma_d^2 = (n-1)p(1-p)$

If $np = \lambda = const.$

$d(v) \sim Poisson(\lambda)$ $\mu_d = np$ $\sigma_d^2 = np$.

Connectedness

Consider $G(n,p)$ models.

◇ If $p < \frac{(1-\varepsilon)\ln n}{n}$ then $G(n,p)$ a.s. is not connected – i.e. it has isolated vertices.

◇ If $p > \frac{(1+\varepsilon)\ln n}{n}$ then $G(n,p)$ a.s. is connected.

Questions about Random Social Network Models.

1) Does the graph have isolated nodes?
Cycles?
Giant connected components?

2) What are the probabilities of such events?

3) Asymptotic Analysis
Compute probabilities, as $n \to \infty$

TIPPING POINTS

Phase Transition

0-1 Laws

Either a probability Approaching 1 }

Or a probability Approaching 0 }

Asymptotically (In the limit $n \to \infty$).

THRESHOLD FUNCTIONS FOR CONNECTIVITY.

Erdös - Rényi 1961.

A threshold function for the connectivity of the Erdös-Rényi model $G(n,p)$ is

$$t(n) = \frac{\log n}{n}.$$

In other words, for $G(n, \lambda \frac{\log n}{n})$

- If $\lambda < 1$, $Pr(\text{Connectivity}) = 0$
- If $\lambda > 1$, $Pr(\text{Connectivity}) = 1$.

$$I_i = \begin{cases} 1, & \text{if node } i \text{ is isolated.} \\ 0, & \text{otherwise.} \end{cases}$$

I_i = Bernoulli r.v. $\sim$ Bernouli (π)

$$\pi = \Pr[I_i = 1] = (1-p)^{n-1} = (1-p)^{\frac{1}{p}\cdot(n-1)p}$$
$$= e^{-np}$$
$$= e^{-n\lambda\frac{\log n}{n}}$$
$$= e^{-\lambda \log n} = n^{-\lambda}$$

$I_i \sim$ Bernoulli $(n^{-\lambda})$

$X = \sum I_i$ = Total # Isolated Nodes

$$E[X] = n \cdot n^{-\lambda} \longrightarrow \begin{cases} \infty, & \text{if } \lambda < 1; \\ 0, & \text{if } \lambda > 1. \end{cases}$$

Problems. :

(1) Need to show that $\Pr[X=0] = 0$, if $\lambda < 1$.

(2) Note, also that $\Pr[X=0] \neq 0$ does not necessarily imply that the graph is connected.

Let's consider the case, $\lambda < 1$ ...

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

$$= n\, var(I_1) + n(n-1)\, cov(I_1, I_2)$$

$$var(I_1) = \pi(1-\pi)$$

$$cov(I_1, I_2) = E(I_1 I_2) - E[I_1]\, E[I_2]$$

$$(1-p)^{2n-3} = \frac{\pi^2}{(1-p)} \qquad \pi^2$$

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

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

$$\sim n\, n^{-\lambda} = E(X)$$

$$Var(x) \geq (0 - E(x))^2\, Pr[x = 0]$$

$$\Rightarrow Pr[X=0] \leq \frac{Var(x)}{E(x)^2} = \frac{1}{E(x)} \to 0$$

"Graph is disconnected"

$\Rightarrow \exists_{V' \subset V,\ |V'| = k,\ k \leq n/2}$ Disconnect $[V', V \setminus V']$

$Pr[V'$ is not connected to $V \setminus V']$

$= (1-p)^{k(n-k)}$

$Pr[\exists_{V',\ |V'| = k}\ V'$ is not connected to $V \setminus V']$

$= \binom{n}{k} (1-p)^{k(n-k)}$

$Pr[$Graph is disconnected$]$

$$\approx \underbrace{\sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)}}_{\lim_{n\to\infty} \to 0}$$

when $p = \frac{\lambda \log n}{n}, \lambda > 1.$

◇ GIANT COMPONENT.

$G(n, p(n))$ - Erdös-Rényi Models.

◇ We know that if $p(n) \ll \frac{\ln n}{n}$ then the graph is a.s. disconnected.

In fact, the graph has an arbitrarily large number of connected components.

◇ TWO REGIMES:

$$p(n) = \frac{\lambda}{n} \quad \begin{cases} \lambda < 1 \\ \text{vs} \\ \lambda > 1. \end{cases}$$

1) For $\lambda < 1$, all components of the graph are "small."

2) For $\lambda > 1$, one component of the graph is a unique "giant" component.