Day 7
Probabilistic method
確率的手法-2

- Tools in probability
  - Random graph
  - Coupon collector problem クーポン収集問題
  - Random walk and algorithms 乱歩とアルゴリズム

Probability is not linear

- The probability of A or B happens is not always Pr(A) + Pr(B)
- The probability of A and B happne is not always Pr(A)Pr(B)
  - When it happens??
  - You will see that you should always aware of above, even in real life….

Puzzle of 100 prisoners

- Story: The names of 100 prisoners are placed in 100 wooden boxes, one name to a box, and the boxes are lined up on a table in a room. One by one, the prisoners are led into the room, each may look in at most 50 boxes, but must leave the room then, and is permitted no further communication with the others. The prisoners have a chance to plot their strategy in advance, and they are going to need it, because unless every single prisoner finds his own name all will subsequently be executed.
- Problem: Find a strategy for them which has probability of success exceeding 30%

100人の囚人の問題

- 100 人の囚人の氏名が、テーブルに一列に並んだ100 個の木箱に一つづつ入っている。囚人は一人づつ呼ばれて、高々50個の箱の中身を見ることができる。終了後は退室し、他の囚人との情報交換はできない。囚人たちは前もって作戦を立てることができる。囚人全員が自分の名前を見つけない限り、全員の囚人が一斉に処刑される運命にある
- 問題: 成功確率が30%以上になる作戦を立案せよ

Intuition

- The success probability of each prisoner is exactly 0.5.
- Thus, the probability that all prisoners find their names is $(0.5)^{100}$, which is less than 0.000,000,000,000,000,000,000,000,000,001.
- How to increase it to 0.3?
- Expectation number of names found is 50.
- Concentrate the success!
  - If we make “at least 75 prisoners fail” or “all prisoners success”, we are done. (why??)

Random graph
ランダムグラフ

- Random regular graph with degree d. A "random" graph with constraint that each node degree is d.
- 次数dのランダム正則グラフ
- The key of the puzzle is the probability that a random graph with degree 2 has a "giant component".
- 次数２のランダム正則グラフが大きい連結成分を持つ確率を考える

Use of random graph

- Expander : d-regular graph with high expansion ratio エクスパンダグラフ
  - h(G) = min { |Γ(S)|/|S| : |S| > |V|/2}
    - |Γ(S)| = cut size = number of edges in cut
  - Means that it is well connected
  - “Ramanujan graph”
- Super concentrator：超凝縮グラフ

A graph with a set O of n input and I of n output nodes, such that for any O’ ⊂O and I’ ⊂I we have vertex disjoint paths connecting O’ and I’.

Coupon collector problem
クーポン収集問題

- There are n types of coupons and at each trial a coupon is chosen at random. Each trial is independent from others. Let X be the random variable showing the number of trials required to have all kinds of coupons. n 種類のクーポンを揃えるために、何枚のクーポンを集めるだろうか。
  - Find the expected value E(X) of X
  - Find the variance E(X) of X

Analysis of coupon collector

- x(i): the time to find a new coupon after having i different coupons.
  - $p_i = (n-i)/n$: probability to find a new coupon
- $E(x(i)) = 1/p_i = n/(n-i)$
- $V(x(i)) = (1-p_i)/p_i^2 = ni/(n-i)^2$

$= n^2/(n-i)^2 - n/(n-i)$

- Expectation and variance are linear!, so add x(i) for i=1,2,..,n

http://www-stat.stanford.edu/~susan/surprise/Collector.html

Some beautiful formulas

$$1+1/2+1/3+...+1/n \le n \ln n + \gamma$$

$$\varsigma(s) = 1+1/2^s+1/3^s+1/4^s+.... = \sum_n 1/n^s$$

Rieman zeta function

$$\varsigma(2) = \pi^2/6$$

Chebyshev's inequality: For a random variable X with standard deviation σ and expectation μ, Prob(|X – μ | > t σ) < $1/t^2$

Stronger tools: Chernoff's ineuality, Azuma's inequality, which I omit here

Proof of Chebyshev's inequality

- Consider $Y = |X – \mu|^2$
- $E(Y) = V(X) = \sigma^2$
- $Prob(E(Y) > a) \quad > E(Y)/a$
  - This is called Markov's inequality
- Set $a = \sigma^2 t^2$ and we have Chevshev's inequality.
  - $Prob(E(|X-\mu|) > \sigma t) = Prob(E(Y)> \sigma^2 t^2)$
  - $> E(Y)/ \sigma^2 t^2 = \sigma^2 / \sigma^2 t^2 = 1/ t^2$

Collecting 52 cards

Euler constant γ is 0.577

Expectation is n (ln n + 0.577)

If n = 52, E(X) ~250

Variance is $(\pi n)^2/6 - n \ln n \sim 2500$

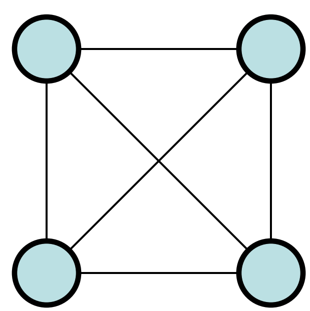
$\sigma \sim 50$

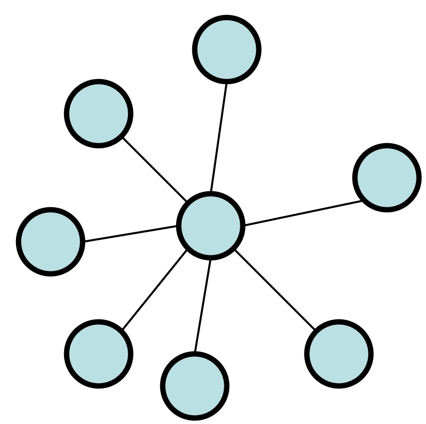
Use Chebyshev's inequality for t= 4:

We need more than 450 steps with probability less than 1/16

Ramdom walk on a graph

- We start at a node u of a graph and select a neighbor node with uniform probablity.
  - Expected time to visit a node v (reaching time)
  - Expected time to visit all vertices (cover time)
- Solve this problem if G is
  - Clique
  - Star graph
  - Path





Commuting time

- Let G be a graph with m edges
- For each vertex u and v, let h(u,v) be the reaching time from u to v in a random walk
- C(u,v) = h(u,v)+ h(v,u): commuting time

Theorem. C(u,v) = m R(u,v), where R(u,v) is the effective resistance if we consider G as an electric circuit

交換時間は、グラフを電気回路と見たときの有効抵抗にグラフの辺数をかけたもの

A "Magical" Proof

- By definition, if x is not v,
  - $h(x, v) = (1/d(v)) \Sigma_{y \in \Gamma(x)} (h(y, v)+1)$
  - $\Gamma(x)$: neighbors of x
- Regard G as a network such that each edge has unit resistance
- Inject d(x) amperes from each x, and remove 2m amperes from v
  - $d(v)= \Sigma_{y \in \Gamma(x)} (\varphi(x, v) - \varphi(y,v))$
- Then, $\varphi(x,v)=h(x,v)$ (Why??)

- Then, $\varphi(x,v)=h(x,v)$ (Why??)
- If we remove $d(x)$ amperes from x and inject 2m amperes from v,
- $-\psi(x, u) = h(x,u)$
- Thus, $\varphi(u,v) + \psi(u,v) = h(u,v) + h(v,u)$
- The left side: Inject 2m amperes from v and remove 2m amperes from u
  - Amperes from other nodes are cancelled.
- Thus, $2m\, R(u,v) = h(u,v) + h(v,u)$

Coin flip Game (why rich person wins)

Alice and Bob play a coin flipping game.

Each has n dollars, and if the coin is head, Alice get 1 dollar, if tail Bob get 1 dollar.

How may expected tosses will be done until one of them bankrupt?

How about Alice has 100 dollars and Bob has 50 dollars?

WalkSAT Algorithm

- 2-CNF(conjunctive normal form) formulas
- F: AND of (x or y) clauses (m clauses)
- x , y are among x(i) or $\overline{x(i)}$ for i=1,2,..n
- Judge whether there is Boolean assignment of variables to make F “true”.

Algorithm: If there is an clause that is currently “false”, flip one of variables.

Problem: Analyze its running time.