CS 174

What is a Randomized Algorithm?

Traditionally: $x \longrightarrow \boxed{\mathcal{A}} \longrightarrow f(x)$ (deterministic)

Randomized: $\xrightarrow{x}$, $\xrightarrow[r \in \{0,1\}^T]{} \boxed{\mathcal{A}} \longrightarrow \tilde{f}(x)$ (depends on random bits r)

$\hookrightarrow$ r is T bits $\to$ T choices $\to$ 2^T possibilities $\to$ $\Pr[\tilde{f}(x) = \text{True}]$?

Often, we use higher level randomizations rather than binary.

ex) "pick a random element from set S", "permuter S randomly"

$\hookrightarrow$ These can be simulated using binary decisions!

Example) Is x prime? (Primality Testing)

Want: x is prime $\Rightarrow \Pr[\tilde{f}(x) = \text{Yes}] \simeq 1 \ (\geq 3/4)$

x not prime $\Rightarrow \Pr[\tilde{f}(x) = \text{Yes}] \simeq 0 \ (\leq 1/4)$

amplified using repetition

Ideally: x is prime $\Rightarrow \Pr[\tilde{f}(x) = \text{Yes}] = 1$

x not prime $\Rightarrow \Pr[\tilde{f}(x) = \text{Yes}] \simeq 0 \ (\leq 1/2)$

$\Rightarrow$ one-sided errors

$\hookrightarrow$ run T trials, and only output "Yes" iff all trials output "Yes".

$\Rightarrow$ Error probability of <u>false positive $\leq 2^{-T}$</u>!

Amplification: $\Pr[\text{error in } 2k+1 \text{ trials}] \leq \sum_{i=0}^{k} \binom{2k+1}{i} \left(\frac{3}{4}\right)^i \left(\frac{1}{4}\right)^{2k+1-i} \leq \frac{1}{2} \cdot 2^{2k+1} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{k+1}$

$\leq 4^k \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^k \leq \left(\frac{3}{4}\right)^k$ $\Rightarrow$ every repetition, error reduces geometrically!

Application) Polynomial Identity Testing (Is $F(x) \equiv G(x)$?)

ex) $(x-2)(x-1)(x+4)(x+1) = x^4 - 2x^3 - 9x^2 - 2x + 8$?

Naively, expand the LHS $\rightarrow O(d^2)$ time, where $d :=$ degree of polynomial

Randomized: Pick a random integer $r \in [1, R]$.

If $F(r) = G(r)$, output "Yes". Else, output "No".

Analysis: One-sided error for "No" (could have common root).

What is $\Pr[F(r) = G(r) \mid F(x) \not\equiv G(x)]$? $H(x) := F(x) - G(x)$.

Then, degree$(H) \leq d \rightarrow$ at most d points where $F(x) - G(x) = 0$!

$\Rightarrow$ If we set $\underline{R = 2d}$, $\Pr[\text{error}] \leq \frac{d}{R} \leq \frac{d}{2d} = \frac{1}{2}$. //

Extra: for multivariate functions $F(x_1, \ldots, x_m) \equiv G(x_1, \ldots, x_m)$,

the gain from randomization becomes better! (exponential $\rightarrow$ polynomial)

Probability

Probability Space: finite/countably infinite set of results Ω

$\forall \omega \in \Omega$, $\exists\, 0 \leq \Pr[\omega] \leq 1$. $\sum_{\omega \in \Omega} \Pr[\omega] = 1$.

Event: Some subset $E \subseteq \Omega$. $\Pr[E] = \sum_{\omega \in E} \Pr[\omega]$.

Examples) (1) $\Omega = [n]$, $\Pr[\omega] = \frac{1}{n}$. (2) Roll 2 fair dice. $\Omega = [6] \times [6]$.

$\Pr[(a,b)] = \frac{1}{36}$.

(3) Balls & Bins. Toss m balls into n bins indep. & u.a.r.

$\Omega = \{1, \dots, n\}^m$ (m-tuple of bin choices). $\Pr[(i_1, \dots, i_m)] = \frac{1}{n^m}$.

(4) Random Permutations. $\Omega = \{$ set of all permutations of n items $\}$.

$\Pr[\pi] = \frac{1}{n!}$ where $\pi(i) :=$ position of i-th item in permutation π.

* Dual views: permutation, or sequence of choices (trivially equivalent)

(5) Poker Hands. Choose 5 cards (unordered) from 52-card deck.

$|\Omega| = \binom{52}{5}$. $\Pr[\omega] = 1/\binom{52}{5}$.

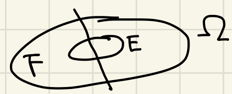
Calculating Probabilities: 1) If $\{E_i\}$ is disjoint, $\Pr[\bigcup_i \{E_i\}] = \sum_i \Pr[E_i]$.

2) For any E_1, E_2: $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2]$.

For any $E_1, \dots, E_n$: $\Pr[\bigcup_i E_i] = \sum_i \Pr[E_i] - \sum_{i<j} \Pr[E_i \cap E_j] + \sum_{i<j<k} \Pr[E_i \cap E_j \cap E_k] \cdots$

3) Union Bound: $\Pr[\bigcup_i E_i] \leq \sum_i \Pr[E_i]$ (RHS always overcounts / is exact).

4) Complement: $\Pr[\bar{E}] = 1 - \Pr[E]$.



Conditional Probability: New Probability Space of $\tilde{\Pr}[\omega] = \begin{cases} 0 & \text{if } \omega \notin F \\ \frac{\Pr[\omega]}{\Pr[F]} & \text{if } \omega \in F. \end{cases}$

For $E \subseteq \Omega$: $\tilde{\Pr}[E] = \frac{\Pr[E \cap F]}{\Pr[F]} = \Pr[E \mid F]$.

Independence: $E \perp\!\!\!\perp F$ if $\Pr[E \mid F] = \Pr[E]$ / $\Pr[E \cap F] = \Pr[E]\Pr[F]$.

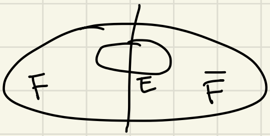
Bayes' Rule: $\Pr[E \cap F] = \Pr[E \mid F]\Pr[F] = \Pr[F \mid E]\Pr[E]$.

$\hookrightarrow \boxed{\Pr[E \mid F] = \frac{\Pr[F \mid E]\Pr[E]}{\Pr[F]}}$ ☆

Iterating $\Pr[E \cap F] = \Pr[E|F]\Pr[F]$ results in

$$\Pr\left[\bigcap_{i=1}^{n} E_i\right] = \Pr[E_1] \cdot \Pr[E_2|E_1] \cdot \Pr[E_3|E_1, E_2] \cdots\cdots \Pr\left[E_n \,\middle|\, \bigcap_{i=1}^{n-1} E_i\right].$$

Law of Total Probability: $\Pr[E] = \Pr[E \cap F] + \Pr[E \cap \bar{F}]$



$$= \Pr[E|F]\Pr[F] + \Pr[E|\bar{F}]\Pr[\bar{F}].$$

Can be generalized to any partition $\{F_i\}$ $(i \le k)$

$$\Pr[E] = \sum_{i=1}^{k} \Pr[E|F_i]\Pr[F_i].$$

Examples) Dice: $\Pr[\text{sum of 2 dice roll} = 10] = \frac{3}{36}$ (counting)

$\Pr[\text{second die} > \text{first die}] = (1 - \Pr[\text{equal values}])/2 = \frac{5}{12}$ (symmetry)

Balls & Bins: $\Pr[\text{first bin empty}] = \left(\frac{n-1}{n}\right)^m = \left(1 - \frac{1}{n}\right)^m$ (independence)

$(n-1)^m / n^m$ (counting), $\sim e^{-m/n}$ as $m, n \to \infty$ (asympotics)

Permutations: $\Pr[1 \text{ is a fixed point}] = \frac{1}{n}$ (sequence intepretation)

$(n-1)!/n!$ (counting)

$\Pr[7 \,\&\, 17 \text{ are fixed points}] = \frac{1}{n(n-1)}$ (sequence interpretation)

$\Pr[\pi \text{ contains a fixed point}] = \Pr\left[\bigcup_{i=1}^{n} E_i\right]$ where $E_i := i$ is fixed.

Define $P_1 = \frac{1}{n}$. $P_2 = \frac{1}{n(n-1)}$. $\cdots$ $P_n = \frac{1}{n!}$ where $P_x :=$ x points are fixed.

$$\to \Pr\left[\bigcup_{i=1}^{n} E_i\right] = \sum_i \Pr[E_i] - \sum_{i<j} \Pr[E_i \cap E_2] + \cdots$$

$$= n \cdot P_1 - \binom{n}{2} P_2 + \binom{n}{3} P_3 \cdots = 1 - \frac{1}{2!} + \frac{1}{3!} \cdots \sim e^{-1}$$

$\Rightarrow \Pr[\pi \text{ is a derangement}] \sim 1-\frac{1}{e}$ (asymptotic)

Poker Hands: $\Pr[\text{two-pair hand}] = \dfrac{\#\text{ of 2pair hands}}{\#\text{ of all hands}}$

$$= \frac{\binom{13}{2}\times\binom{4}{2}^2\times(52-8)}{\binom{52}{5}} \quad \left(\frac{(\text{values})\times(\text{suits})\times(\text{remaining one card})}{(\text{all hands})}\right) \simeq 0.0475$$

Application) Bayesian Inference: 3 coins with $\Pr[H]$ of $\overset{C_1}{\frac{1}{2}}, \overset{C_2}{\frac{2}{3}}, \overset{C_3}{1}$.

We pick a coin at random and toss it. It comes up heads.

What can we conclude about the probability of the coin we chose?

$\Pr[C_1] = \Pr[C_2] = \Pr[C_3] = 1/3$. (prior distribution)

Want: $\Pr[C_i|H]$. This is easy to calculate using $\Pr[H|C_i]$!

$$\rightarrow \Pr[C_i|H] = \frac{\Pr[H|C_i]\Pr[C_i]}{\Pr[H]} = \frac{1}{3}\,\frac{\Pr[H|C_i]}{\sum_{i=1}^{3}\Pr[H|C_i]\Pr[C_i]} \rightarrow \tfrac{1}{3}\left(\tfrac{1}{2}+\tfrac{2}{3}+1\right)$$

$$\Rightarrow \Pr[C_i|H] = \left\{C_1 = \frac{3}{13},\ C_2 = \frac{4}{13},\ C_3 = \frac{6}{13}\right\}.$$

Application) Matrix Multiplication Testing: 3 $n\times n$ matrices A, B, C.

Is $AB = C$? Naively, multiply AB and compare with C, $\sim O(n^{2.7})$

Randomized: Pick a random vector $r \in \{0,1\}^n$. If $A(Br) = Cr$, then output "Yes". Else, output "No". $\rightarrow O(n^2)$ with 3 (Mat x vector)

Error Analysis: Only if $AB \neq C$ and $ABr = Cr$. Define $D := AB - C$.

$D \neq 0 \Rightarrow$ WLOG, $d_{11} \neq 0$. $\quad \overset{d_1}{D}\,[r] = \begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix} \Rightarrow \sum_{j=1}^{n} d_{1j} r_j = 0$

(for analysis' sake) $\quad \Rightarrow r_1 = -\frac{1}{d_{11}}\sum_{j=2}^{n} d_{1j} r_j$

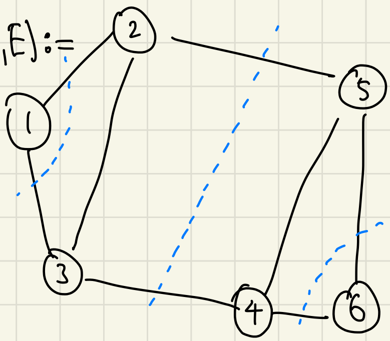
$\Rightarrow r_1$ s.t. $D=0$ is unique $\Rightarrow Pr[D_r = 0] \leq \frac{1}{2}$. (deferred decision!)

Concretely, $Pr[D_r = 0] \leq Pr[r_1 = -\frac{1}{d_{11}} \sum_{j=2}^{n} d_{1j} r_j]$

$$= \sum_{(r_2 \cdots r_n)} \underbrace{Pr[r_1 = \cdots | r_2 \ldots r_n]}_{\leq 1/2} \cdot Pr[r_2 \ldots r_n] \leq \frac{1}{2}.$$

Algorithm) Karger's Randomized Min-Cut

G(V,E):=



Min-Cut: Partiton (S, V-S) s.t. # of edges crossing S and V-S is minimized

* max s-t flow $\Leftrightarrow$ min s-t cut

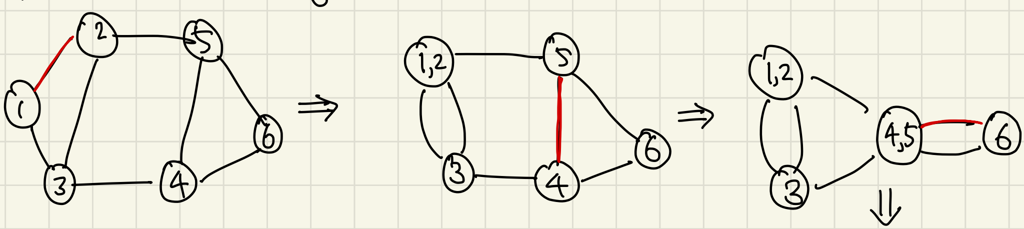
$\hookrightarrow$ can be done in $O(n \cdot n^3 \log n) \simeq O(n^4)$

while # vertices > 2:

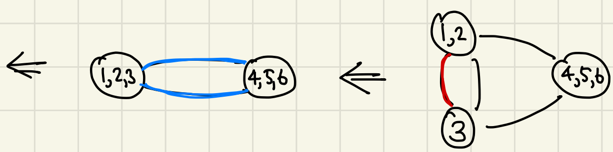
pick an edge u.a.r and contract it.

Output the remaining cut.

ex)



Termination, return (1,2,3) (4,5,6).



Claim: Let C be any min-cut in G. Then Pr[Algorithm outputs C] $\geq \frac{2}{n(n-1)}$.

Corollary: Run $T = O(n^2)$ indep. trials and output the best result.

$Pr[\text{fail to find } C] \le \left(1 - \frac{2}{n(n-1)}\right)^T \le e^{-\frac{2T}{n(n-1)}} \le e^{-200}$ if $T := 100n^2$.

Proof: Fix C. $c := |C|$. Let event $E_i :=$ C survives the i-th round.

Observe that the min degree of $G \ge c$ (otherwise, that is the min-cut).

↳ # of edges in $G \ge \frac{nc}{2}$ (divide by 2 for double counting)

$Pr[\bar{E}_1] = \frac{\#\text{ of edges in } C}{\#\text{ of all edges}} \le \frac{c}{nc/2} = \frac{2}{n} \rightarrow Pr[E_1] \ge 1 - \frac{2}{n}$.

$Pr[\bar{E}_2 | E_1] \le \frac{c}{(n-1)c/2} = \frac{2}{n-1} \rightarrow Pr[E_2 | E_1] \ge 1 - \frac{2}{n-1}$.

$Pr[\bigcap_{i=1}^{n-2} E_i] = Pr[E_1] \times Pr[E_2 | E_1] \times \cdots \times Pr[E_{n-2} | E_1 \cap \ldots \cap E_{n-3}]$.

$\ge \left(1 - \frac{2}{n}\right) \times \left(1 - \frac{2}{n-1}\right) \times \cdots \times \left(1 - \frac{2}{3}\right)$

$= \left(\frac{\cancel{n-2}}{n}\right)\left(\frac{\cancel{n-3}}{n-1}\right)\left(\frac{\cancel{n-4}}{\cancel{n-2}}\right)\left(\frac{\cancel{n-5}}{\cancel{n-3}}\right) \cdots \left(\frac{\cancel{2}}{\cancel{4}}\right)\left(\frac{1}{\cancel{3}}\right) = \frac{2}{n(n-1)}$. //

Runtime: $O(n^2)$ per output, $O(n^2)$ iterations $\rightarrow O(n^4)$ time

* Karger optimizes the procedure upto $O(n^2 \log n)$.

Random Variables & Expectation

Def) Random Variable: a function $X : \Omega \rightarrow \mathbb{R}$ on a prob. space Ω.

↳ $Pr[X = a] = \sum_{\omega \in \Omega \text{ s.t. } X(\omega) = a} Pr[\omega]$. $X \perp Y$ if $Pr[X = a] \perp Pr[Y = b]$ $\forall a, b$.

ex) $X :=$ Sum of 2 dice rolls. $Pr[X = 2] = \frac{1}{36}$, $Pr[X = 4] = \frac{3}{36} = \frac{1}{12}$.

Def) Expectation: for a RV X, $E[X] = \sum_a a \cdot Pr[X=a]$.

ex) $\Omega = \mathbb{N} \setminus \{0\}$, $Pr[X=i] = \frac{1}{c} \cdot \frac{1}{i^2} \rightarrow E[X] = \frac{1}{c} \cdot \sum_i \frac{1}{i} \rightarrow \infty$

Linearity of Expectation: $\overbrace{\forall \text{RV } X, Y}^{\text{does not need to be } X \perp Y!}$, $E[X+Y] = E[X] + E[Y]$

* $E[XY] = E[X]E[Y]$ only if $X \perp Y$!

ex) $E[\overbrace{\text{sum of 2 dice rolls}}^{X}] = E[X_1] + E[X_2]$ where $X_1 = X_2 :=$ value of a roll

$\rightarrow E[X] = 2 \cdot E[X_1] = 2 \cdot \frac{7}{2} = 7$. //

ex) Balls & Bins: m balls, n bins, $X :=$ # of empty bins.

$X = X_1 + X_2 + \cdots + X_n$ wher $X_i := \mathbb{1}\{\text{bin } i \text{ is empty}\}$. $\rightarrow$ indicator RV

$\rightarrow E[X] = E[X_1 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$.

$\forall i \in [n]$, $E[X_i] = \left(\frac{n-1}{n}\right)^m = \left(1 - \frac{1}{n}\right)^m \Rightarrow \underline{E[X] = n\left(1 - \frac{1}{n}\right)^m}$. //

* if $n = m$ is large, $E[X] = n\left(1 - \frac{1}{n}\right)^n \sim n/e$.

ex) $X :=$ # of fixed points in a random permutation π.

$X = \sum_{i=1}^{n} X_i$ where $X_i := \mathbb{1}\{i \text{ is a fixed point}\}$.

$\rightarrow E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{n} = n \cdot \frac{1}{n} = 1$. //

ex) $Y :=$ # of cycles in a random perm. π.

Fix an element i. Let $L_i :=$ length of cycle containing i.

Claim: $Pr[L_i = k] = \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{1}{n-1}\right) \times \cdots \times \left(1 - \frac{1}{n-k+2}\right) \times \frac{1}{n-k+1} = \frac{1}{n}$.

$\because$ until the k-th hop, we have $1, 2, \ldots, (k-1)$ forbidden elements. On the k-th hop, we have to pick exactly i out of $(n-k+1)$ elements.

$\rightarrow \underbrace{Y = \sum_{i=1}^{n} \frac{1}{L_i}}$ $\rightarrow E[Y] = \sum_{i=1}^{n} E\left[\frac{1}{L_i}\right] = \sum_{i=1}^{n}\sum_{k=1}^{n}\left[\frac{1}{n}\cdot\frac{1}{k}\right] = \sum_{i=1}^{n}\frac{H_n}{n} \sim \underline{\log n + C}$.

$\hookrightarrow$ this way, each cycle contributes exactly 1 to Y.

Binomial Distribution

Toss a coin with heads probability p, n times. $X :=$ # of heads.

$\underline{\Pr[X=k] = \binom{n}{k} p^k (1-p)^{n-k}}$ for $0 \le k \le n$.

$E[X] = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k}$, or $X = \sum_{i=1}^{n} X_i$ where $X_i := \mathbb{1}\{i\text{th coin is heads}\}$.

$\rightarrow \underline{E[X]} = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = \underline{np}$.

Geometric Distribution

Same experiment, but n is unbounded. $Y :=$ # of flips until first heads.

$\underline{\Pr[Y=k] = (1-p)^{k-1} p}$ for all $1 \le k$.

$E[Y] = \sum_{k=1}^{\infty} k(1-p)^k p$. Let $S := \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{p}$. $\frac{dS}{dp} = -\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$.

Since $E[Y] = p \cdot \frac{dS}{dp}$, $\underline{E[Y] = \frac{1}{p}}$.

Alteratively, we claim that $E[Y] = \sum_{k=1}^{\infty} \Pr[Y \ge k]$. (tail sum)

Then, $E[Y] = \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{1}{p}$.

Proof of Claim: $E[X] = \sum_{k=1}^{\infty} k \cdot Pr[X=k] =$

$Pr[X=1] +$

$Pr[X=2] + Pr[X=2] +$

$Pr[X=3] + Pr[X=3] + Pr[X=3] +$

$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \cdots$

$Pr[X \geq 1] + Pr[X \geq 2] + \cdots = \sum_{k=1}^{\infty} Pr[X \geq k]$. //

Coupon Collection

n coupons, $X :=$ # of boxes purchased until we have n distinct coupons

$X = X_1 + X_2 + \cdots + X_n$ where $X_i :=$ # of boxes until a new coupon from last

Trivially, $X_1 = 1$. $X_i \sim Geo\left(\frac{n-i+1}{n}\right)$ to account for repeated pulls.

$\rightarrow E[X_i] = \frac{n}{n-i+1} \rightarrow E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot \sum_{j=1}^{n} \frac{1}{j} = \underline{\underline{n \cdot H_n}}$.

Application) Quicksort: pick a pivot $x^* \in \{X_i\}_n$ u.a.r. compare all elements with x^*, and partition them into $X_{<x^*}$, X^*, $X_{>x^*}$. recursively apply Quick sort to $X_{<x^*}$ and $X_{>x^*}$ and concatenate.

Claim: $E[T_n] = 2n\log n + \Theta(n)$ where $T_n :=$ runtime of QS on $|A| = n$.

Proof of Claim: Let $y_1 < y_2 < \cdots < y_n$ be the ordering of the <u>sorted</u> list.

$Z :=$ # of comparisons made by QS on input $X_1, \ldots, X_n$.

$Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Z_{ij}$ where $Z_{ij} := \mathbb{1}\{$the pair (Y_i, Y_j) is compared by QS$\}$.

Key observation: (Y_i, Y_j) is compared iff either Y_i or Y_j is the first pivot selected in the set $\{Y_i, Y_{i+1}, \ldots, Y_j\}$. This holds since if any other element is selected, Y_i and Y_j will be put in opposite bins for the next QS iterations, never to be compared. //

$\rightarrow Pr[(Y_i, Y_j) \text{ is compared}] = \frac{2}{|\{Y_i, \ldots, Y_j\}|} = \underline{\frac{2}{j-i+1}}$.

$\rightarrow E[Z] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$. reparameterizing $k = j-i+1$, $E[Z] = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k}$.

Claim: $E[Z] = \sum_{k=2}^{n} (n+1-k)\frac{2}{k} = 2(n+1)\sum_{k=1}^{n} \frac{1}{k} - 2(n-1) \sim \underline{2n\log n + \Theta(n)}$. //

* This analysis still holds for a deterministic QS for a random permutation.

Concentration Inequalities

How likely is the RV X to deviate a lot from its mean $E[X]$?

Def) Moments: ith moment of $X = E[X^i]$

Def) Markov's Inequality: $\forall X \geq 0$, $\underline{Pr[X \geq \alpha] \leq \frac{1}{\alpha} E[X]}$.

Proof: Assume that $Pr[X \geq \alpha] > \frac{1}{\alpha} E[X]$. $E[X] = \sum_{k \geq 0} k Pr[X=k] \geq \alpha \cdot Pr[X \geq \alpha] > E[X]$. Contradiction. //

ex) $X := Bi(n, \frac{1}{2})$. $Pr[X \geq \frac{3n}{4}] \leq \frac{4}{3n} \cdot E[X] = \frac{4}{3n} \cdot \frac{n}{2} = \boxed{\frac{2}{3}}$

ex) $Y := \#$ of fixed points in π. $E[Y]=1$. $\Pr[Y \geq 10] \leq \boxed{\frac{1}{10}}$ //

Def) Variance: $E[(X-E[X])^2] = \underbrace{E[X^2]-E[X]^2}\geq 0$.

Def) Standard Deviation: $\sigma(X) = \sqrt{Var(X)}$.

Fact: For any RV X,Y, $Var(X+Y) = Var(X)+Var(Y)+Cov(X,Y)$

where $\underbrace{Cov(X,Y) := E[(X-E[X])(Y-E[Y])]}$.

Justification: Expand $Var(X+Y) = E[(X+Y-E[X]-E[Y])^2)$.

Fact: If $X \perp Y$, $Cov(X,Y)=0$. $\because E[XY]-E[X]E[Y]=0$.

$\Rightarrow$ For $X \perp Y$, $Var(X+Y) = Var(X)+Var(Y)$!

Examples) $X \sim Bi(n,p)$. $E[X]=np$. $Var(X)$?

$Var(X) = Var(\sum_{i=1}^{n} X_i)$ where $X_i := \mathbb{1}\{\text{heads for ith flip}\}$.

$\rightarrow Var(X) = \sum_{i=1}^{n} Var(X_i)$ because each flip is independent.

$Var(X_i) = E[X_i^2]-E[X]^2 = (p)-(p^2) = p(1-p)$.

$\Rightarrow \underbrace{Var(X)} = n\cdot(p(1-p)) = np(1-p) \leq np$ //

$X := \#$ of fixed points in a random perm. π

$X = \sum_{i=1}^{n} X_i$ where $X_i = \mathbb{1}\{i \text{ is a fixed point}\}$

$E[X] = n \times \frac{1}{n} = 1$. $Var(X)$? $Var(X) = E[X^2] - \underbrace{E[X]^2}_{\rightarrow 1}$

$$E[X^2] = E\Big[\Big(\sum_{i=1}^{n} X_i\Big)^2\Big] = \underbrace{\sum_{i=1}^{n} E[X_i^2]}_{\to 1} + \sum_{i\neq j} E[X_i X_j]$$

$$E[X_i X_j] = \Pr[i \text{ and } j \text{ are both fixed points}] = \left(\frac{1}{n}\right)\cdot\left(\frac{1}{n-1}\right)$$

$$\to \sum_{i\neq j}^{n} E[X_i X_j] = \sum_{i\neq j}^{n} 2\binom{n}{2}\frac{1}{n(n-1)} = 1 \Rightarrow \mathrm{Var}(X) = 1 + 1 - 1 = \underline{1} \,//$$

$X \sim \mathrm{Geo}(p)$. $E[X] = \frac{1}{p}$. $\mathrm{Var}(X) = E[X^2] - \underbrace{E[X]^2}_{\to \frac{1}{p^2}}$

$E[X^2] = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p$. $\to$ can be calculated with 2nd derivative

$$E[X^2] = E[X^2 \mid 1\text{st flip is tails}]\cdot \Pr[1\text{st flip is tails}] + E[X^2 \mid 1\text{st fip is heads}]\cdot \Pr[1\text{st flip is heads}]$$

$$= (1-p)E[X^2 \mid 1\text{st flip is tails}] + p\cdot \underbrace{1}_{\to \text{experiment terminates}}$$

$$= (1-p)E[(Z+1)^2] + p \quad \text{where } Z \sim \mathrm{Geo}(p)$$

$$= (1-p)\left[E[X^2] + 2E[X] + 1\right) + p \quad \text{since } Z = X$$

$$\to pE[X^2] = 2\left(\frac{1-p}{p}\right) + (1-p) + p = \frac{2-2p+p}{p} \to E[X^2] = \frac{2-p}{p^2}$$

$$\Rightarrow \underline{\mathrm{Var}(X)} = \left(\frac{2-p}{p^2}\right) - \left(\frac{1}{p^2}\right) = \frac{1-p}{p^2} \leq \frac{1}{p^2} \,//$$

Def) Chebeshev's Inequality: $\Pr[|X - E[X]| \geq \alpha] \leq \frac{\mathrm{Var}(X)}{\alpha^2}$

Proof: $\Pr[|X - E[X]| \geq \alpha] = \Pr[(|X - E[X]|)^2 \geq \alpha^2]$.

Let $Y := (|X - E[X]|)^2$, which is nonnegative.

$$\to \Pr[Y \geq \alpha^2] \leq \frac{E[Y]}{\alpha^2} = \underline{\frac{\mathrm{Var}(X)}{\alpha^2}} \,. //$$

Observations) $\Pr[|X-E[X]| \geq \beta\sigma(X)] \leq \frac{Var(X)}{\beta^2\sigma^2(X)} = \frac{1}{\beta^2}$.

$\Pr[|X-E[X]| \geq \gamma \cdot E[X]] \leq \frac{Var(X)}{\gamma^2 E[X]^2}$. $\frac{Var(X)}{E[X]^2} \rightsquigarrow$ critical ratio,

if the critical ratio doesn't "blow up" deviations approach $\emptyset$ as $\gamma \to \infty$.

Examples) $X \sim Bi(n, 1/2)$. $E[X] = \frac{n}{2}$, $Var(X) = \frac{n}{4}$.

$\Pr[X \geq \frac{3n}{4}] \leq \Pr[|X-E[X]| \geq \frac{n}{4}] \leq \frac{Var(X)}{(n/4)^2} = \frac{4}{n}$.

$X :=$ # of fixed points in π. $E[X] = 1$, $Var(X) = 1$.

$\Pr[X \geq 10] \leq \Pr[|X-E[X]| \geq 9] \leq \frac{Var(X)}{9^2} = 1/81$.

Application) Randomized Median Finding Algorithm

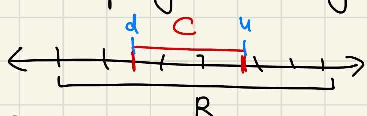
Median: $\lceil n/2 \rceil$-th element in a sorted list (assume unique elements)

Obviously, sorting in $O(n \log n)$ solves finding the median

Can we find the median without sorting the entire list?

Proposal: Simple $O(n)$ randomized algorithm via sampling/sketching

↳ pick a random sample of $n^{3/4}$ elements, R.



Sort R, then find the "central section" C of R. $(O(n^{3/4} \log(n^{3/4})))$

Hopefully, the median m of S is contained in $[d, u]$ while

C is narrow enough to only have $o(n)$ elements!

Given that $m \in [d, u]$, the median can be located by sorting C and get the $(\lceil n/2 \rceil - l_d + 1)$-th element where $l_d :=$ # of elements of S smaller than d, which is in $O(n)$ time.

Pseudocode: 1) Pick random sample $R \subseteq S$ of size $n^{3/4}$ (with repl.)

2) Sort R.

3) Locate interval of size $[\frac{1}{2}n^{3/4} - \sqrt{n}, \frac{1}{2}n^{3/4} + \sqrt{n}]$ at the center of R.
Call the bounding elements u and d, respectively.

4) Find all elements $C \subseteq S$ that lie between $[d, u]$. Find l_d, the # of elements in S less than d (l_u is for elements $> u$)

5) Sort C and output the $(\lceil n/2 \rceil - l_d + 1)$-th element in C.

* If $|C| \geq 4n^{3/4}$, then FAIL. ($\mathcal{E}_c := |C| \geq 4n^{3/4}$)

* Also, if m doesn't lie between $[d, u]$, algo fails.

$\rightarrow m < d$

Claim: Let $\mathcal{E}_d := |\{r \in R \mid r \leq m\}| \leq \frac{1}{2}n^{3/4} - \sqrt{n}$,

$\rightarrow m > u$

$\mathcal{E}_u := |\{r \in R \mid r \geq m\}| \leq \frac{1}{2}n^{3/4} + \sqrt{n}$. If neither $\mathcal{E}_d$ nor $\mathcal{E}_u$ happens, then the second mode of failure doesn't happen.

$\hookrightarrow \Pr[\text{FAIL}] = \Pr[\mathcal{E}_u \cup \mathcal{E}_d \cup \mathcal{E}_c] \leq \Pr[\mathcal{E}_u] + \Pr[\mathcal{E}_d] + \Pr[\mathcal{E}_c]$

Analysis of $\mathcal{E}_d$: Let $X :=$ # of elements in R that are $\leq m$.

$X \sim Bi(n^{3/4}, 1/2)$ since X_i are basically coin flips of $p = 1/2$.

$\rightarrow E[X] = \frac{1}{2}n^{3/4}$. $Pr[\mathcal{E}_d] \leq Pr[|X - E[X]| \geq \sqrt{n}] \leq \frac{Var(X)}{n}$

$= \frac{(1/4)n^{3/4}}{n} = \frac{1}{4n^{1/4}}$. Same applies to $\mathcal{E}_u$.

Analysis of $\mathcal{E}_c$: $\mathcal{E}_c' :=$ $\geq 2n^{3/4}$ elements of $C \leq m$,

$\mathcal{E}_c'' := \geq 2n^{3/4}$ " $C \geq m$, $\mathcal{E}_c = \mathcal{E}_c' \cup \mathcal{E}_c''$.

$\mathcal{E}_c' \Rightarrow$ rank of d in $S \leq \frac{1}{2}n - 2n^{3/4} \rightarrow R$ must have at least $\frac{1}{2}n^{3/4} - \sqrt{n}$ elements within the first $\frac{1}{2}n - 2n^{3/4}$ elements of S.

Let $Y :=$ # of elements of R that lie in first $\frac{1}{2}n - 2n^{3/4}$ elements of S.

$\rightarrow Y \sim Bin(n^{3/4}, \frac{\frac{1}{2}n - 2n^{3/4}}{n}) = Bin(n^{3/4}, \frac{1}{2} - \frac{2}{n^{1/4}})$.

$\rightarrow E[Y] = \frac{n^{3/4}}{2} - 2\sqrt{n}$, $Var(X) = n^{3/4}(\frac{1}{2} - \frac{2}{n^{1/4}})(\frac{1}{2} + \frac{2}{n^{1/4}})$

$= n^{3/4}(\frac{1}{4} - \frac{4}{n^{1/2}}) = \frac{n^{3/4}}{4} - 4n^{1/4} \leq \frac{1}{4}n^{3/4}$.

$\rightarrow Pr[\mathcal{E}_c'] = Pr[Y \geq \frac{1}{2}n^{3/4} - \sqrt{n}] \leq Pr[|Y - E[Y]| \geq \sqrt{n}] \leq \frac{Var(Y)}{n} = \frac{1}{4n^{1/4}}$

Same applies to $\mathcal{E}_c'' \rightarrow Pr[\mathcal{E}_c] \leq Pr[\mathcal{E}_c'] + Pr[\mathcal{E}_c'']$

$\Rightarrow Pr[FAIL] \leq Pr[\mathcal{E}_d] + Pr[\mathcal{E}_u] + P[\mathcal{E}_c'] + Pr[\mathcal{E}_c''] \leq 4 \cdot \frac{1}{4n^{1/4}} = \underline{\frac{1}{n^{1/4}}}$ //

Chernoff (Hoeffding) Bounds

Motivation) Consider $X_1, \dots, X_n$ of coin tosses of heads prob. p.

$\rightarrow X = \sum_{i=1}^{n} X_i \sim Bi(n, p) \rightarrow Pr[|X - E[X]| \geq \varepsilon np] \leq \underline{\frac{1}{\varepsilon^2 np}}$. <u>Can we do better?</u>

↳ Given that X_is are independent, Chernoff gives $\leq \exp(-cn)$ bound!

Def) Moment Generating Functions: $M_X(t) := E[e^{tX}]$. An MGF exists if $M_X(t)$ is finite in a small region $[-\delta, \delta]$ around 0.

Claim: $\forall k \geq 0$, the k-th moment of X is determined by $M_X^{(k)}(0)$.

"Proof": $M_X(t) = E[e^{tX}] = E[1 + tX + \frac{t^2X^2}{2!} + \cdots]$

$= 1 + E[tX] + \frac{t^2}{2!}E[X^2] + \cdots$

$\rightarrow M_X'(t) = E[X] + t(E[X^2] + \cdots) \rightarrow M_X'(0) = E[X]$.

Similarly, differentiating $M_X(t)$ k times yields $E[X^k]$ at $M_X(0)$.

Example) $X \sim Geo(p)$. $M_X(t) = E[e^{tX}] = \sum_{k=1}^{\infty} (1-p)^{k-1} p \cdot e^{tk}$

$= \frac{p}{1-p}\sum_{k=1}^{\infty}(1-p)^k e^{tk} = \underline{\left(\frac{p}{1-p}\right)\left(\frac{1}{1-(1-p)e^t} - 1\right)}$ //

$M_X'(t) = \frac{p}{1-p} \cdot \frac{1}{(1-(1-p)e^t)^2} \cdot (1-p)e^t = \left(\frac{p}{1-p}\right)\left(\frac{1}{p^2}\right)(1-p) = \frac{1}{p}$. //

$X_i \sim Ber(p)$. $M_{X_i}(t) = E[e^{tX_i}] = pe^t + (1-p)$.

Fact) For independent X, Y, $M_{X+Y}(t) = M_X(t) M_Y(t)$.

Proof) $M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}] \overset{\text{by independence.}}{=} E[e^{tX}] \cdot E[e^{tY}]$. //

$\rightarrow X \sim Bi(n, p) \rightarrow M_X(t) = \prod_{i=1}^{n} M_{X_i}(t) = (pe^t + (1-p))^n$. //

Fact) If $M_X(t)$ & $M_Y(t)$ exist and $M_X(t) = M_Y(t)$, then X and Y have the same distribution (MGF uniquely defines a RV)

Obs) $\Pr[X \geq \alpha] = \Pr[e^{tX} \geq e^{t\alpha}] \leq \frac{E[e^{tX}]}{e^{t\alpha}} = \frac{M_X(t)}{e^{t\alpha}}$ for $\forall t > 0$.

Def) Chernoff-Type Bounds: $\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}]$ $(t>0)$ $\leq \frac{E[e^{tX}]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$.

$\Pr[X \leq a] = \Pr[e^{tX} \geq e^{ta}]$ $(t < 0)$ $\leq \frac{E[e^{tX}]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$.

$X := \sum_{i=1}^{n} X_i$ where $X_i :=$ coin flips where $\Pr[X_i] = p_i$, $\mu = \sum_{i=1}^{n} E[X_i]$.

$\rightarrow \Pr[X \geq (1+\delta)\mu] \leq \frac{M_X(t)}{e^{t(1+\delta)\mu}}$ for any $t > 0$.

$M_X(t) = \prod_{i=1}^{n} (p_i e^t + (1-p_i)) = \prod_{i=1}^{n} (1 + p_i(e^t - 1)) \leq \prod_{i=1}^{n} \exp(p_i(e^t-1))$

$= \exp(\sum_{i=1}^{n} p_i(e^t - 1)) = \exp(\mu(e^t-1))$. $\rightarrow \frac{M_X(t)}{e^{t(1+\delta)\mu}} = \frac{\exp(\mu(e^t-1))}{\exp(t(1+\delta)\mu)}$.

Set $t = \ln(1+\delta)$. Then, $\frac{\exp(\mu\delta)}{(1+\delta)^{(1+\delta)\mu}} = \underline{\left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu}$. //

Claim: $\left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right)$. $\rightarrow$ $\underline{\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \leq \exp\left(\frac{-\delta^2}{2+\delta}\right)}$.

Taking logs, $\delta - (1+\delta)\ln(1+\delta) \leq -\frac{\delta^2}{2+\delta}$. Take the fact that $\ln(1+\delta) > \frac{\delta}{1+\delta/2}$ for $\delta \geq 0$ (simple calculus). $\rightarrow (1+\delta)\ln(1+\delta) \geq \delta + \frac{\delta^2}{2+\delta}$

$\rightarrow (1+\delta)\ln(1+\delta) > (1+\delta)\frac{\delta}{1+\delta/2} = \frac{2(1+\delta)\delta}{2+\delta} = \delta + \frac{\delta^2}{2+\delta}$. //

$\Rightarrow \underline{\Pr[X \geq (1+\delta)\mu] \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right)}$ (for all $\delta > 0$),

$\star$ $\underline{\leq \exp\left(-\frac{\delta^2}{3}\mu\right)}$ (for $0 < \delta \leq 1$).

Lower Tail Case: $\Pr[X \leq (1-\delta)\mu] \leq \frac{\exp(\mu(e^t - 1))}{\exp(t(1-\delta)\mu)}$.

Set $t = \ln(1-\delta) < 0$. $\rightarrow \frac{\exp(-\delta\mu)}{(1-\delta)^{(1-\delta)\mu}} = \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$

Claim: $\left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \leq \exp\left(-\frac{\delta^2\mu}{2}\right)$. $\rightarrow \frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq e^{-\delta^2/2}$. Taking logs,

$\delta + (1-\delta)\ln(1-\delta) \geq \frac{\delta^2}{2}$. Take the fact that $(1-\delta)\ln(1-\delta) \geq -\delta + \frac{\delta^2}{2}$,

the proof is trivial.

$\Rightarrow \underline{\Pr[X \leq (1-\delta)\mu] \leq \exp\left(\frac{-\delta^2\mu}{2}\right)}$ $\star$ (for $0 < \delta < 1$).

Corollary: $\underline{\Pr[|X-\mu| \geq \delta\mu] \leq 2\exp\left(\frac{\delta^2\mu}{3}\right)}$ $\star$ (for $0 < \delta < 1$).

Def) Hoeffding's Bound: Let $X_1, \ldots, X_n$ be indep. r.v. s.t. $E[X_i] = \mu_i$ and s.t. $a_i \leq X_i \leq b_i$ for some constants a_i, b_i.

Let $X := \sum_{i=1}^{n} X_i$, then $E[X] = \sum_{i=1}^{n} \mu_i$.

Then, $\underline{\Pr[|X-\mu| \geq \lambda] \leq 2\exp\left(-\frac{2\lambda^2}{\sum_i (b_i - a_i)^2}\right)}$. $\star$

e.g. If $\lambda = \delta\mu$, $\Pr[|X-\mu| \geq \delta\mu] \leq 2\exp\left(-\frac{2\delta^2\mu^2}{\sum_i (b_i - a_i)^2}\right)$.

if X_i is a 0-1 r.v., $2\exp\left(\frac{2\delta\mu^2}{n}\right) = 2\exp(2\delta\mu p)$.

Examples) Fair Coin $Bi(n, \frac{1}{2})$. $\mu = E[X] = \frac{n}{2}$, $Var(X) = \frac{n}{4}$.

Chernoff: $Pr[|X - n/2| \geq \delta \frac{n}{2}] \leq 2\exp(-\frac{\delta^2 n}{6})$.

1) $\delta n/2 = cn. \to 2\exp(-\frac{4c^2 n}{6}) = 2e(-\frac{2c^2 n}{6})$.

2) $\delta n/2 = c\sqrt{n}. \to 2\exp(-\frac{2c^2}{3})$.

2)' $\delta n/2 = c\sqrt{n \log n} \to 2\exp(-\frac{2c^2}{3}\log n) \leq 2n^{-2c^2/3}$

Comparison with Chebyshev: $Pr[|X - \mu| \geq \frac{\delta n}{2}] = \frac{Var(X)}{\delta^2 n^2/4} = \frac{1}{\delta^2 n}$.

↳ This gives a looser bound w.r.t. bounds on δ.

Goal: estimate p = proportion of Democrats in population.

↳ pick n people u.a.r. (with repl.) $X_i := \mathbb{1}\{i \text{ is Democrat}\}$. $X := \sum_{i=1}^{n} X_i$.

output estimate $\tilde{p} := \frac{X}{n}$. $E[\tilde{p}] = p$. We want n large enough s.t.

$Pr[|\tilde{p} - p| \geq \varepsilon p] \leq \delta \Rightarrow Pr[|X - \mu| \geq \varepsilon\mu] \leq \delta$.

Chernoff: $Pr[|X - \mu| \geq \varepsilon\mu] \leq 2\exp(-\frac{\varepsilon^2\mu}{3})$ (want $\leq \delta$)

$\to$ require $\frac{\varepsilon^2\mu}{3} \geq \ln(\frac{2}{\delta}) \to n \geq \frac{3}{\varepsilon^2 p}\ln(2/\delta)$

Assume $p \geq 1/4$, we want $\leq 1\%$ of absolute error, confidence 95%.

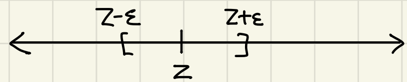
$\to \delta = 0.05$, $\varepsilon = 0.04 \Rightarrow n \geq 3(\frac{100}{4})^2 \times \frac{1}{1/4} \times \ln(40) \simeq 28{,}000$.

($\varepsilon = \frac{0.01}{p}$, if $p \geq 25\%$, $\varepsilon \leq \frac{0.01}{1/4} = 0.04$)

* could also sample with Chebyshev then compute expected number of trials with Chernoff. should give the same order of magnitude.

Ex) Algorithm outputs estimate $\hat{Z}$ of quantity Z.

Suppose $\Pr[|\hat{Z}-Z| \geq \varepsilon] \leq \frac{1}{4}$.



Perform t independent trials, and output the median. $\leq 1-$(more than half falls inside of $Z \pm \varepsilon$)

↳ $\Pr[\text{new algo is bad}] = \Pr[\text{median falls outside of } Z \pm \varepsilon] < \delta$

(if we set $t = O(\log(1/\delta))$ due to Chernoff)

Routing in a Hypercube

Hypercube: $V = \{0,1\}^n$, $(u,v) \in E$ iff u and v differ by only one bit

↳ $N = 2^n$ vertices, nN directed edges.

Setting: A packet at each vertex i & a permutation π of $\{0,1\}^n$.

Goal: Send each packet i to its destination $\pi(i)$.

Synchronous: One packet may move across a directed edge at any given timestep.

Queuing: FIFO (or any well-defined "lively" queuing protocol)

Ideally, the path of packet i should <u>only depend on i and $\pi(i)$</u>
↳ obliviousness

Theorem) For any <u>deterministic</u> oblivious scheme, $\exists \pi$ that requires $\Omega(\sqrt{N/n}) = \Omega(2^{n/2})$ steps.

Theorem) [Valiant, Brebner] There exists a simple oblivious scheme s.t.

for every π, $\Pr[\text{takes more than } 9n \text{ steps}] \leq 2^{-n} \to 0$ as $n \to \infty$.

The Scheme: Phase 1 $\to$ each packet i chooses a destination $\sigma(i)$ u.a.r. and proceeds to $\sigma(i)$ using a bit-fixing path.

Phase 2 $\to$ each packet proceeds from $\sigma(i)$ to $\pi(i)$ using a bit-fixing path. $*\sigma(i)$ is NOT a permutation!

Bit-Fixing Path: Correct any leftmost differing bit at all times

Let $D(i) :=$ delay suffered by packet i in Phase 1.

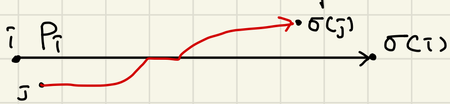
Claim A) $\forall$ packet i, $\Pr[D(i) \geq \frac{7}{2}n] \leq e^{-2n}$

Corollary) $\Pr[\max_i D(i) \geq \frac{7}{2}n] \leq 2^{-n}$

Proof: $\Pr[\exists i \text{ s.t. } D(i) \geq \frac{7}{2}n] \leq \sum_i \Pr[D(i) \geq \frac{7}{2}n] \leq \underbrace{2^n \cdot e^{-2n}}_{\leq 2^n \cdot 2^{-2n}} \leq 2^{-n}$.

Corollary) Whole Phase 1 terminates in $\leq 9/2\, n$ steps with probability $\geq (1 - 2^{-n})$.

Claim B) $\overbrace{D(i) \leq |S(i)|}^{\text{charge a different packet } j \in S(i) \text{ for every delay}}$ where $S(i) :=$ set of packets j whose paths intersect packet i.



Define $H_{ij} := \mathbb{1}\{j \in S(i)\}$. Then, $D(i) \leq \sum_{j \neq i} H_{ij}$.

Observe that for a fixed i, H_{ij} are independent. $\to$ Chernoff?

Fix the path $P_i = [e_1, e_2, \cdots, e_m]$. Focus on one edge $e \in P_i$.

Suppose edge e is: $(b_1, b_2, \cdots, b_{\ell-1}, a_\ell, a_{\ell+1}, \cdots, a_n) \rightarrow$

$(b_1, b_2, \cdots, b_{\ell-1}, b_\ell, a_{\ell+1}, \cdots, a_n)$ (flips the ℓ-th bit).

How many packets j have a path that uses this edge e?

↳ j must be of form $(*, *, \ldots, *, a_\ell, a_{\ell+1}, \cdots, a_n)$, so $\exists\, 2^{\ell-1}$ possible js.

↳ $\Pr[\text{such } j \text{ actually uses } e] = \Pr[\sigma(j) \text{ is of form } (b_1, \cdots, b_\ell, *, \cdots, *)]$

$= 2^{-\ell}$. $\Rightarrow E[\#\text{ of } j \text{ using edge } e] = 2^{\ell-1} \cdot 2^{-\ell} = 1/2$.

$\Rightarrow E[\sum_{j \neq i} H_{ij}] \leq n/2$. //

$\Pr[\sum_{j \neq i} H_{ij} \geq (1+\delta)\mu] \leq \exp(-\frac{\delta^2}{2+\delta}\mu)$. Plug in $\mu = n/2$, $\delta = 6$.

↳ $\Pr[\sum_{j \neq i} H_{ij} \geq 7/2\, n] \leq \exp(-\frac{36}{8} \cdot \frac{n}{2}) \leq \exp(-2n)$.

* Plugging in $\mu = n/2$ is justified by tuning $(1+\delta)\mu = A \Rightarrow \delta = \frac{A-\mu}{\mu}$

$\rightarrow$ bound becomes $\exp\left(-\frac{(A-\mu)^2/\mu^2}{2+(A-\mu)/\mu} \cdot \mu\right) = \exp\left(-\frac{(A-\mu)^2}{A+\mu}\right) \rightarrow$ is decreasing

$\rightarrow$ worst case is when μ is maximized $\Rightarrow \mu = n/2$ is a valid upper bound.

Taking the union bound for all 2^n vertices, $\Pr[\text{any delay} \geq 7/2\, n] \leq 2^{-n}$. //

Balls and Bins Method

m balls, n bins. each ball chooses a bin i.i.d. u.a.r.

Some Fundamental Questions:

1) Collisions: How big does m have to be before a collision?

↳ Birthday Problem: $n=365$ bins, m people $\to \Pr[\text{collision}] \geq \frac{1}{2}$ for $m=23$.

In general, $\Pr[\text{no collision}] = \prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right) \simeq \prod_{j=1}^{m-1}\exp(-j/n)$ ($1-x \approx e^{-x}$ for $x \ll 1$)

$\to \exp\left(-\frac{1}{n}\sum_{j=1}^{m-1} j\right) = \exp\left(-\frac{1}{n}\cdot\frac{m(m-1)}{2}\right) \simeq \exp\left(\frac{-m^2}{2n}\right) \to$ want $\leq \frac{1}{2}$

$\to \exp\left(-\frac{m^2}{2n}\right) = \frac{1}{2} \Rightarrow m = \sqrt{(2\ln 2)n} = O(\sqrt{n})$. // (For $n=365$, $m \simeq 22.5$!)

2) Empty Bins: What is the expected # of empty bins?

$X :=$ # of empty bins $= \sum_{i=1}^{n} X_i$ where $X_i := \mathbb{1}\{\text{bin } i \text{ is empty}\}$.

↳ $E[X_i] = \left(1-\frac{1}{n}\right)^m \to E[X] = n\left(1-\frac{1}{n}\right)^m \simeq ne^{-m/n}$.

↳ If $m = cn$, $E[X] \simeq ne^{-c}$.

3) Maximum Load: $X :=$ max load in any bin. Assume $\underline{m=n}$.

↳ W.h.p., $X \simeq \frac{\ln n}{\ln\ln n} + O(\text{lower order})$ (e.g. $m=n=10^6$, then $\frac{\ln n}{\ln\ln n} \simeq 5$).

Poisson Approximation: $X :=$ load in bin 1. $X \sim Bi(n, 1/n)$.

↳ X is $Bi(n,p)$ where $np = \lambda$ is constant ($\lambda = 1$ in this case) as $n \to \infty$.

Claim: $Bi(n, \frac{\lambda}{n})$ for any fixed λ converges in distribution to $Po(\lambda)$.

↳ i.e. $\forall k$, $Pr[X=k] \to Pr[Y=k]$ for $Y \sim Po(\lambda)$ $\left(\underline{Pr[Y=k] := \frac{\lambda^k}{k!} e^{-\lambda}}\right)$

↳ $Pr[X=k] = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!} e^{-\lambda} = Pr[Y=k]$

$$M_Y(t) = E[e^{Yt}] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \cdot e^{tk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

↳ $\underline{E[Y] = \lambda}$ by $M_Y'(0)$ and $E[Y^2] = \lambda^2 + \lambda \to \underline{Var(Y) = \lambda}$.

Fact: If $X \sim Po(\lambda_1)$, $Y \sim Po(\lambda_2)$, then $X+Y \sim Po(\lambda_1 + \lambda_2)$.

Proof: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$ //

Chernoff for $Po(\lambda)$: $Pr[X \geq (1+\delta)\lambda] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\lambda$ where $X \sim Po(\lambda)$.

Proof: $Pr[X \geq a] = Pr[e^{Xt} \geq e^{at}] \leq \frac{M_X(t)}{e^{at}}$ for $t > 0$.

$\to \frac{e^{\lambda(e^t-1)}}{e^{at}} = e^{\lambda(e^t-1) - ta}$. Choose $t = \ln(a/\lambda) > 0$ when $a > \lambda$

$\to \exp(a - \lambda - a\ln(\frac{a}{\lambda})) = \frac{e^{-\lambda}(e\lambda)^a}{a^a}$. Set $a = (1+\delta)\lambda \to \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\lambda$. //

Let $[X_1, \dots, X_n]$ be the real bin loads, $[Y_1, \dots, Y_n]$ be i.i.d. $Po(\lambda)$.

↳ observe that $E[Y_i] = E[X_i] = \lambda$.

Claim: Distribution of $[X_1, \dots, X_n]$ is the same as $[Y_1, \dots, Y_n]$ given that $\sum_{i=1}^{n} Y_i = m$ (sum of Y_i equals the # of balls).

Proof: For any $k_1,\dots,k_n$ s.t. $\sum k_i = m$, $\Pr[\vec{X}=\vec{k}] = \dfrac{\binom{m}{k_1,\dots,k_m}}{n^m}$

where $\binom{m}{k_1,\dots,k_n}$ is a multinomial coefficient of $\dfrac{m!}{k_1!\cdots k_n!} \rightarrow \left(\dfrac{m!}{k_1!(m-k_1)!}\cdot\dfrac{(m-k_1)!}{k_2!(m-k_1-k_2)!}\cdots\right)$

Now consider $\Pr[\vec{Y}=\vec{k} \mid \sum Y_i = m] = \prod_{i=1}^{n} \Pr[Y_i = k_i] \Big/ \Pr[\sum Y_i = m]$. ($Y_i$ is iid)

Observe that $\sum Y_i \sim Po(m)$. Then, $\prod_{i=1}^{n} e^{-m/n}\left(\frac{m}{n}\right)^{k_i}\cdot\frac{1}{k_i!} \Big/ e^{-m}\frac{m^m}{m!} = \dfrac{m!}{n^m k_1!\cdots k_n!}$. //

Corollary: Let $\mathcal{E}$ be any event depending only on the bin loads. Then,

$$\Pr_{real}[\mathcal{E}] \le e\sqrt{m}\,\Pr_{Po}[\mathcal{E}].$$

Proof: $\Pr_{Po}[\mathcal{E}(\vec{Y})] = \sum_{k=0}^{\infty} \Pr_{Po}[\mathcal{E}(\vec{Y}) \mid \sum Y_i = k]\Pr_{Po}[\sum Y_i = k] \ge$

$\Pr_{Po}[\mathcal{E}(\vec{Y}) \mid \sum Y_i = m]\Pr_{Po}[\sum Y_i = m] = \Pr_{Re}[\mathcal{E}(\vec{X})]\Pr_{Po}[\overbrace{\sum Y_i}^{\sim Po(m)} = m]$

$= \Pr_{Re}[\mathcal{E}(\vec{X})]\cdot e^{-m}\dfrac{m^m}{m!}$. Use Stirling upper bound of $m! \le e\sqrt{m}\left(\frac{m}{e}\right)^m$.

$\Rightarrow \Pr_{Po}[\mathcal{E}(\vec{Y})] \ge \Pr_{Re}[\mathcal{E}(\vec{X})]\cdot\dfrac{1}{e\sqrt{m}}$. //

* Improvement: If $\mathcal{E}$ is monotonically increasing / decreasing w.r.t. m (balls), then we can improve to $\Pr_{Re}[\mathcal{E}] \le 2\cdot\Pr_{Po}[\mathcal{E}]$. (take the entire tail)

Proof of Maximum Load: Set $m=n$ for ease of calculation. Then, max load $\sim \dfrac{\ln n}{\ln\ln n}$ w.h.p. Concretely, $\mathcal{E}_1 :=$ some bin contains $\ge \dfrac{(1+\varepsilon)\ln n}{\ln\ln n}$ balls, and $\mathcal{E}_2 :=$ no bin contains $\le \dfrac{(1-\varepsilon)\ln n}{\ln\ln n}$ balls, then $\Pr[\mathcal{E}_1]$ and $\Pr[\mathcal{E}_2] \to 0$ as $n \to \infty$. (sufficient to prove that $\Pr_{Po}[\mathcal{E}_1] \to 0$ and $\Pr_{Po}[\mathcal{E}_2] \to 0$ since $\mathcal{E}_1$ & $\mathcal{E}_2$ are monotone)

Define $P_k := \Pr_{Po}[Y_1 \geq k]$ where $Y_1 \sim Po(1)$. $\rightarrow P_k = \sum_{j \geq k} e^{-1} \frac{1}{j!}$.

Claim: $\frac{1}{ek!} \leq P_k \leq \frac{1}{k!}$. Proof: $P_k = \frac{1}{e} \sum_{j \geq k} \frac{1}{j!} \geq \frac{1}{e} \cdot \frac{1}{k!}$, $P_k = \frac{1}{ek!} \underbrace{\left(1 + \frac{1}{k+1} + \frac{1}{(k+1)(k+2)} \cdots\right)}_{\leq e}$.

$\mathcal{E}_1$ case: $\Pr[\mathcal{E}_1] \leq n \cdot \underbrace{\Pr\left[Y_1 \geq \frac{(1+\varepsilon)\ln n}{\ln \ln n}\right]}$ (by union bound).

Then, consider P_k where $k = \frac{(1+\varepsilon)\ln n}{\ln\ln n} \rightarrow \Pr[\mathcal{E}_1] \leq \frac{n}{k!}$

$\ln(P_k) \leq -\ln(k!) \sim -k\ln(k!) = -\frac{(1+\varepsilon)\ln n(\ln(1+\varepsilon) + \ln\ln n - \ln\ln\ln n)}{\ln \ln n}$

$\sim -(1+\varepsilon)\ln n \rightarrow P_k \sim n^{-(1+\varepsilon)} \rightarrow n \cdot P_k \sim n^{-\varepsilon}$ goes to 0 as $n \rightarrow \infty$. //

$\mathcal{E}_2$ case: $\Pr[\mathcal{E}_2] = (1-P_k)^n$ where $k = \frac{(1-\varepsilon)\ln n}{\ln\ln n}$ (bins are indep.!)

$\leq \left(1 - \frac{1}{ek!}\right)^n \leq e^{-n/ek!}$, so we want $\frac{n}{ek!} \rightarrow \infty$, i.e. $\ln\left(\frac{n}{ek!}\right) \rightarrow \infty$.

$\rightarrow \ln n - 1 - \ln(k!) \sim \ln n - 1 - k\ln k = \ln n - 1 - (1-\varepsilon)\ln n \sim \varepsilon \ln n$

goes to ∞ as $n \rightarrow \infty$. //

Hashing: big universe U and a small subset S. Any $x \in U$ gets stored in hash table T through perfectly random function $h: U \rightarrow T$.

↳ This maps to $S \rightarrow$ balls, $T \rightarrow$ bins. If we want no collisions, we need $|T| = \Omega(|S|^2)$. $E[\text{keys in a location}] = \frac{|S|}{|T|}$. Two issues:

1) Worst-case search time is bad (max load is $O(\ln|S|)$).

2) Random hash functions are unwieldy ("huge").

↳ # of bits to specify h is $O(\underbrace{|U|}_{\propto |U|}\log_2|T|)$!

↳ One solution is "double hashing".

→ We can use "pairwise independent" hash functions $O(|T|)$ size!

Random Graphs

Erdös–Rényi Model: $\mathcal{G}(n,p)$, $G \in \mathcal{G}(n,p) :=$ put down n (labeled) vertices, include each edge $\{i,j\}$ indep. w.p. p.

↳ $\Pr[G] = p^{|E|} \cdot (1-p)^{\binom{n}{2}-|E|}$ (prob. of a specific graph appearing)

$X := |E|$, then $X \sim Bi(\binom{n}{2}, p) \to E[X] = \binom{n}{2} p$.

$Y :=$ # of neighbors for a vertex, $Y \sim B(n-1, p) \to E[Y] = (n-1)p$.

$Z :=$ # of common neighbors for 2 vertices, $E[Z] = (n-2)p^2$.

$T :=$ # of triangles, $E[T] = \binom{n}{3} p^3 \sim \frac{1}{6} n^3 p^3$.

$I :=$ # of isolated (no neighbors) vertices, $E[I] = n(1-p)^{n-1}$.

* $p = \frac{1}{n}$ is a threshold for the "appearance" of triangles, i.e. if $p = o(\frac{1}{n})$, then $\Pr[G \text{ has a } \Delta] \to 0$ as $n \to \infty$, $p = \omega(\frac{1}{n})$, then $\to \infty$ as $n \to \infty$.

* $p = \frac{\ln n}{n}$ is a threshold for connectivity, Hamiltonian cycle, perfect matching...

Standard Regimes: $p = 1/2$ (uniform random graphs, dense)

$p = \frac{d}{n}$ (sparse random graphs, "real world"? avg. degree $= d$)

$p = c \frac{\ln n}{n}$ (sparse connected networks)

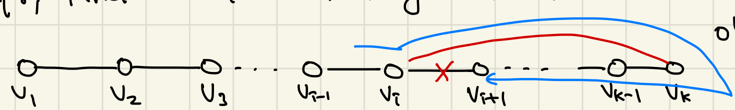
$\mathcal{G}(n,m)$ model is a random graph with n vertices, m edges at uniform prob.

↳ Roughly, $\mathcal{G}(n,m)$ behaves like $\mathcal{G}(n,p)$ with $m = \binom{n}{2} \cdot p$ (poissonization)

For sparse graphs, $\mathcal{G}_{reg}(n,d)$, random d-regular graphs, is also used.

Claim: Let $G \in \mathcal{G}(n,p)$ with $p \geq \frac{40 \ln n}{n}$. Then $\exists$ a poly time algorithm that, whp., finds a Hamiltonian cycle in G.

Idea:



(v_{i+1} becomes head) → rotate(P, v_i)

reverse(P) → v_1 becomes head, extend(P, u) → new vertex u becomes head

Assumption: G is represented as follows: (avoiding dependencies)

- each vertex v has a list unused(v) of neighbors, which includes each possible neighbor u independently w.p. q and in random order → $\geq \frac{20 \ln n}{n}$
- lists unused(v) are all independent of each other

Algorithm: 1) Start with $P = \{v_1\}$.

2) Repeat until $\underbrace{\text{HC is found}}_{(A)}$ or $\underbrace{\text{unused(current head)} = \emptyset}_{(B)}$ or $\underbrace{\geq 3n \ln n \text{ iterations}}_{(C)}$: →(ideally, Pr[(A)] dominates the other two)

- w.p. $1/n$, reverse(P).
- w.p. $\frac{|\text{used}(v_k)|}{n}$, pick edge (v_k, v_i) u.a.r. from used(v_k).
- w.p. $1 - \frac{1}{n} - \frac{|\text{used}(v_k)|}{n}$, pick first edge (v_k, u) from unused(v_k).

If $u=v_i$, then rotate(P, v_i). Else, extend(P, u).

3) If $k=|P|=n$ & $u=v_1$, output HC and halt.

↳ used(v) are all edges that were shifted from unused(v) after use.

Analysis: At any step t, let h_t be head(P). Provided unused$(h_t)\neq\emptyset$, then all vertices are equally likely to be the next head, i.e.

$\Pr[h_{t+1}=u \mid \text{history of algorithm}]=\frac{1}{n}\ \forall u\in V$.

Proof: Let $P=(v_1,\dots,v_k)$ s.t. $h_t=v_k$. $\Pr[h_{t+1}=v_1]=1/n$ via reverse(P).

If $u=v_{i+1}$ and $(v_k,v_i)\in$ used(v_k), $\Pr[h_{t+1}=v_{i+1}]=\frac{|\text{used}(v_k)|}{n}\cdot\frac{1}{|\text{used}(v_k)|}=\frac{1}{n}$.

If $u=v_{i+1}$ and $(v_k,v_i)\notin$ used(v_k) or $u\notin P$, $\Pr[h_{t+1}=v_{i+1}]=\frac{n-1-|\text{used}(v_k)|}{n}\cdot\frac{1}{n-1-|\text{used}(v_k)|}=\frac{1}{n}$. // (think unused$(v_k)$ as a black box outputting a "new" u)

⇒ By coupon collecting, w.p. $\geq 1-\frac{1}{n}$, $2n\ln n$ iterations will see all vertices.

($\because \Pr[u\notin P]=(1-\frac{1}{n})^{2n\ln n}\leq \exp(-2\ln n)=n^{2}\to \Pr[\exists u\notin P]\leq\frac{1}{n}$)

Then, ~$n\ln n$ iterations to close the cycle from $v_n\to v_1$. (let $q\geq\frac{20\ln n}{n}$)

Analysis 2: what is $\Pr[\text{unused}(v_k)\neq\emptyset]$? $|\text{unused}(v)|\sim Bi(n-1,q)$ at the start $\to E[|\text{unused}(v)|]=q(n-1)=\frac{20\ln n}{n}\cdot(n-1)\leq 19\ln n$.

Claim: $\Pr[\text{unused}(u)=\emptyset \text{ within } 3n\ln n \text{ iterations}]=\Pr[\mathcal{E}]\leq\frac{1}{n}$.

Proof: $\mathcal{E}' :=$ at least one vertex has $\leq 10\ln n$ edges in its initial unused list

$\mathcal{E}'' :=$ at least one vertex has $\geq 9 \ln n$ neighbors removed from its unused list

$\rightarrow \Pr[\mathcal{E}] \leq \Pr[\mathcal{E}'] + \Pr[\mathcal{E}'']$ since for $\mathcal{E}$ to happen, $\mathcal{E}'$ or $\mathcal{E}''$ must happen

$\mathcal{E}'$: $X = |\text{unused}(u)| \rightarrow E[X] \leq 19 \ln n$. $\Pr[X \leq 10 \ln n] = \Pr[X \leq (1-\frac{9}{19})\mu]$

$\leq \exp(-\frac{(9/19)^2}{2} \cdot 19 \ln n) = \exp(-\frac{81}{38} \ln n) \leq n^{-2} \rightarrow \Pr[\mathcal{E}'] \leq n \cdot \frac{1}{n^2} = \frac{1}{n}$.

$\mathcal{E}''$: $Y = \#$ of vertices removed from $\text{unused}(v) \preceq \text{Bi}(3n \ln n, \frac{1}{n}) \rightarrow E[Y] = 3 \ln n$

$\Pr[Y \geq 9 \ln n] \leq \Pr[Y \geq (1+2)\mu] \leq \exp(-\frac{2^2}{2+2} 3 \ln n) = n^{-3} \rightarrow \Pr[\mathcal{E}''] \leq \frac{1}{n^2}$.

Analysis 3: Preprocessing $G \in \mathcal{G}(n,p)$, $p \geq \frac{40 \ln n}{n}$. $\rightarrow$ let $q \in [0,1]$

satisfy $p = q(2-q)$, i.e. $q = 1 - \sqrt{1-p} \geq p/2 \Rightarrow p \geq \frac{40 \ln n}{n}$ ensures $q \geq \frac{20 \ln n}{n}$

$\forall (u,v) \in E(G)$: w.p. $\frac{q(1-q)}{q(2-q)}$ ($u \in \text{unused}(v)$, $v \notin \text{unused}(u)$), same for

vice versa, and $1 - \frac{2q(1-q)}{q(2-q)} = \frac{q^2}{q(2-q)}$ (both are in each other's list).

$\Pr[u \in \text{unused}(v)] = p\left[\frac{q(1-q)}{q(2-q)} + \frac{q^2}{q(2-q)}\right] = q$. $\rightarrow$ equal probability

$\Pr[u \in \text{unused}(v) \wedge v \in \text{unused}(u)] = p\left[\frac{q^2}{q(2-q)}\right] = q^2$. $\rightarrow$ i.i.d. //

(* can also analyze when both are not in each other's list, omitted here.)

The Probabilistic Method

Erdős [1947], Ramsey Numbers: $R_k :=$ smallest n s.t. in any 2-coloring of the edges of a complete graph K_n must include a monochromatic k-clique.

(party of n people must contain either a set of k mutual friends or k mutual strangers)

Claim: $R_3 = 6$.

Proof: For some vertex v, there are at least 3 edges of the same coloring. Then, we cannot avoid a monochromatic triangle. //

$R_4 = 18$, $R_5 \in [43, 48]$, ..., $R_{10} \in [798, 23556] \rightarrow$ intractable!

Theorem) [Erdös] $R_k > 2^{k/2}$.

Proof: Color edges of K_n u.a.r. $\Pr[\exists$ a (monochromatic) k-clique$] \leq$ # of k-cliques $\cdot$ $\Pr[$a k-cliques is monochromatic$] = \binom{n}{k} \cdot \frac{1}{2^{\binom{k}{2}+1}} \leq \frac{n^k}{k!} \cdot 2^{1-k^2/2+k/2}$. If $n = 2^{k/2}$, $\Pr[\varepsilon] \leq \frac{(2^{k/2})^k}{k!} \cdot 2^{1-k^2/2+k/2} = \frac{2^{k/2+1}}{k!} < 1$ for $\forall k \geq 3$.

$\rightarrow$ there must exist a coloring that contains <u>no monochromatic k-clique</u>!

Max-Cut Problem: Maximize crossing edges, NP-Hard.

Theorem) Any graph $G(V,E)$ contains a cut of size $\geq \frac{|E|}{2}$.

Proof: Pick a cut u.a.r. Let $X :=$ # of cut edges. $E[X] = \frac{|E|}{2}$.

$\Rightarrow \Pr[X \geq \frac{|E|}{2}] > 0 \Rightarrow \exists$ a cut of size at least $\frac{|E|}{2}$. //

Obs) $E[X] = \frac{1}{2} E[X \mid v_1 \in L] + \frac{1}{2} E[X \mid v_1 \in R]$. Similarly,

$E[X \mid (v_1, \dots, v_k) \text{ fixed}] = \frac{1}{2} E[X \mid (v_1, \dots, v_k), v_{k+1} \in L] + \frac{1}{2} E[X \mid (v_1, \dots, v_k), v_{k+1} \in R]$.

$\Rightarrow$ Take the choice that maximizes the next expected value, and by induction, we can always maintain $E[X \mid v_1 \dots v_k] \geq \frac{1}{2}|E|$. (<u>Conditional Expectation Method</u>)

MAX k-SAT: CNF boolean formula (ANDs of ORs) $\varphi \to$ assignment that maximizes the # of clauses satisfied.

Claim: Every k-SAT CNF formula has an assignment that satisfies at least $(1-\frac{1}{2^k})$ fraction of the clauses.

Proof: Pick a random assignment u.a.r. $X :=$ # of satisfied clauses.

$E[X] = \sum_{c \in C} E[X_c] = (1 - \frac{1}{2^{-k}}) \cdot m$ where $m :=$ # of clauses. //

↳ Also yields a deterministic greedy algorithm maximizing $E[X \mid \varphi_{x_1 \ldots x_k}]$.

Computing $E[X]$ is a bit more subtle, which depends on # of variables in clause.

Independent Set Problem: $G(V,E) \to$ largest indep. set in G (no edges btwn)

Theorem) $\forall G(V,E)$ of max degree d, $\exists$ an indep. set of size $\geq \frac{n}{d+1}$.

Proof: Assign a real value $r_v \in [0,1]$ to each vertex $v \in V$. v is a local minimum if $r_v \leq r_u \; \forall u \in N(v)$. Observe that the set of local minima is an independent set since local minima cannot be neighbors. Let

$E[X] = \sum_{v \in V} E[X_v] \geq n \cdot \frac{1}{d+1}$ since $E[X_v] \geq \frac{1}{d+1} \; \forall v \in V$. //

Theorem) Assume $m \geq n/2$. Such G contains an indep. set of size $\geq \frac{n^2}{4m}$.

Proof: Let $d := \frac{2m}{n}$ (average degree). Delete each vertex $v \in V$ w.p. $(1-\frac{1}{d}) = (1-\frac{n}{2m})$. For each remaining edge, remove it and one of its endpoints. Output the remaining vertices. → Produces an indep. set.

Let $X :=$ # of vertices remaining after deleting vertices → $E[X] = \frac{n}{d}$.

$Y :=$ # of edges remaining after deleting vertices → $E[Y] = \frac{nd}{2} \cdot \frac{1}{d^2} = \frac{n}{2d}$.

$E[\text{size of indep. set}] = E[X-Y] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d} = \frac{n^2}{4m}$. //

Graph Crossing Number: $c(G) :=$ minimum # of crossings in a planar drawing of G.

Euler's Formula: If G is a connected planar, $m \leq 3n - 6$.

Claim: $\forall$ connected G, $c(G) \geq m - 3n + 6$.

Proof: Take an optimal embedding of G (# of crossing $= c(G)$). Make the drawing planar by adding vertex at every crossing. $n \to n+c$. $m \to m+2c$.

By Euler, $m + 2c \leq 3(n+c) - 6 \to c \geq m - 3n + 6$. //

Theorem) Assume $m \geq 4n$. $c(G) \geq \frac{m^3}{64n^2}$.

Proof: Choose a random subgraph of G by picking each vertex w.p. p.

Let $n_p, m_p, c_p :=$ remaining vertices, edges, and crossings remaining. Then,

$c_p \geq m_p - 3n_p + 6 \to E[c_p] \geq E[m_p] - 3E[n_p] = np - 3mp^2$.

$E[c_p] = cp^4$ since it remains only if 4 vertices survive. → $c \geq \frac{m}{p^2} - \frac{3n}{p^3}$.

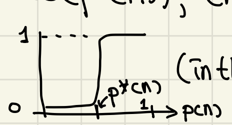
→ choose $p = \frac{4n}{m}$, then $c \geq \frac{m^3}{64n^2}$. //

Thresholds in Random Graphs

For $G \in \mathcal{G}(n,p)$, we can ask questions such as:

Is G connected? Contains a HC? Contain a subgraph, say, k-clique?

Def) Threshold (informal): value $p^*(n)$ s.t. if $p = o(p^*(n))$, then $\Pr[G \text{ has property}] \to 0$ as $n \to \infty$, and if $p = \omega(p^*(n))$, then $\Pr[G \text{ has property}] \to 1$ as $n \to \infty$.



(in the limit)

Ex) $X :=$ # of 4-cliques in $G(n,p)$. $X = \sum_c X_c$ where $X_c := \mathbb{1}\{c \text{ is a 4-clique}\}$

$E[X] = \sum_c E[X_c] = \binom{n}{4} p^6 = \Theta(n^4 p^6)$. $\to$ (i) If $p(n) = o(n^{-2/3})$, $E[X] \to 0$.

(ii) If $p(n) = \omega(n^{-2/3})$, $E[X] \to \infty$ (as $n \to \infty$). $p^*(n) = n^{-2/3}$?

↳ (i) $\Rightarrow \Pr[X > 0] = \Pr[X \geq 1] \leq \frac{E[X]}{1} \to 0$ as $n \to \infty$. ✓

↳ (ii) use Chebyshev (Second Moment Method) in this way:

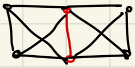
$$\Pr[X = 0] \leq \Pr[|X - E[X]| \geq E[X]] \leq \frac{Var(X)}{E[X]^2} \xrightarrow[\text{prove}]{\text{to}} 0.$$

Alternatively, since $Var(X) = E[X^2] - E[X]^2$, show that $\frac{E[X^2]}{E[X]^2} \to 1$.

$$Var(X) = Var(\sum_c X_c) = \underbrace{\sum_c Var(X_c)}_{(A)} + \underbrace{\sum_{c,c'} Cov(X_c, X_{c'})}_{(B)}.$$

(A) $Var(X_c) = E[X_c^2] - E[X_c]^2 \leq E[X_c^2] = E[X_c] \Rightarrow \sum_c Var(X_c) = E[X]$.

(B) $Cov(X_c, X_{c'})$, use case analysis:

- $|C \cap C'| \leq 1 \Rightarrow X_c \perp\!\!\!\perp X_{c'} \Rightarrow \mathrm{Cov}(X_c, X_{c'}) = \emptyset.$  $\to$ 11 edges

- $|C \cap C'| = 2 \Rightarrow \mathrm{Cov}(X_c, X_{c'}) = E[X_c X_{c'}] - E[X_c]E[X_{c'}] \leq E[X_c X_{c'}]$

 $= \Pr[C \text{ and } C' \text{ are both cliques}] = \underbrace{p^{11}}.$ # of such (C, C') pairs $= \underbrace{\Theta(n^6)}.$

9 edges
- $|C \cap C'| = 3 \Rightarrow \mathrm{Cov}(X_c, X_{c'}) \leq E[X_c X_{c'}] = \underbrace{p^9}$, # of pairs $= \underbrace{\Theta(n^5)}.$

$$\Rightarrow \frac{\mathrm{Var}(X)}{E[X]^2} \leq \frac{1}{E[X]^2}\left[E[X] + \Theta(n^6 p^{11}) + \Theta(n^5 p^9)\right] = \frac{\Theta(n^4 p^6 + n^6 p^{11} + n^5 p^9)}{\Theta(n^8 p^{12})}$$

$$= \frac{1}{\Theta(n^4 p^6)} + \frac{1}{\Theta(n^2 p)} + \frac{1}{\Theta(n^3 p^3)} \to \emptyset \text{ as } n \to \infty \text{ and } p(n) = \omega(n^{-2/3}). \quad //$$

What about when $p(n) = c \cdot p^*(n) = \Theta(p^*(n))$?

For 4-cliques, if $p(n) = c \cdot n^{-2/3}$, then $X \sim Po(c^6/24)$ as $n \to \infty$.

In the scale of $n^{-2/3}$, the threshold look like a smooth increasing function again $\to$ "coarse network"

given w/o proof.

For a threshold for a "giant" connected component of size $\Theta(n)$, $p^*(n) = \frac{1}{n}$.

If $p(n) = \frac{c}{n}$, then: 1) if $c < 1$, largest component has size $\Theta(\ln n)$, 2) if $c > 1$, $\Theta(n)$, 3) if $c = 1$, $\Theta(n^{2/3})$. $\to$ "sharp threshold"!

Theorem) Every monotone graph property has a (not necessarily sharp) threshold.

Does G contain a fixed subgraph H of v vertices, e edges?

$\hookrightarrow E[\#\text{ of copies of } H] = \binom{n}{v} p^e = \Theta(n^v p^e) \to p^*(n) = n^{-v/e}$?

↳ This only holds when edge density of $H \geq$ edge density of subgraph of H.

In fact, $p^*(n) = n^{-1/d}$ where $d :=$ max edge density of subgraph of H.

For many properties, $p^*(n) = \frac{\ln n}{n}$, such as isolated vertex, HC, connected... (in HW!)

$X :=$ # of HC, $E[X] = \frac{1}{2}(n-1)!\,p^n \rightarrow p^* = \Theta(\frac{1}{n})$, although real threshold is $\frac{\ln n}{n}$.

For $G \in G(n, \frac{1}{2})$ (dense network), we can ask questions such as:

Largest k-clique in G? $\rightarrow E[X_k] = \binom{n}{k} 2^{-\binom{k}{2}}$ crosses 1 at $k = k^* \sim 2\log_2 n$

↳ For $k \leq k^*(n) - 2$, $\Pr[G$ contains k-clique$] \rightarrow 1$, and for $k \geq k^*(n) + 2$,

$\Pr[G$ contains k-clique$] \rightarrow 0$ as $n \rightarrow \infty$, i.e. if $n = 1000$, largest

k-clique is size 15 or 16 w.h.p.

Challenge: Algorithm that finds clique of size $k \geq (1+\varepsilon)\log_2 n$?

Pairwise Independence

Def) Pairwise Independence: Family of RV $\{X_1, \ldots, X_n\}$ s.t. $\forall i \neq j \in [n]$,

$\forall x, y$, $\Pr[X_i = x \cap X_j = y] = \Pr[X_i = x] \cdot \Pr[X_j = y]$.

Claim) We can obtain $2^b - 1$ pairwise independent bits given only

b mutually independent bits $\{X_1, \ldots, X_b\}$.

Proof: For each $2^b - 1$ nonempty subsets of b bits $S_i := \{X_\alpha, \ldots\}$, define bit $y_i := \sum_{x \in S_i} X \pmod 2$. Then, y_i is uniform because we can use deferred decisions, $y_i = \left(\sum_{x \in S_i \setminus \{X_k\}} X\right) + X_k$, so $y_i = 0$ or 1 w.p. ½ each.

$y_j \neq y_k$ are PWI because $\exists \bar{X}$ s.t. $\bar{x} \notin S_j \wedge \bar{x} \in S_k$ (WLOG). Fix all other bits, and consider $\Pr[y_j = c \wedge y_k = d] = \Pr[y_k = d \mid y_j = c] \cdot \Pr[y_j = c]$ $= \frac{1}{2} \cdot \frac{1}{2}$ by deferred decisions again. //

Max Cut Revisited: $G(V,E) \to$ Find a maximum cut in G.

↳ $\exists$ a cut with $\geq \frac{|E|}{2}$ edges ($\because E[\text{size of cut}] = \frac{|E|}{2}$ when randomized).

New Challenge: Construct the cut using only PWI bits $\{y_1, \ldots, y_n\}$?

$X := \sum_e X_e$ where $X_e := \mathbb{1}\{y_{e_1} \neq y_{e_2}\}$, and y_v is a membership in 0-set or 1-set for vertex v. By this construction, we only need $\log_2 n$ MI bits.

Now, we can actually derandomize the algorithm by enumerating the entire $2^{\log_2 n} = n$ prob. space to get a $O(n(n+m)) = O(nm)$ determ. algorithm!

PWI RVs under (mod p), i.e. under field $\{0, 1, \ldots, p-1\}$ where p is prime: $\mathbb{F}(p)$

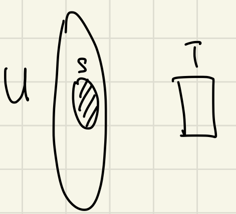
Let X_1, X_2 be uniform indep. RVs in $\mathbb{F}(p)$. Define $\underline{y_j := X_1 + j X_2 \pmod p}$.

Claim) y_j is uniform & PWI.

Proof: $\Pr[y_j = k] = \Pr[X_1 = k - jX_2]$. Fix RHS, then $\Pr[X_i = y]$ is uniformly $\frac{1}{p}$. $\Pr[y_i = k \wedge y_j = l] = \Pr[X_1 + iX_2 = k \wedge X_1 + jX_2 = l]$.
The explicit solution is a unique pair (X_1, X_2) $(X_2 = (l-k)(j-i)^{-1}, X_1 = \sim)$
$\Rightarrow \Pr[y_i = k \wedge y_j = l] = \frac{1}{p^2}$ since there is one pair out of p^2 pairs. //

Chebyshev Revisited: Still holds even if the X_is in $X = \sum_i X_i$ are only PWI, as in $Var(X) = \sum_i Var(X_i)$ since $Cov(X_i, X_j)$ are all $\emptyset$.

Universal Hash Functions: Universe U, Hash Table T. $|U| = M$. We want



$n = |T| \simeq |S| = n$ where we store $S \subseteq U$. Hash functions $h: U \to T$ exist, and we want to pick any h u.a.r. But # of bits to write down hs $\geq \log_2(n^M) = O(\underbrace{M}_{\text{bad!}} \log n)$

Def) 2-Universal: A family $\mathcal{H}$ of hash function $h: U \to T$ s.t. $\forall x_1 \neq x_2 \in U$, $\Pr_{h \sim \mathcal{H}}[h(x_1) = h(x_2)] \leq \frac{1}{n}$. ($\mathcal{H}$ is strongly 2-universal if $\Pr[h(x_1) = y_1 \wedge h(x_2) = y_2] = \frac{1}{n^2}$.)

Claim) Let $U = \{0, 1, \dots, (M-1)\}$, $T = \{0, \dots, (n-1)\}$. Let $p \gg M$ be prime.
Define $h_{a,b}(x) := ((ax+b) \bmod p) \bmod n$ and $\mathcal{H} := \{h_{a,b} \mid \underbrace{1 \leq a \leq (p-1), 0 \leq b \leq (p-1)}_{\text{u.a.r.}}\}$.
Then, $\mathcal{H}$ is 2-universal. require only $O(\log M)$ bits!

Proof) Suff. to prove that $\Pr_{a,b}[h_{a,b}(x_1) = h_{a,b}(x_2)] \le \frac{1}{n}$ $\forall x_1 \ne x_2$.

Consider the events $(ax_1 + b = u)$ and $(ax_2 + b = v)$. This has a unique solution for a and b since $x_1 \ne x_2$, so $u \ne v$ necessarily. How many pairs (u,v) have the property s.t. $u \ne v \pmod p$ but $u = v \pmod n$? $\to p(\lceil p/n \rceil - 1)$ $\le \frac{p(p-1)}{n}$. $\Rightarrow \Pr_{a,b}[h_{a,b}(x_1) = h_{a,b}(x_2)] \le \frac{p(p-1)/n}{p(p-1)} = 1/n$. //

Obs) Let $X :=$ # of collisions (pairs (x,y) s.t. $x \ne y$ and $h(x) = h(y)$). Assuming $\mathcal{H}$ is 2-universal, $E[X] \le \binom{m}{2}\frac{1}{n} \le \frac{m^2}{2n}$. If we set $n = m^2$, then $\Pr[X \ge 1] \le \frac{1}{2}$. If we set $n = m$, $E[X] \le n/2 \to$ max load $= O(\sqrt{n})$?

(For fully indep. hash functions, max load $\le \frac{\ln n}{\ln \ln n}$ w.h.p.)

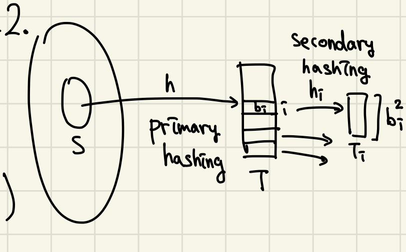
Def) Perfect Hashing: Assume static dictionary (no additional data).

Claim) For a static dictionary $U \to T$, we can achieve no collisions ($O(1)$ search) with $|T| \le 5|S|$. $\to \underbrace{|S|}_{\text{prim.}} + \underbrace{4|S|}_{\text{second.}}$

$\sum_i |\text{table}(h_i)| = O(\sum_i b_i^2)$

↳ Secondary hash functions h_i: set of size $b_i \longrightarrow$ set of size b_i^2.

$\Rightarrow \Pr[\exists$ any collisions for $h_i] \le 1/2$. $\Rightarrow E[\#$ of trials$] \le 2$.



↳ Primary hash functions h: we want h to have $\Pr_h[\sum_i b_i^2 \ge 4m] \le 1/2$. # of collisions under $h = \sum_i \binom{b_i}{2}$

$= \frac{1}{2}(\sum_i b_i^2 - \sum_i b_i) = \frac{1}{2}(\sum_i b_i^2 - m)$. Take expectations $\to E[\sum_i b_i^2] =$

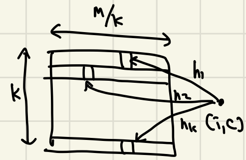
$2 \cdot E[\#\text{ of collisions}] + m \le 2\binom{m}{2}\cdot\frac{1}{n} + m \le m + m \le 2m$ if $n = m$.

$\Rightarrow \Pr[\sum_i b_i^2 \ge 4m] \le 1/2. \Rightarrow E[\#\text{ trials}] \le 2.$

Application) Heavy Hitters in Streams: Stream of data items $(\overset{\text{index}}{i_1}, \overset{\text{count}}{c_1}), \dots (i_T, c_T)$.

Define count$(i,T) = \sum_{t: i_t = i} c_t$. Output all "heavy hitters", i.e.: for a tuple (q, ε, δ),

- if count$(i,T) \ge q$, i must be outputted. $Q := \sum_{t=1}^{T} c_t$
- if count$(i,T) \le q - \varepsilon Q$, i may only be outputted w.p. $\le \delta$.



Idea: maintain a set of m counters in $\underbrace{k}_{\text{TBD}}$ rows and $(\frac{m}{k})$ columns (assume $m|k$).

$C_{a,j}$ (initially all 0) is the value at row a, column j $(1 \le a \le k, 0 \le j \le m/k - 1)$. Use k

2-universal hash functions $h_a : U \to [0, m/k - 1]$ where $U :=$ set of indices.

When data (i_t, c_t) arrives, compute $h_a(i_t)$ for $1 \le a \le k$ and increment

$C_{a,h_a(t)}$ by c_t. Define $C_{a,j}(T) :=$ volume of counter $C_{a,j}$ at time T.

Claim) (i) $\forall$ items i, $\min_{j = h_a(i)\, \forall a \in [k]} \{C_{a,j}(T)\} \ge$ count(i,T). (trivially true)

(ii) w.p. $\ge 1 - (\frac{k}{m\varepsilon})^k$ (over choice of hash functions), $\min_{j = h_a(i)\, \forall a \in [k]} \{C_{a,j}\} \le$ count$(i,T) + \varepsilon Q$.

Proof of (ii): Fix(i,t). Consider the first counter $C_{1,h_1(i)}$ (others follow by symmetry)

$C_{1,h_1(i)} \ge$ count(i,T) at time T. Define $Z_1 :=$ amount increased by items

other than i. $\to Z_1 = \sum_{t=1}^{T} X_t c_t$ where $X_t := \mathbb{1}\{i_t \ne i \wedge h_1(i_t) = h_1(i)\}$.

Since h_1 is 2-universal, $E[X_t] = \Pr[h_1(i_t) = h_1(i)] \le k/m$. So $E[Z_1] \le \frac{k}{m}\sum_{t=1}^{T} c_t = \frac{k}{m}Q$. By Markov, $\Pr[Z_1 > \varepsilon Q] \le \frac{k/m}{\varepsilon} = \frac{k}{m\varepsilon}$. Same applies to $h_2, \dots, h_k$. So, $\Pr[\min_j Z_j \ge \varepsilon Q] = \prod_j \Pr[Z_j \ge \varepsilon Q] \le \left(\frac{k}{m\varepsilon}\right)^k$. //

→ Choose $m = \ln(\frac{1}{\delta})\frac{\varepsilon}{e}$ (total # counters), $k = \ln(\frac{1}{\delta})$ (# of hash functions).

Output all items with min. count$(i,T) \ge q$ and no others. Claim: this works.

∵ If count$(i,T) \ge q$, definitely output i. If count$(i,T) \le q - \varepsilon Q$, i is output w.p. $\le \left(\frac{k}{m\varepsilon}\right)^k \le e^{-\ln(-1/\delta)} = \delta$. (just plug in values into above bounds) //

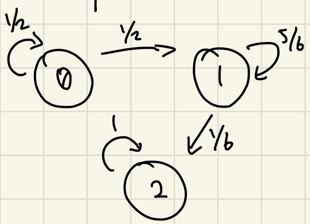
Markov Chains

Def) Stochastic Process: index set T and a set of r.v.s (X_i)

Def) Markov Chain: a stochasic process with Markovian property:

$$\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \dots, X_1 = a_1] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}] = P_{a_t, a_{t-1}}.$$

Graph View ⟷ Matrix View



$$\begin{bmatrix} P_{00} & P_{01} & \cdots \\ P_{10} & \ddots & \\ \vdots & & \ddots \end{bmatrix}$$

Current PMF: $p(0) = [1 \ 0 \ 0]$

Next time step: $p(t) = p(t-1)P$

↳ $\underbrace{p(t) = p(0)P^t}$

Application) 2-SAT: ANDs of OR clauses of at most 2 boolean variables

Idea: Pick an unsatisfied clause. Flip a variable in that clause u.a.r.

Repeat T times. If the statement is ever satisfied, return True. Else, False.

↳ Naively, this takes $O(T \cdot m)$ where $m :=$ # of clauses.

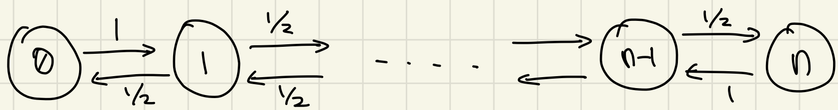
Analysis: This is a one-sided error algo when $\exists$ a sat. assignment.

Let $X_t :=$ # of variables agreeing with a particular sat. assignment $\in [0, n]$.

Observe that $\Pr[X_{i+1} = j+1 \mid X_i = j] \geq \frac{1}{2}$ and $\Pr[X_{i+1} = j-1 \mid X_i = j] \leq \frac{1}{2}$.

Take a $\overbrace{\text{pessimistic view}}^{\text{stochastic domination}}$ and set both inequalities to strict equalities.

For edge cases, $\Pr[X_{i+1} = 1 \mid X_i = 0] = \Pr[X_{i+1} = n-1 \mid X_i = n] = 1$.



Consider hitting times, $h_{n,n} = 0$, $h_{i,n} = \frac{1}{2} h_{i+1,n} + \frac{1}{2} h_{i-1,n} + 1$, $h_{0,n} = h_{1,n} + 1$.

↳ Solve to get $h_{i,n} = n^2 - i^2 \leq n^2$. $\rightarrow$ If $T = 2kn^2$, $\underbrace{\Pr[\text{error}] \leq 2^{-k}}_{\text{Markov \& union}}$.

Classification of States: how to think about long-term behaviors of MC

Def) Accessible: $i \rightarrow j$ if $P^n_{i,j} > 0$ for some n. ($\exists$ a path $i \rightsquigarrow j$)

Def) Communicate: $i \leftrightarrow j$ if $i \rightarrow j$ and $j \rightarrow i$. ($i, j \in$ SCC, equivalence)

Def) Irreducible MC: $\forall i, j$, $i \leftrightarrow j$ (MC is one SCC)

Let $r^t_{i,j} :=$ prob. being at j for the first time after t steps starting from i.

Def) Recurrent State: $\sum_{t=1}^{\infty} r^t_{i,i} = 1$. Def) Transient State: $\sum_{t=1}^{\infty} r^t_{i,i} < 1$.

For a transient state, $X_i :=$ # of times hitting $i \sim$ Geom. r.v.

An alternative interpretation of hitting time of itself is $h_{i,i} = \sum_{t=1}^{\infty} t \cdot r_{ii}^{t}$.

↳ Surprisingly, $h_{i,i}$ can be infinite even if i is recurrent (null recurrent) if the MC is infinite. Otherwise, all states are positive recurrent.

ex) $\mathbb{Z}^d$, choose a random dimension, perturb by ± 1. For $d = 1, 2$, every state is (null) recurrent. For $d \geq 3$, every state is transient.

Def) Periodic State: $\exists \Delta > 0$ s.t. $\Pr[X_{t+s} = i \mid X_t = i] > 0$ only if $s \bmod \Delta = 0$.

Def) Ergodic State: a positive reccurent and aperiodic state.

each with \$$d_1$, \$$d_2$

Ex) Gambler's Ruin: Two players play a fair game until one goes broke.

From player 1's view, $\overset{1}{\circlearrowleft} (-d_1) \xleftarrow{1/2} \bigcirc \xleftrightarrow{1/2} \cdots (0) \cdots \xleftrightarrow{1/2} \bigcirc \xrightarrow{1/2} (d_2) \overset{1}{\circlearrowright}$

Let $q_i :=$ prob. of "winning" from state i. $q_{-d_1} = 0$, $q_{d_2} = 1$, $q_i = \frac{1}{2} q_{i+1} + \frac{1}{2} q_{i-1}$.

↳ Solve to get $q_0 = \frac{d_1}{d_1 + d_2}$, so chance of winning is proportional to betting!

Recall that $p(t) = p(t-1) P$. If $\pi = \pi P$, π is a stationary distribution.

Theorem) Fundamental Theorem of MC: Every finite, irreducible, aperiodic MC 1) has a unique stationary distribution π, 2) $\forall j, i$, $\lim_{t \to \infty} P_{j,i}^{t} \to \pi_i$ is independent of the state j, 3) $\pi_i = \frac{1}{H_{ii}}$ where $H_{ii} :=$ E[return time to i]

Ex1) P is bistochastic (columns also sum to 1): $\sum_i P_{ij} = 1 \ \forall j$.

Then, π is uniform. $\because$ we want $\sum_i \pi_i P_{ij} = \pi_j \ \forall j$. Set $\pi_i = \frac{1}{N} \ \forall i$.

Then, $\sum_i \frac{1}{N} P_{ij} = \frac{1}{N} \sum_i P_{ij} = \frac{1}{N}$. // (If $P_{ij} = P_{ji}$, P is symmetric & bistochastic)

Ex2) P is <u>reversible</u> w.r.t. some dist. π: $\pi_i P_{ij} = \pi_j P_{ji}$, i.e. transitions on the chain are "balanced" (detailed balance). Then, π is the stationary.

$\because$ we need $\sum_i \pi_i P_{ij} = \pi_j \ \forall j$. $\sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j$. //

Random Walks on a Graph

Undirected graph $G(V,E)$, $P_{ij} = \frac{1}{\deg(i)}$ if $(i,j) \in E$, $\emptyset$ otherwise.

↳ P is irreducible iff G is connected, aperiodic iff G is not bipartite.

Claim) P is reversible w.r.t. $\pi(i) = \frac{\deg(i)}{2|E|}$.

$\because$ we want $\pi_i P_{ij} = \pi_j P_{ji} \ \forall i,j$. $\frac{\cancel{\deg(i)}}{2|E|} \times \frac{1}{\cancel{\deg(i)}} = \frac{\cancel{\deg(j)}}{2|E|} \times \frac{1}{\cancel{\deg(j)}} = \frac{1}{2|E|}$.

$H_{ij} :=$ E[time to reach j from i], $C(G) := \max_i$ E[time to visit all nodes starting from i]

↳ this is always modeled by system of lin.eq., $H_{ij} = 1 + \sum_k H_{kj} P_{ik}$.

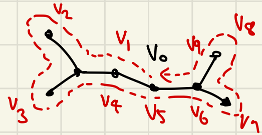
Lemma) $\forall i,j$ s.t. $(i,j) \in E$, $H_{ij} + H_{ji} \leq 2|E|$.

Proof: Replace G by a random walk on directed edges. This is a directed graph with $2|E|$ states. The stationary dist. is uniform, as it is bistochastic

(Observe that any incoming edge for state (i,j) has prob. $\frac{1}{\deg(i)}$). Hence,

$\forall (i,j)$, $\pi(i,j) = \frac{1}{2|E|} = \frac{1}{H_{(i,j),(i,j)}} \to H_{(i,j)(i,j)} = 2|E|$. This is $\underset{i}{\bullet}\to\underset{j}{\bullet}\rightsquigarrow\underset{i}{\bullet}\to\underset{j}{\bullet}$,

which contains a commute $i \to j \to i$. Thus $H_{ij} + H_{ji} \le H_{(i,j)(i,j)} = 2|E|$. //



Theorem) $\forall$ connected G, $C(G) \le 2|E|(|V|-1)$.

Proof: Choose any spanning tree T of G. Pick a vertex v_0 and an Eulerian tour around from it. Label vertices in order of visit, possibly labeling vertices twice. Then, $C \le \sum_{i=0}^{2n-2} H_{v_i, v_{i+1}} = \sum_{(k,j)\in E} (H_{kj} + H_{jk})$

$\le 2|E|\cdot(|V|-1)$ since all $v_i \to v_{i+1}$ are covered at least twice. //

Application) S-T connectivity: start a random walk at S. If it reaches T within $4|E||V|$ steps, output "Yes". Otherwise, output "No". Then, $\Pr[\text{error}] \le \Pr[\text{cover time} > 4|E||V|] \le 1/2$ by Markov. Amplify.

↳ DFS: $O(|E|)$ time & $O(|V|)$ space VS Random Walk: $O(|E||V|)$ time & $O(1)$ space, so we are "trading" time for space, and (time)x(space) is conserved!

Theorem) Mattews Theorem: $\forall$ connected G, $C(G) \le a \cdot \ln n \cdot \overbrace{\max_{i,j} H_{ij}}^{H_{max}}$.

Proof: Consider a r.w. on G of total length $a \cdot H_{max} \cdot \ln n$, divided into "epochs" of length $a \cdot H_{max}$. E[time to visit vertex j in some epoch]

$\leq H_{max}$. Then, $\Pr[j \text{ is not visited in a given epoch}] \leq \frac{H_{max}}{a \cdot H_{max}} = \frac{1}{a}$.

$\Pr[j \text{ is not visited in any epoch}] \leq \left(\frac{1}{a}\right)^{\ln n} = n^{-\ln a}$. By union bound,

$\Pr[\exists \text{ vertex not visited in any epoch}] \leq n \times (n^{-\ln a}) = n^{(1-\ln a)}$.

Now, choose a s.t. $\ln a = 4 \rightarrow \Pr[\text{don't cover graph}] \leq n^{-3}$.

$\rightarrow C(G) \leq a \cdot H_{max} \cdot \ln n + n^{-3} \cdot \underbrace{2n^3}$ worst possible cover time (deterministically)

$\leq a' \cdot H_{max} \cdot \ln n$ where $a' \simeq a$. //

For a r.w. on a number line $[1, n]$, $H_{1n} = (n-1)^2$. Lemma tells that $C(G) \leq 2 \cdot |E| (|V|-1) = 2n(n-1)$. Matthews tells $C(G) \leq a \cdot \ln n \cdot O(H_{max}) = O(n^2 \ln n)$, which is not as tight.

For a clique K_n, $C(G) \leq 2|E|(|V|-1) = O(n^3)$, which is bad because $C(G) = \overbrace{O(n \ln n)}^{\simeq \text{coupon collecting}}$. Matthews tells $C(G) \leq a \cdot \ln n \cdot (n-1) = O(n \ln n)$.

$\Rightarrow$ Which bound is tighter depends on the specific structure of the graph!

MCMC

Goal: Given a prob. dist. π on Ω, sample randomly from Ω w.r.t. π.

Method: Construct a MC on Ω which is ergodic and has SD π.

Then simulate the chain for "suff. many steps" until dist. is close to π. Output.

Ex) Shuffling a deck of n cards: Ω = set of all permutations of deck

1. Riffle shuffle → split L/R deck w.r.t. $Bi(n, \frac{1}{2})$, drop each card from L/R according to current value $\frac{|L|}{|L|+|R|}, \frac{|R|}{|L|+|R|}$ (equivalently, all interleavings of L/R are equally likely).

↳ Ergodicity: Aperiodicity follows from a trivial self-loop step
Irreducibility follows from a mechanical one-card at a time construction.

↳ What is π? Consider the "inverse" riffle shuffle of assigning $\{0,1\}$ to all cards and pulling out all 0s and putting them on top.
This means that P is bistochastic ⇒ π is uniform.

2. Random-to-Top → pick a card u.a.r., put on top.

↳ Ergodic by obvious reasons. Is π uniform? Yes, there are n last steps that have $1/n$ prob. of transitioning to the current step, so P is bistochastic.

3. Random Transposition → Pick two indices i, j u.a.r w/ replacement. Switch the cards at positions i and j.

↳ Ergodic? Yes. Is π uniform? Yes, in fact π is symmetric!

Ex) Random Walk in a Hypercube: on $\{0,1\}^n$, $|\Omega| = 2^n$. Pick a bit u.a.r and flip it. Ergodic? Not aperiodic, but we make it "lazy" by rather than flipping, we set the bit to 0 or 1 u.a.r. π is uniform since a random walk on a hypercube is uniform.

Ex) Graph Coloring Sampling: $G(V,E)$, # of colors q. Ω = set of all proper q-colorings of G. We want π to be uniform.

↳ Pick a vertex v and color c u.a.r, recolor v with c if possible.

↳ Ergodic? aperiodic, but irreducible only if $q \geq \overbrace{\Delta}^{\text{max deg. of } G}+2$ since we can resolve conflicts by offering a "temp" greedy coloring

↳ π uniform? Yes, P is symmetric.

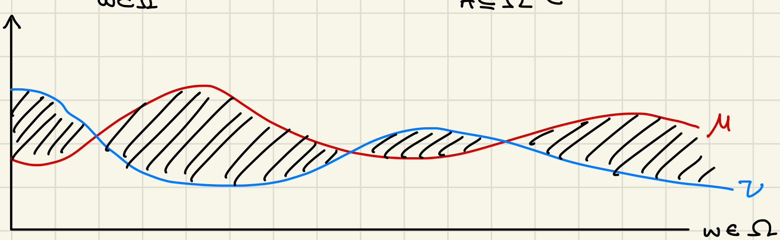
... How do we analyze the appropriate mixing time?

Let P_x^t denote the distribution of MC in t steps starting from x.

Of course, $P_x^t \to \pi$ as $t \to \infty$, but how far is it from π?

Def) Total Variation Distance: for two distributions μ, ν on Ω,

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)| = \max_{A \subseteq \Omega}\{\mu(A) - \nu(A)\}.$$



ex) $\Omega = n!$ permutations, μ = uniform, ν = uniform except Q♡ is on top.

$\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} \{\mu(A) - \nu(A)\}$ where $A :=$ Q♡ is on top $\rightarrow \underline{1 - \frac{1}{52}}$.

Notation) $\Delta_x(t) := \|p_x^t - \pi\|_{TV}$, $\Delta(t) = \max_x \Delta_x(t)$,

$T_x(\varepsilon) = \min\{t \mid \Delta_x(t) < \varepsilon\}$, $T(\varepsilon) = \max_x T_x(\varepsilon)$. (Note: $\Delta_x(t)$ monotone decreases)

Def) Coupling: Any joint distribution ξ on $(\Omega \times \Omega)$ s.t. marginals μ, ν are preserved, i.e. $\forall \omega \in \Omega$, $\sum_{\omega'} \xi(\omega, \omega') = \mu(\omega)$, and vice versa.

↳ For MC, we can design a coupling $Z_t := (X_t, Y_t)$ s.t. X_t & Y_t both behave like the original MC starting from x, y respectively. Let T be the first time s.t. $X_T = Y_T$.

Lemma) Suppose $\exists$ a coupling $Z_t = (X_t, Y_t)$ s.t. $\forall x, y$, $\Pr[X_T \neq Y_T \mid X_0 = x, Y_0 = y] \leq \varepsilon$. Then, $T(\varepsilon) \leq T$.

Proof: For any coupling of r.v.s $X \sim \mu$, $y \sim \nu$, $\Pr[X \neq y] \geq \|\mu - \nu\|_{TV}$, $\cdots (1)$ and $\exists$ a coupling that achieves equalities. Then, $\underline{\Delta(t)} = \max_x \|p_x^t - \pi\|_{TV}$

$\leq \max_{x,y} \|p_x^t - p_y^t\| \overset{(1)}{\leq} \max_{x,y} \Pr[X_t \neq Y_t \mid X_0 = x, Y_0 = t] \underline{\leq \varepsilon} \rightarrow T(\varepsilon) \leq T$. //

Ex) Analysis of Random-to-Top: Couple s.t. we pick the same card.

$\Rightarrow$ The first s cards will always be the same, where $s :=$ # of different cards we have picked so far. When $s = n$, we would have seen every card, so the two decks are equal. This is coupon collecting, so the expected coupling time is $(n \log n + o(n))$, so $\Pr[T > n \ln n + cn] \le e^{-c}$.
$\Rightarrow T = O(n \log n)$, Riffle shuffle is $O(\log n)$ (w/o proof).

Ex) Analysis of Graph Coloring Sampling: Recall that the MC was picking a random vertex v & color c, then recolor v to c if no conflicts. To ensure irreducibility, we have $q \ge \Delta + 2$. π is uniform b/c P is symmetric.

Theorem) If $q \ge 4\Delta + 1$, then mixing time is $O(n \log n)$.

Proof: We design a coupling first. In both X_t & Y_t, we always pick the same v and c at every step. Let $d_t :=$ # of disagreeing vertices at time t. Unfortunately, not all moves are "good", in that they do not decrease d_t. Specifically, a good move is when v is a disagreeing vertex <u>and</u> c is not in the neighborhood of v in both X_t and Y_t. Observe that

# of good moves $\ge d_t(q - 2\Delta)$ is a lower bound of worst case.

A "bad" move that increases d_t happens when v is an agreeing vertex <u>and</u> c is in only <u>one</u> of the neighborhood of X_t or Y_t. # of bad moves $\le$

$\overbrace{d_t \Delta}^{\#\text{vertices}} \cdot \overbrace{2}^{\text{coloring}}$. Thus, (# good moves) − (# bad moves) $\geq d_t(q-2\Delta-2\Delta) = d_t(q-4\Delta)$ $\geq d_t$ since $q \geq 4\Delta+1$. Now, $E[d_{t+1} | d_t] = d_t + Pr[d_{t+1} = d_t + 1 | d_t]$ $- Pr[d_{t+1} = d_t - 1 | d_t] = d_t + \frac{1}{nq} \cdot (2\Delta d_t) - \frac{1}{nq}(q-2\Delta)d_t = d_t(1 - \frac{q-4\Delta}{nq})$ $\leq d_t(1 - \frac{1}{nq})$. Finally, $E[d_{t+1}] = E[E[d_{t+1} | d_t]] = (1-\frac{1}{nq})E[d_t]$.

By induction on t, $E[d_t] \leq (1-\frac{1}{nq})^t \overbrace{E[d_0]}^{\text{bounded}} \leq (1-\frac{1}{nq})^t n$. Set $t = Cqn\ln n \rightarrow E[d_t] = (1-\frac{1}{nq})^{Cqn\ln n} \cdot n \leq e^{-C\ln n} \cdot n \leq n^{1-c}$. Since $Pr[X_T \neq Y_T] = Pr[d_t > 0] = Pr[d_t \geq 1] \leq E[d_t]$, which is a value we can arbitrarily decrease with C, $T = O(n\log n)$.

Generalization: given a weight function $w: \Omega \rightarrow \mathbb{R}^+$, sample accordingly.
↳ As before, construct an ergodic MC s.t. $\pi(\omega) \propto W(\omega)$ and (hopefully) has a rapid mixing time. Sample every T steps.

Metropolis Algorithm: Given $w: \Omega \rightarrow \mathbb{R}^+$, design a connected neighborhood structure (undirected graph) on Ω. Proposal is $K_{xy} = K_{yx}$.
The metropolis rule is that in state $x \in \Omega$, pick a neighbor y w.p. K_{xy}.
Move to y w.p. $\min\{\frac{\omega(y)}{\omega(x)}, 1\}$.
Claim) MC is ergodic & $\pi(\omega) \propto W(\omega)$.

Proof: We can show that $\forall x, y$, $\pi(x) P(x,y) = \pi(y) P(y,x)$. (reversible)
That is, WLOG assuming $\omega(y) \leq \omega(x)$, $\omega(x) P(x,y) = \omega(y) P(y,x)$? By definition, $\omega(x) \cdot K_{xy} \cdot \frac{\omega(y)}{\omega(x)} = \omega(y) \cdot K_{xy} \cdot 1$, so it is satisfied. //

Remark: If we can <u>sample</u> $\underbrace{\pi \propto \omega}$ → we can <u>appoximate</u> $\underbrace{Z = \sum_{\omega} w(\omega)}$.
↳ Concretely, if we can uniformly sample from valid colorings of G → we can get an approximate # of colorings of G. Consider a sequence of graphs, $\emptyset = G_0 \subseteq G_1 \subseteq \dots \subseteq G_m = G$ and $G_i = G_{i-1} + \{e_i\}$. Let $C(G) :=$ # of colorings of G. Then, $C(G) = \frac{C(G_m)}{C(G_{m-1})} \cdot \frac{C(G_{m-1})}{C(G_{m-2})} \cdots \frac{C(G_1)}{C(G_0)} \cdot C(G_0)$.
$C(G_0)$ is trivially q^n. $\frac{C(G_i)}{C(G_{i-1})}$ is the ratio of colorings in G_i that also satisfy with an extra edge e_i. This can be emperically found by running MCMC on each G_{i-1} and checking e_i. $C(G)$ is found. //

Theorem) [Valiant '79?] $\exists$ problems in #P-Comp. whose decision problems are in P.
ex) #-DNF (ORs of AND clauses). Decision for DNF is trivial. However, counting it is hard.
Fact) #-CNF is #P-Complete. (due to Cook's Theorem)
⇒ Suppose φ is DNF. $\overline{\varphi}$ + DeMorgan ⇒ CNF. Observe that #DNF(φ) + #CNF($\overline{\varphi}$) = 2^n. → #DNF is also #P-Complete. //

Def) Fully Polynomial Randomized Approximation Scheme (FPRAS): for a nonnegative function $f: \Sigma^* \to \mathbb{N}$ that, on input (x, ε), runs in time$(|x|, 1/\varepsilon)$ and outputs a value $A(x,\varepsilon)$ s.t. $\Pr[|A(x,\varepsilon) - f(x)| > \varepsilon \cdot f(x)] \leq 1/4$.

(If $\Pr[\text{error}\% \geq \varepsilon] \leq 1/4$, we can boost this via median finding in $O(\log(1/\delta))$ trials to reduce the error prob. to δ)

"Folklore" Statement: Almost all natural #P-Complete problems either have a FPRAS or provably can't be approximated "in any reasonable way".

Examples of FPRAS for #P-Complete Problems:

1) #DNF (HW4: estimating size of union of sets is a reduction)

↳ we can sample from sets like in the HW to obtain a FPRAS.

2) #q-Coloring (for $q \geq 4\Delta + 1$) → assume $\exists$ poly(n) time algorithm that outputs "uniformly" random q-colorings. Then, the remark above is FPRAS.

↳ Analysis: each error must be in bound $\leq \varepsilon/2m$ to get overall error $\leq \varepsilon$.

$$\frac{C(G_i)}{C(G_{i-1})} = \frac{\#\text{ colorings in } G_i \text{ where } e_i = (u,v) \text{ differs}}{\#\text{ colorings in } G_{i-1}} = \Pr[c(u) \neq c(v) \mid c \in C(G_i)].$$

Using Unbiased Estimator Theorem, we need $O(\frac{m^2}{\varepsilon^2} \cdot \frac{1}{p} \cdot \log(m/\delta))$ trials.

We claim that $\forall i, \frac{C(G_i)}{C(G_{i-1})} \geq 2/3$. Set up a multi-mapping $g: \overbrace{C(G_{i-1}) - C(G_i)}^{\text{bad colorings}} \to C(G_i)$.

For a coloring $\sigma \in C(G_{i-1}) - C(G_i)$, just recolor v with some valid color. Observe that each coloring in $C(G_i)$ is hit at most once by g!

→ There are $q-\Delta$ outgoing edges from $C(G_{i-1}) - C(G_i)$ and at most 1 incoming edge to $C(G_i)$. → $\frac{C(G_i)}{C(G_{i-1})} \geq \frac{q-\Delta}{q-\Delta+1} \geq 2/3$ for $q \geq \Delta+2$.

⇒ ∃ FPRAS.

Martingales

Def) $Z_1, \dots, Z_n$ is a <u>martingale</u> w.r.t. $X_1, \dots, X_n$ if:

i) Z_n is a function of $X_1, \dots, X_n$.

ii) $E[|Z_n|] < \infty$.

iii) $E[Z_{n+1} | X_1, \dots, X_n] = Z_n$.

ex) X_n: bet on game n & result, Z_n: total money of gambler after n games

Def) Doob Martingale: $Z_n = E[Y | X_1, \dots, X_n]$ where Y is any RV.

↳ $Z_0 = E[Y]$. $Z_n = E[Y | X_1, \dots, X_n] = Y(X_1, \dots, X_n)$

edge exposure martingale

ex) $Y :=$ largest clique in graph $G_{n,p} = y(X_1, X_2, \dots, X_{\binom{n}{2}})$.

Alternatively, $Y = y(V_1, V_2, \dots, V_n)$ ← vertex exposure martingale

$$\underbrace{E[Z_{n+1} | X_1, \dots, X_n]} = E[E[Y | X_1, \dots, X_{n+1}] | X_1, \dots, X_n] = E[Y | X_1, \dots, X_n] = \underbrace{Z_n}.$$

Def) Stopping Time: $T \geq 0$ for a sequence $X_0, X_1, \dots$ where event $T=n$ depends only on $X_0, \dots, X_n$. (T is a RV!)

ex) Gambling. $T :=$ first time I have \$1000.

↳ a non-example is the last time I have \$1000 (depends on future)

Theorem) Optional Stopping: If $Z_0, Z_1, \dots$ is a martingale w.r.t. $X_0, X_1, \dots$, T is a stopping time for $\{X_i\}$, and if one of the following holds:

i) Z_i is bounded $\forall n$, i.e. $|Z_i| < C$ for some C $\forall i$.

ii) T is bounded.

iii) $E[T] < \infty$ and $E[|Z_{i+1} - Z_i| \mid X_0, \dots, X_i] \leq C$ $\forall i$.

Then, $E[Z_T] = E[Z_0]$.

* For the \$1000 stopping, the optional stopping theorem does not apply!

↳ All three conditions are not met, and $Z_0 = 0$, but $Z_T \geq 1000$.

Gambler: starts at 0, stops when hits $\overset{\text{bankrupt}}{-a}$ or b. What is $\Pr[\text{bankrupt}] =: p$, and what is $E[\text{finish time}]$? Analysis with martingales:

$Z_i :=$ position at time i. → condition (i) holds (Z_i is bounded)

$T :=$ time to reach $-a$ or b. ⇒ By optional stopping, $E[Z_T] = Z_0 = 0$.

$E[Z_T] = p\cdot(-a) + (1-p)\cdot b = 0 \Rightarrow \underbrace{p = \frac{b}{a+b}}, \ (1-p) = \frac{a}{a+b}.$

Define $Y_i := Z_i^2 - i$. We claim that $\{Y_i\}$ is a martingale.

$$E[Y_{i+1}|X_0,\dots,X_i] = E[Z_{i+1}^2 - (i+1)|X_0,\dots,X_i] = E[Z_{i+1}^2|X_0,\dots,X_i] - (i+1)$$

$$= \frac{1}{2}(Z_i+1)^2 + \frac{1}{2}(Z_i-1)^2 - i + 1 = Z_i^2 - i = Y_i.$$

$(Z_{i+1}^2 - i + 1) - (Z_i^2 - i)$

Observe that $E[T] < \infty$, and $E[|Y_{i+1} - Y_i||X_0,\dots,X_i] \le c$ (condition (iii))

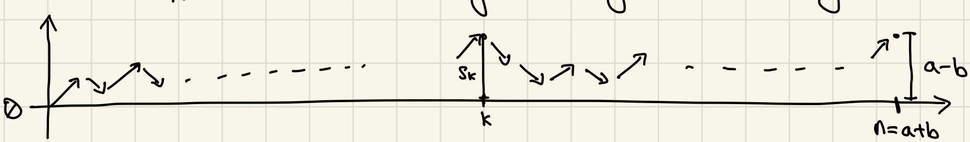
$$\to E[Y_T] = \left(\frac{b}{a+b}\right)\cdot\left[(-a)^2 - E[T]\right] + \left(\frac{a}{a+b}\right)\cdot\left[b^2 - E[T]\right] = Y_0 = 0.$$

$$\to E[T] = \frac{ab}{a+b}(a+b) = \underline{ab}.$$

Ballot Theorem: 2 candidates A, B. A gets a votes, B gets b votes.

WLOG, say A wins the election. We count votes in random order.

What is Pr[A remains winning at every time during counting]?



Define $S_k :=$ A's lead after k votes counted. Define $Z_k := \frac{S_{n-k}}{n-k}$.

Claim) (Z_k) is a martingale w.r.t. sequence of votes read backwards.

Stopping Time $T := \min_k \{Z_k = 0\}$ if such k exists, else $n-1$.

↳ OST condition (ii) applies $\to E[Z_T] = E[Z_0] = \frac{S_n}{n} = \frac{a-b}{a+b}$.

case i: A remains ahead at all times. $\to T = n-1$, $Z_T = \frac{S_1}{1} = 1$.

case ii: A doesn't remain ahead $\to \exists\, k<n$ s.t. $S_{n-k} = 0 \Rightarrow Z_T = 0$.

Let $p := Pr[\text{case } 1]. \to E[Z_T] = p\cdot 1 + (1-p)\cdot 0 = \underline{p = \frac{a-b}{a+b}}.$ //

Proof of Claim: $E[Z_k \mid \text{last } k-1 \text{ votes}] = E[Z_k \mid S_{n-k+1}]$. Conditioning on S_{n-k+1}, we want to show $E[Z_k \mid S_{n-k+1}] = Z_{k-1} \to \underline{\frac{S_{n-k}}{n-k} = \frac{S_{n-k+1}}{n-k+1}}$. Observe that at time $n-k-1$, A had $\frac{(n-k+1)+S_{n-k+1}}{2}$ votes, and B had $\frac{(n-k+1)-S_{n-k+1}}{2}$ votes.

$$E[S_{n-k} \mid S_{n-k+1}] = (S_{n-k+1}+1)\cdot \frac{(n-k+1)-S_{n-k+1}}{2\cdot(n-k+1)} + (S_{n-k+1}-1)\cdot\frac{(n-k+1)+S_{n-k+1}}{2\cdot(n-k+1)}$$

$$= \underline{S_{n-k+1}\cdot\frac{n-k}{n-k+1}}.$$ This satisfies $E[Z_k \mid S_{n-k+1}] = Z_{k-1}$. //

Wald's Equation: Let (X_i) be ind. r.v.s with common finite mean $E[X_i] = \mu^{<\infty}$. Let T be a stopping time for (X_i) s.t. $E[T] < \infty$. Then $\underline{E[\sum_{i=1}^{T} X_i] = E[T]\cdot\mu}$.

Proof: [Assume all X_is are nonnegative.] Define $Z_i = \sum_{j=1}^{i}(X_j - \mu)$. This is obviously a martingale since $E[Z_i \mid X_0 \dots X_{i-1}] = E[X_i + \sum_{j=1}^{i-1} X_i - i\mu] = Z_{i-1}$. Take stopping time as T. $E[T] < \infty$ by definition, and $E[|Z_{i+1} - Z_i| \mid X_0 \dots X_i] = E[|X_{i+1} - \mu|] \le E[|X_{i+1}|] + \mu = 2\mu \ \forall i. \to$ condition (iii)!

$\Rightarrow E[Z_T] = E[Z_0] = 0$. $E[Z_T] = E[\sum_{i=1}^{T} X_i - \mu] = E[\sum_{i=1}^{T} X_i] - E[T]\mu$. //

Theorem) Azuma's Inequality: Let $X_0, \dots, X_n$ be a martingale with bounded differences, i.e. $|X_{i+1} - X_i| \le C_i \ \forall i$. Then, $Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} C_i^2}\right)$.

Proof: Let $D_i := |X_i - X_{i-1}|$. Then $E[D_i \mid X_0, \dots, X_{i-1}] = 0$. We will proceed to prove $Pr[X_n - X_0 \geq \lambda]$. Other tail follows by symmetry.

$\rightarrow$ for $\alpha > 0$, $Pr[e^{\alpha(X_n - X_0)} \geq e^{\alpha\lambda}] \leq e^{-\alpha\lambda} E[e^{\alpha(X_n - X_0)}] = e^{-\alpha\lambda} E[e^{\alpha(D_n + X_{n-1} - X_0)}]$

$= e^{-\alpha\lambda} E[E[e^{\alpha(D_n + X_{n-1} - X_0)} | X_0, \dots, X_{n-1}]]$ $\quad e^{\alpha(X_{n-1} - X_0)} E[e^{\alpha D_n} | X_0, \dots, X_{n-1}]$.

Lemma) $E[e^{\alpha D_n} | X_0 \dots X_{n-1}] \leq e^{(\alpha c_n)^2/2}$. (Seperate proof)

$\rightarrow Pr[X_n - X_0 \geq \lambda] \leq e^{-\alpha\lambda} \cdot e^{(\alpha c_n)^2/2} \cdot E[e^{\alpha(X_{n-1} - X_0)}]$. By iteration,

$\leq e^{-\alpha\lambda} \cdot e^{\frac{\alpha^2}{2}\sum_{i=1}^{n} c_i^2}$. Set $\alpha = \frac{\lambda}{\sum c_i^2} \rightarrow \leq \exp(-\frac{\lambda^2}{2\sum c_i^2})$. //

Proof of Lemma: We shall prove such fact: Let Y be a r.v. taking values in $[-1, +1]$ and $E[Y] = 0$. Then $\forall \alpha > 0$, $E[e^{\alpha Y}] \leq e^{\alpha^2/2}$. This suffices since setting $Y = \frac{D_n}{c_n}$ implies $E[e^{(c_n\alpha) D_n / c_n}] \leq e^{\alpha^2 c_n^2/2}$. Observe that $e^{\alpha x}$ is convex. Thus $e^{\alpha x} \leq \frac{1}{2}(1+x)e^{\alpha} + \frac{1}{2}(1-x)e^{-\alpha}$. Then $E[e^{\alpha Y}]$

$\leq \frac{1}{2}e^{\alpha} + \frac{1}{2}e^{-\alpha} + \cancelto{0}{(\frac{1}{2}e^{\alpha} - \frac{1}{2}e^{-\alpha})E[Y]} = \frac{1}{2}[1 + \alpha + \frac{\alpha^2}{2!} + \cdots] + \frac{1}{2}[1 - \alpha + \frac{\alpha^2}{2!} - \cdots]$

$= 1 + \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} + \cdots = \sum_{i=0}^{\infty} \frac{\alpha^{2i}}{(2i)!}$. Using $(2i)! \geq 2^i \cdot i!$, $\leq \sum_{i=0}^{\infty} \frac{(\alpha^2/2)^i}{i!} = e^{\alpha^2/2}$. //

Suppose we have $f(X_1, \dots, X_n)$ where X_i are indep. $Z_i = E[f(\cdots) | X_0 \dots X_i]$ is a martingale. Azuma tells that $Pr[|f(\dots X_n) - E[f]| \geq \lambda] \leq \exp(-\frac{\lambda^2}{2\sum c_i^2})$. If we insist that f is c-Lipschitz (changing one X_i deviates $f(\cdot)$ by no more than $\pm c$), then Azuma's is naturally useful, $\leq \exp(\frac{-\lambda^2}{2nc^2})$.

Claim) If f is c-Lipschitz and X_i are indep., then $|Z_i - Z_{i-1}| \leq c$ $\forall i$.

Proof: $Z_{i-1} = E[f(X_1 \ldots X_i \ldots X_n) | X_1 \ldots X_{i-1}] = E[f(X_1 \ldots \hat{X}_i \ldots X_n) | X_1 \ldots X_{i-1}]$ where $\hat{X}_i$ has the same distribution as X_i but is independent of all X_j.
$= E[f(X_1 \ldots \hat{X}_i \ldots X_n) | X_1 \ldots X_{i-1}, X_i]$. Now, $|Z_i - Z_{i-1}| = |E[f(X_1 \ldots X_i \ldots X_n) - f(X_1 \ldots \hat{X}_i \ldots X_n) | X_1 \ldots X_i]| \leq C$. //

Application) Pattern Matching: Let $X :=$ uniformly random string $x_1 \ldots x_n \in \Sigma^n$. Let $B :=$ fixed pattern $b_1 \ldots b_k \in \Sigma^k$ where $k << n$. How many times does B appear in X? Let $f(X_1 \ldots X_n) := \#$ of times B appears in X. Define $Z_i = E[f(\cdot) | X_1 \ldots X_i]$. $E[f(\cdot)] = (n-k+1) \cdot |\Sigma|^{-k}$. Azuma's tells that $\Pr[|f - E[f]| \geq \lambda] \leq \exp(-\frac{\lambda^2}{2\sum c_i^2})$. Now observe that f is k-Lipschitz. Thus $\leq \exp(-\frac{\lambda^2}{2k^2 n})$. If k is regarded as a constant, $E[f] = O(n)$, and $\Pr[|f - E[f]| \gg \sqrt{n}]$ is very small.

Application) Balls and Bins: Let $X_i :=$ bin chosen by ball i. $f(X_1 \ldots X_m) := \#$ of empty bins. $E[f] = n(1 - \frac{1}{n})^m \sim ne^{-m/n}$. If $m \sim cn$, $E[f] = O(n)$. Observe that f is 1-Lipschitz. By Azuma, $\Pr[|f - E[f]| \geq \lambda] \leq \exp(\frac{-\lambda^2}{2n})$.

Application) Chromatic Number of $G \in G_{n,p}$: $X_i :=$ expose vertex i and all neighbors of i. $f(X_1 \ldots X_n) := \chi(G)$ defined by $X_1 \ldots X_n$. Observe that f is 1-Lipschitz. Thus, $\Pr[|f-E[f]| \geq \lambda] \leq \exp(\frac{\lambda^2}{2n})$, and deviations are no worse than $O(\sqrt{n})$. For demonstration, let $p = 1/2$.

Fact) For $p = 1/2$, size of max. indep. set of $G \in \mathcal{G}_{n,1/2}$ is $2\log_2 n + $ (low order) w.h.p. $\rightarrow \chi(G) \geq \frac{n}{2\log_2 n}$ w.h.p. A more difficult fact is that $\chi(G) \lesssim \frac{n}{2\log_2 n}$ w.h.p., and thus that $E[\chi(G)]$ is tight. Set $m = \frac{n}{(\log_2 n)^2}$.

While $\exists > m$ uncolored vertices in G, pick an arbitrary subset S of size m of them. Take $I \subseteq S$, the max. indep. set of S. Color I with a new color. When $\leq m$ vertices are uncolored, color them all with seperate new colors. Observe that $|I| = O(\log_2 n))$ w.h.p.

The claim in question is w.h.p. whether <u>every</u> subset $S \subseteq V$ has $|I| = 2\log_2 n + o(\sim)$ i.e., we want $2^n \Pr[S \text{ does not satisfy}] \rightarrow 0$. Turns out that $\Pr[S \sim]$ is $\lesssim \exp(-\Omega(n^2))$, so it works. We use an edge exposure martingale of $f(X_1 \dots X_{\binom{n}{2}}) := $ size of a max family of edge-disjoint indep. sets of size $2\log_2 n$. Azuma's proves a good bound.

Fingerprinting (Out of Scope)

Scenario: Alice & Bob have large files. Can we check whether the two files are equal without communicating too many bits?

Let Alice's file be $\overset{a:=}{a_1,\ldots,a_n}$, Bob's be $\overset{b:=}{b_1,\ldots,b_n}$, n-bit binary strings.

→ Alice: Pick a random prime $\in [2,T]$. Compute fingerprint $F_p(a) = a \pmod p$. Send $F_p(a)$ & p to Bob.

→ Bob: Compute $F_p(b) = b \pmod p$. If $F_p(b) = F_p(a)$, accept. Else, reject.

If $a=b$, the algorithm always suceeds. However, it could be the case that $a \neq b$ but $a \equiv b \pmod p$. → One-sided error

Analysis of error: Error happens when $|a-b| \equiv 0 \pmod p$. $|a-b|$ is an n-bit number. How many distinct prime divisors? → at most n

→ $\Pr[\text{error}] \leq \frac{n}{\#\text{primes} \in [2,T]}$. By the Prime Number Theorem, # of primes in $[2,T]$, $\pi(T) \sim \frac{T}{\ln T}$ as $T \to \infty$. In fact, $\frac{T}{\ln T} \leq \pi(T) \leq 1.26\frac{T}{\ln T}$ $\forall T \geq 17$. Thus, $\Pr[\text{error}] \leq \frac{n}{\pi(T)} = \frac{n \ln T}{T}$. If $T = cn\ln n$, then $\Pr[\text{error}] \leq \frac{n(\ln n + \ln\ln n + \ln c)}{cn\ln n} = \frac{1}{c} + o(1) \Rightarrow T = O(n\ln n)$ suffices, so we only need to send $O(\ln n)$ bits to reach a constant error!

Remark: # of prime divisors can be replaced to $\pi(n)$. Then, the bound improves to $1.26\frac{n}{\ln n} \cdot \frac{\ln T}{T}$. Setting $T = cn$, this becomes $\frac{1.26}{c} + o(1)$!

Ex) $n = 2^{23}$ (~1Mb), $T = 2^{32}$ (32-bit fingerprint) → $\Pr[\text{error}] \leq 1.26\frac{n\ln T}{T\ln n} = 1.26 \cdot \frac{2^{23}}{2^{32}} \cdot \frac{32}{23} \lesssim 0.0035$, so works well empirically!

Application) Pattern Matching: $X := x_1, \dots, x_n$, a long string. $Y := y_1, \dots, y_m$, a short pattern. Does Y appear in X? → Naively, it takes $O(nm)$ time.

* There exist nontrivial deterministic algorithms in $O(n+m)$ time (KMP, etc.)

→ We can develop a simple $O(n+m)$ randomized algorithm via fingerprinting!

- pick a random prime $p \in [2, T]$.
- compute $F_p(Y) = Y \pmod p$.
- for $j \in [1, n-m+1]$:
  - compute $F_p(X[j])$ where $X[j] := X_j \dots X_{j+m-1}$.
  - if $F_p(X[j]) = F_p(Y)$, output "match" and half. ↳ could be wrong
- output "no match".

Choice of T: Naively, $\Pr[\text{error}] \le n \cdot \frac{\pi(m)}{\pi(T)}$ by union bound. However, observe that p is bad if $p \mid |Y - X(j)|$ for some $j \Leftrightarrow p \mid \prod_j |Y - X(j)| \to \le \frac{\pi(nm)}{\pi(T)}$

⇒ $T = cnm$ suffices, so only $\underline{O(\ln n)}$ bits.

Runtime: $O(m)$ for $F_p(Y)$, $O(n)$ iterations, and after one $O(n)$ computation for $F_p(X[1])$, we can find $F_p(X[j+1]) = F_p(2(X[j] - 2^{m+1} X_j) + X_{j+m})$ in (near) constant time. ⇒ Total $O(m) + O(n) \cdot O(1) = \underline{O(n+m)}$. //

Ex) $n = 2^{12}$, $m = 2^8$, $T = 2^{32}$. $\Pr[\text{error}] \le \frac{\pi(nm)}{\pi(T)} \le 1.26 \frac{nm}{\ln(nm)} \cdot \frac{\ln T}{T} = 1.26 \cdot$

$\frac{2^{20}}{2^{32}} \cdot \frac{32}{20} \simeq 0.0005.$

Primality Testing

Question: for an integer n, is n prime?

↳ checking $i = 2 \dots \sqrt{n}$ does not work because we still need $2^{n/2}$ iterations

sampling $i \in [2, \sqrt{n}]$ still does not work because divisors are usually sparse.

Theorem) FLT: If n is prime, then $a^{n-1} \equiv 1 \pmod{n}$ $\forall a \in [1, n-1]$.

Fermat Test: pick $a \in [2, n-1]$ u.a.r. If $\gcd(a, n) = 1$, output "no".

Else, if $a^{n-1} \not\equiv 1 \pmod{n}$ then output "no". Else, output "yes".

Claim) If n is not prime & has a witness (i.e. $a^{n-1} \not\equiv 1 \pmod{p}$), then $\Pr[\text{error}] \leq \frac{1}{2}$ for the Fermat Test, i.e. $\Pr_a[a \in \mathbb{Z}_n^* \text{ is a witness}] \geq \frac{1}{2}$.

Proof: $\mathbb{Z}_n^* :=$ multiplicative group of integers coprime to n. Let $S \subseteq \mathbb{Z}_n^*$ be the non-witnesses. S is a <u>proper subgroup</u> of $\mathbb{Z}_n^*$ (closed under multiplication $\pmod{n}$, $a^{n-1} \equiv 1 \wedge b^{n-1} \equiv 1 \Rightarrow (ab)^{n-1} \equiv 1 \pmod{n}$). Lagrange's Theorem tells that $\frac{|\mathbb{Z}_n^*|}{|S|}$ is an integer, and $|S| < |\mathbb{Z}_n^*|$, so $\frac{|S|}{|\mathbb{Z}_n^*|} \leq \frac{1}{2}$. //

Caveat: $\exists$ non-primes n s.t. $a^{n-1} \equiv 1 \pmod{n}$ $\forall a \in \mathbb{Z}_n^*$ (Carmichael #s) such as 561, 1105, 1729, ... → This is a problem!

Claim) 561 is a Carmichael #. $\Rightarrow a^{560} \equiv 1 \pmod{561}$ $\forall a$.

Proof: $561 = 3 \times 11 \times 17$. It suffices to show that $a^{560} \equiv 1$ (mod 3, 11, 17) by the Chinese Remainder Theorem. By FLT, we know that $a^2 \equiv 1 \pmod 3$, $a^{10} \equiv 1 \pmod{11}$, $a^{16} \equiv 1 \pmod{17}$. $\Rightarrow a^{560} \equiv 1 \pmod{561}$ //

Fact: If n is prime, then 1 has no non-trivial square roots (mod n), i.e. If $a^2 = 1 \pmod n$, then $a = \pm 1$.

Proof: Since $GF(n)$ is a field, the polynomial $x^2 = 1$ has 2 roots. //

* This doesn't work for composites! (e.g. $6^2 = 1 \pmod{35}$)

Algorithm) Miller-Rabin: Assume n is odd (WLOG) & not a prime power (verifiable).

$(n-1) = 2^r \cdot R$ where R is odd. We will test $a^R, a^{2R}, \dots, a^{2^r R} = a^{n-1}$.

Ex) $n = 561$. $(n-1) = 560 = 2^4 \times 35$. Take $a = 2$. $2^{35} \pmod{561} = 263$, $2^{70} \pmod{561} = 166$, $\underbrace{2^{140} \pmod{561} = 67}_{\text{If } n \text{ were prime, this should be } \pm 1!}$, $2^{280} \pmod{561} = 1$, $2^{560} \equiv 1$.

- If n is even or is a prime power, output "no".
- Compute r, R s.t. $(n-1) = 2^r R$ where R is odd.
- Pick $a \in [2, n-1]$ u.a.r.
- If $\gcd(a, n) \neq 1$, output "no".

- Compute $a^R, a^{2R}, \dots, a^{2^r R} = a^{n-1} \pmod{n}$.
- If $a^{n-1} \not\equiv 1 \pmod{n}$, output "no".
- If $a^R \equiv 1 \pmod{n}$, output "yes".
- Else, let $j = \max\{i \mid a^{2^i R} \not\equiv 1 \pmod{n}\}$.
- If $a^{2^j R} \not\equiv -1$, output "no". $\to$ we found a non-trivial sqrt of $n \neq \pm 1$!
- Else, output "yes".

Claim) If n is odd, composite, & not a prime power, then $\Pr_a[a \in \mathbb{Z}_n^*$ is a witness$] \geq 1/2$.

Let $s = 2^r R$ a <u>bad power</u> if $\exists x \in \mathbb{Z}_n^*$ s.t. $x^s \equiv -1 \pmod{n}$.

Lemma) $\forall$ bad power s, $S_n := \{x \in \mathbb{Z}_n^* \mid x^s \equiv \pm 1 \pmod{n}\}$ is a proper subgroup of $\mathbb{Z}_n^*$. [proof at end]

Proof of Claim: We shall show that all non-witnesses belong to S_n.

Suppose a is a non-witness. Then either: 1) $a^R = a^{2R} = \cdots = a^{n-1} = 1$, or 2) $a^{2^i R} = -1$, $a^{2^{i+1} R} = \cdots = a^{n-1} = 1$. Let s^* be the largest bad power in the sequence $R, 2R, \dots, 2^r R$. In case 1), $a^{s^*} = 1$. In case 2), $a^{s^*} = \pm 1$ since $s^* \geq 2^i R$. In both cases, $a^{s^*} = \pm 1$, so $a \in S_n$. By the lemma and Lagrange's Theorem, we are done. //

Proof of Lemma: Need to show that $S_n := \{x \in \mathbb{Z}_n^* \mid x^s \equiv \pm 1 \pmod{n}\}$ is a proper subgroup of $\mathbb{Z}_n^*$. We just need to find some $y \in \mathbb{Z}_n^*$ s.t. $y \notin S_n$. Since s is a bad power, we can find $x \in \mathbb{Z}_n^*$ s.t. $x^s \equiv -1 \pmod{n}$. Since n is odd, composite, & not a prime power, we can write $n = n_1 \times n_2$ where n_1, n_2 are odd and coprime. By CRT, $\exists$ unique $y \in \mathbb{Z}_n^*$ s.t. $y \equiv x \pmod{n_1}$, $y \equiv 1 \pmod{n_2}$. We claim that $y \in \mathbb{Z}_n^* \setminus S_n$. Observe that $\gcd(y, n_1) = \gcd(x, n_1) = 1$. Also, $\gcd(y, n_2) = 1$. So $y \in \mathbb{Z}_n^*$. Next, $y^s \equiv x^s \equiv -1 \pmod{n_1}$. Also, $y^s \equiv 1 \pmod{n_2}$. Suppose $y \in S_n$. Then $y^s \equiv \pm 1 \pmod{n}$. If $y^s = +1$, then $y^s = 1 \pmod{n_1}$. If $y^s = -1$, then $y^s = -1 \pmod{n_2}$. Both cases cause contradiction, so $y \notin S_n$. //