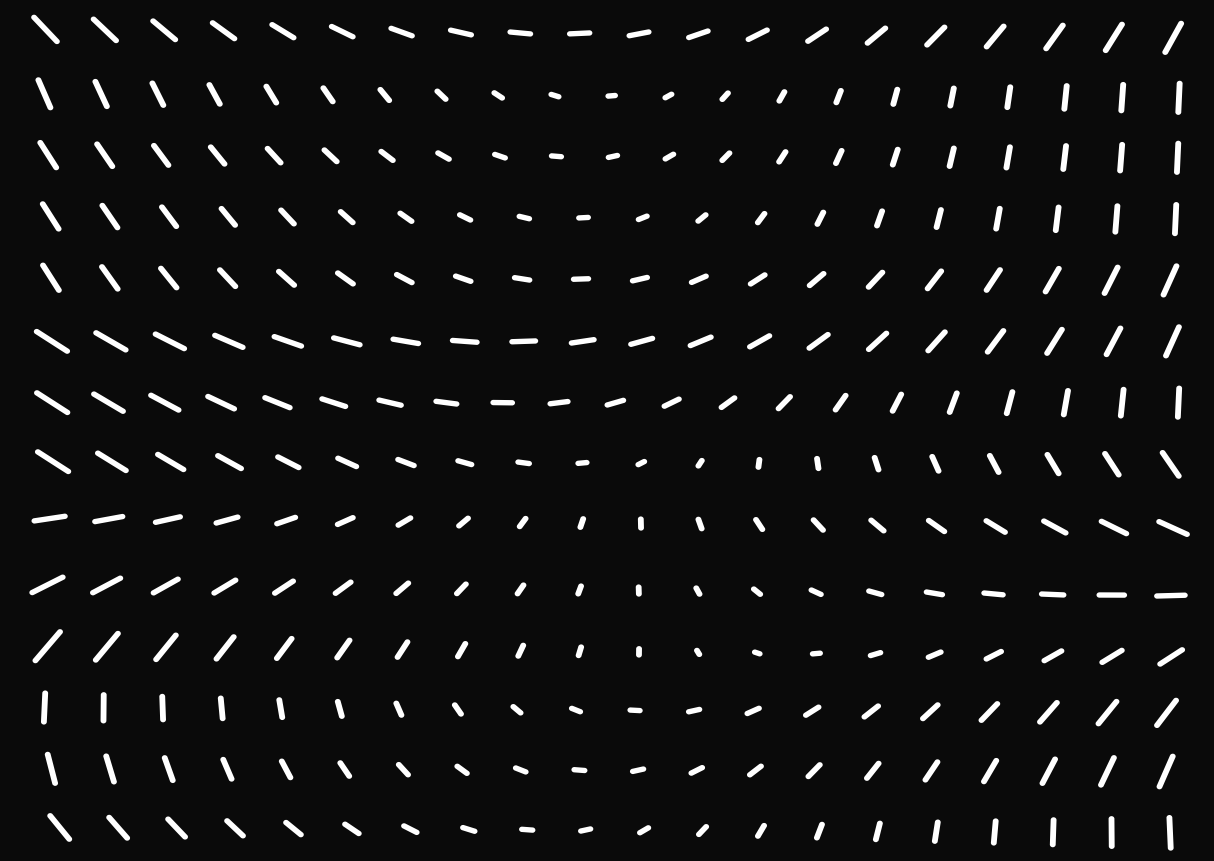


Notes for Quiz

Chernoff Bound:

(Deviation above mean)

$$X = \sum_{i=1}^{n} X_i, \quad \mu = E(X), \quad Pr(X_i = 1) = p_i, \quad \delta > 0$$

$$Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu} \quad (1)$$

$$Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}, \quad 0 < \delta \leq 1$$

$$Pr(X \geq R) \leq 2^{-R}, \quad R \geq 6\mu$$

(Deviation below mean)

For $0 < \delta < 1$

$$Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu}$$

$$Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}$$

Also for $0 < \delta < 1$

$$Pr(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\delta^2/3}$$

Advanced Algorithm

– Sunyajith Chillara

Lecture #1

Assignments	20%
Quiz 1	10%
Mid sem	20%
Quiz 2	10%
End sem	30%
Viva	10%

Book: Probability and Computing Mitzenmacher and Upfal

Verify Matrix Multiplication

Given A, B and C all ($n \times n$ mat)

we need to check if $AB \stackrel{?}{=} C_{n \times n}$

→ Multiply A and B and check if the elements are equal Complexity: n^{ω} → mat. mult exp.

↳ By Vassi, Almans-Williams

So far, only deterministic algorithm

↳ "fixed" Run time → worst case

↳ No Error (always correct)

↓

we want to relax this and make Bounded run time for most of the time.

Randomised Routing

Randomised algorithm : Primality testing, Rabin-Miller

Trend : Hard Problem → Randomization → Approximation

Sn^2 [Blaser] → lower Bound on matrix multiplication.

Updated Question: Is there a rand. algorithm that runs in time $O(n^\delta)$ where $\delta < \omega$. (We want a mostly correct)

If $AB = C$, then $\forall$ vectors v, $ABv = Cv$

else $\exists u$ s.t.
$AB \cdot u \neq C \cdot u \mid u \in$ Nullspace of $AB - C$

$AB\bar{x} = C\bar{x}$

$K \times L$, $L \times S \rightarrow$ KLS complexity.

$AB\bar{x}$ and so $C\bar{x}$ $\Longrightarrow$ $\begin{bmatrix} L_1(\bar{x}) \\ \vdots \\ L_n(\bar{x}) \end{bmatrix} = \begin{bmatrix} L_1'(\bar{x}) \\ L_2'(\bar{x}) \\ \vdots \\ L_n'(\bar{x}) \end{bmatrix}$

$O(n^2)$

$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $\Longrightarrow$ $\begin{bmatrix} L_1(\bar{x}) - L_1'(\bar{x}) \\ \vdots \\ L_n(\bar{x}) - L_n'(\bar{x}) \end{bmatrix}$

If $AB = C$ then $\forall i \in [1, n]$, $L_i(\bar{x}) - L_i'(\bar{x}) = 0$ for every substitution of $\bar{x}$ from S^n

If not, $\exists$ a subset for $\bar{x}$ s.t. not all $L_i(\bar{x}) - L_i'(\bar{x})$ are zero at that point.

$$\Pr_{\bar{a} \in S^n}\left[\bigwedge_{i=1}^{n}\left(L_i(\bar{a}) - L_i'(\bar{a}) = 0\right)\right] \leq \max_{i \in [n]} \Pr\left[L_i(\bar{a}) - L_i'(\bar{a}) = 0\right]$$

where $\bar{a} \in$ Null space of $AB - C$

Bad event: $\forall i,\ L_i(\bar{a}) = L_i'(\bar{a})$

Good Event: $\exists i \in [n]$ s.t. $L_i(\bar{a}) \neq L_i'(\bar{a})$

Prob. of $\bar{a}$ being in null space.

$$\Pr_{\substack{\bar{a} \in S^n \\ R}}\left[\sum_{j=1}^{n} \alpha_j \cdot a_j = 0\right] = \frac{|S|^{n-1}}{|S|^n} = \frac{1}{|S|}, \text{ fix } n-1,\ n^{th} \text{ is fixed}$$

$$\leq \frac{1}{|S|}$$

Algo:

1. Pick $\bar{a}$ uniformly random from S^n (u.a.r.)
2. Compute $AB\bar{a}$ and $C\bar{a}$
3. Check if pointwise equal
4. If yes, return $AB = C$ → 100% correct
   else return $AB \neq C$ → correct with prob. $\geq 1 - \frac{1}{|S|}$

Monte Carlo algorithms: correctness not fixed, time fixed

Las Vegas: correctness 1, time unbounded

Problem: (Polynomial Identity Testing)

Given

$f(x_1, x_2, \ldots x_n)$
$g(x_1, x_2, \ldots, x_n)$ } of degree d

$x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$

$\sum_{i=1}^{n} e_i \leq d$ → no. of coeff. $\binom{n+d}{d}$

if $f(a) = g(a)$ for $d+1$ points then $f(x) = g(x)$

if $f(x) \neq g(x)$ then $\exists\, b$ in $d+1$ distinct evals

$f(b) \neq g(b)$

$$\det(M) = \sum_{\pi \in S_n} \operatorname{sign}(\pi) \prod_{i=1}^{n} M_{i\,\pi(i)}$$

For $M = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$

$$S_3 = \{ \underset{+}{123}, \underset{-}{132}, \underset{-}{213}, \underset{+}{231}, \underset{+}{312}, \underset{-}{321} \}$$

$\operatorname{sign}(123)\, x_{11}x_{22}x_{33} + \operatorname{sign}(132)\, x_{11}x_{23}x_{32}$

$+ \operatorname{sign}(213)\, x_{12}x_{21}x_{33} + \operatorname{sign}(231)\, x_{12}x_{23}x_{31} + \operatorname{sign}(312)\, x_{13}x_{21}x_{32}$

$+ \operatorname{sign}(321)\, x_{13}x_{22}x_{31}$

$x_{11}x_{22}x_{33} - x_{11}x_{23}x_{32} - x_{12}x_{21}x_{33} + x_{12}x_{23}x_{31}$

$+ x_{13}x_{21}x_{32} - x_{13}x_{22}x_{31}$

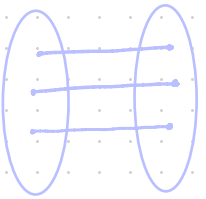
$$\begin{vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{vmatrix} = x_{11}(x_{22}x_{33} - x_{23}x_{32}) - x_{12}(x_{21}x_{33} - x_{23}x_{31}) + x_{13}(x_{21}x_{32} - x_{22}x_{31})$$

$$= \underset{123}{x_{11}x_{22}x_{33}} - \underset{132}{x_{11}x_{23}x_{32}} - \underset{213}{x_{12}x_{21}x_{33}} + \underset{231}{x_{12}x_{23}x_{31}} + \underset{312}{x_{13}x_{21}x_{32}} - \underset{321}{x_{13}x_{22}x_{31}}$$

Lecture #2

Graph Matching - Bipartite

- Matching M is a subset of edges s.t. for every vertex $v \in V(G)$, $|I(v) \cap M| \leq$ where $I(v)$ is the set of edges incident on v.
- Perfect Matching : $\forall v, |I(v) \cap M| = 1$



$V = L \cup R$

$E \subseteq L \times R$

Edmund's Criterion

$$(M_{G_{n \times n}})_{ij} = \begin{cases} x_{ij} & \text{if } (i,j) \in E \\ & i \in L, j \in R \\ 0 & \text{otherwise} \end{cases}$$

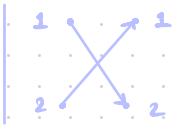
Theorem:
A given bipartite graph has perfect matching (P.M.) if and only if $Det(M_G) \not\equiv 0$

Determinant

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \longrightarrow x_{11}x_{22} - x_{21}x_{12}$$

$1 \longrightarrow 1$

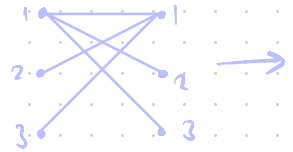
$2 \longrightarrow 2$



Homework: Prove this

$$Det(X) = \sum_{\sigma \in S_n} (-1)^{sgn(\sigma)} \prod_{i=1}^{n} x_{i\sigma(i)}$$

$$\#inv(\sigma) = |\{(i,j) \mid i < j \quad \sigma(i) > \sigma(j)\}|$$



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & 0 & 0 \\ x_{31} & 0 & 0 \end{bmatrix} \end{array} \equiv 0$$ every where

$\therefore$ no perfect matching

PM $\Rightarrow$ $Det(M_G) \not\equiv 0$ | $Det(M_G) \not\equiv 0 \Rightarrow$ PM

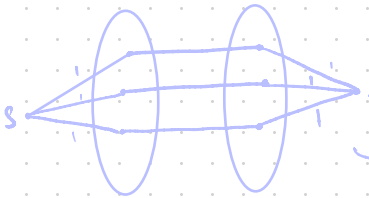
observation: Perfect Matching is infact a permutation.

Further $\prod x_{i\sigma(i)}$ survive iff all $(i, \sigma(i)) \in E$

Sufficient: To find a point where $Det(M_G) = 0$

σ^*: $\forall i$, set $x_{i\sigma^*(i)} = 1$

Existence criteria, not how to find not discussed yet.



complexity: $O(|E| \cdot (m+n))$

$\hookrightarrow$ max flow $\Rightarrow$ Perfect Matching

Problem: Sequential

Want a parallel algorithm.

Matching in RNC^2

claim: For all $e \in E$: check if $G/\{e\}$ has a PM using EC.

$G \leftarrow G/\{e\}$ if Yes

$M \leftarrow M \cup \{e\}$ if No.

Different Ordering will give different perfect Matching

Non-zeroness of a polynomial

Let $f(x_1, x_2, \ldots, x_n)$ be a non-zero polynomial of degree d

Let $S \subseteq R$ of size $> d+1$

Let $(a_1, a_2, \ldots, a_n)$ be chosen s.t. each a_i is picked and uniformly at Random.

$$\Pr_{\bar{a} \sim S^n}\left[f(a_1, a_2, \ldots, a_n)\right] \leq \frac{d}{|S|}$$

$\#\text{Zeros}(f) \leq d \cdot |S|^{n-1}$ over S^n.

$\#\text{Total points} = |S|^n$

Can be proved using induction ← Homework

Algorithm for testing $\text{Det}(M_G) \not\equiv 0$.

1. Pick a Sets of size $100n$
2. Sample $\bar{a}$ independently and uniformly at random from $S^{|E|}$
3. Check if $\text{Det}(M_G(\bar{a})) \stackrel{?}{=} 0$
4. If zero $\Rightarrow$ report $\text{Det}(M_G) \equiv 0$
   Else report $\text{Det}(M_G) \not\equiv 0$

If $Det(M_G) \equiv 0$ then algorithm makes no error

If $Det(M_G) \not\equiv 0$, then the probability of it be being correct is $1 - \frac{1}{100}$

Randomness $= n \cdot \log|S|$

Complexity $= O(|E| \log(100n))$

food for thought for next class:

Having $O(\log(n))$ Randomness = Having no randomness at all.

↳ Area called pseudorandomness

Here n is the number of random bits sampled.

Lecture #3

Graph Matching (Bipartite)

There is a random algorithm that finds a perfect matching in $O(\log^2 n)$ parallel time using $O(m \cdot T_{\det}^{\log^2 n}(n))$ processors.

Outline: Assign wts to edges randomly and ensure and extract unique min. wt. matching

Isolation lemma: Let S be a finite subset of $\mathbb{R}$

Let $T_1, \ldots, T_K \subseteq [m]$ ($[m] = \{1, 2, \ldots, m\}$) be all distinct sets

For each element $i \in [m]$, let wt fn assign a wt ind. and U.A.R. from S. wt of a set T_i is given by

$$\sum_{e \in T_i} wt(e)$$

$$Pr(\text{|min wt set|} = 1) \geq 1 - \frac{m}{|S|}$$

Proof: Let E_i be the event that minimum is obtained by two sets U and V and $i \in U$ but not V.

obs: If min. wt. is not unique then $\exists\, i$, s.t. E_i is true

$$\min_{T_j : i \in T_j} \{wt(T_j)\} = \min_{T_{j'} : i \in T_{j'}} \{wt(T_{j'})\}$$

Bad event: $[\exists\, i \text{ s.t. } E_i]$

Good event: $\Pr\left[\bigcap_{i=1}^{m} \bar{E}_i\right] = 1 - \Pr[\cup E_i] \geq 1 - m \cdot \max_i \{\Pr[E_i]\}$

$\Pr(E_i) \leq \frac{1}{|S|}$

$\therefore \Pr\left(\bigcap_{i=1}^{m} \bar{E}_i\right) \geq 1 - \frac{m}{|S|}$

1. For each each edge $e \in E$, assign a random weight ind. U.A.R. from S.
2. Define matrix W s.t.

$$W_{ij} = \begin{cases} 2^{wt((i,j))} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise} \end{cases}$$

Lemma: Let M_0 be the unique min wt PM in G.

Let $r = wt(M_0)$. Then

$(i,j) \in M_0$ iff $\frac{Det(W^{(i,j)})}{2^r} \cdot 2^{w(i,j)}$ is odd

3. For each edge $(i,j) \in E$, compute $\frac{Det(W^{(i,j)})}{2^r} \cdot 2^{wt((i,j))}$ (in parallel)

$$Det(W) = \sum_{\substack{\sigma \in S_n \\ \sigma \in PM(G)}} (-1)^{sgn(\sigma)} \cdot 2^{\sum_{i=1}^{n} wt(i,\sigma(i))}$$

$M_0, M_1, \ldots, M_K$

$$Det(W) = \sum_{\sigma \in S_n} (-1)^{sgn(\sigma)} \prod_{i=1}^{n} W_{i\,\sigma(i)}$$

$$Det(W) = sgn(M_0) \cdot 2^{wt(M_0)} + sgn(M_1) \cdot 2^{wt(M_1)} + \dots + sgn(M_k) \cdot 2^{wt(M_k)}$$

$$\hookrightarrow \forall i \neq wt(M_i)$$

Lecture #4

$$W_{ij} = \begin{cases} 2^{w(i,j)} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

$wt : E \to S$

u.a.r. + Ind.

Conditioned on

1. PM existence check (Edmond's criteria)
2. Unique min wt guarantee (Isolation lemma)

Claim : $\frac{\det(W)}{2^{wt(M_0)}}$ is odd

Lemma : $M_0 \leftarrow$ min wt PM, $r \leftarrow wt(M_0)$

$(i,j) \in M_0$ if and only if $\det\left(\frac{W^{(i,j)}}{2^r}\right)$

Lecture #5

Randomised Routing (Parallel)

Hypercube: $G = \left(\{0,1\}^n, \{ (u, u \oplus e_i) \mid u \in \{0,1\}^n, i \in [n] \} \right)$ (⊕ = xOR)

e_i → $(0, 0, \ldots, \underset{i^{th}}{1}, \ldots, 0)$ → elementary vector

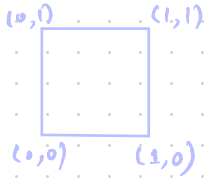
$n = 2$

$\{0,0\}$

$\{0,1\}$

$\{1,0\}$

$\{1,1\}$



We want to route packets on this graph from one vertex to other

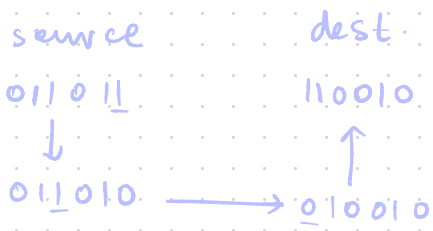
→ Permutation Routing (oblivious/independent Routing)

Bit fixing scheme

source: 0 1 1 0 1 1

destination: 1 1 0 0 1 0 (MSB to LSB)

0 1 1 0 1 1 → 1 1 1 0 1 1 → 1 1 0 0 1 1



Task: Route all packets from every source to every dest as early as possible

There could be waiting at the nodes

Every node can hold unlimited packets and each edge carries at most one packet.

Lemma: There are permutations $\Pi: \{0,1\}^n \to \{0,1\}^n$ for which bit fixing scheme takes $\frac{2^{n/2}}{n}$ time steps to route all the packets

Possible: Hamming distance n

$(a_1, a_2, \ldots, a_n) \to (\bar{a}_1, \ldots, \bar{a}_n)$

$(x_{n/2}, 0_{n/2}) \to (0_{n/2}, x_{n/2}) \quad \forall\ x_{n/2} \in \{0,1\}^{n/2}$ } $2^{n/2}$ many out of 2^n

Claim: Each of these packets need to go through $(0_{n/2}, 0_{n/2})$

At each time step at most n packets can be routed from any node

Time to route all packets $(x_{n/2}, 0_{n/2}) \to (0_{n/2}, x_{n/2})$ is at least $\frac{2^{n/2}-1}{n}$

Theorem: For every deterministic protocol, $\exists$ a permutation that needs at least $\sqrt{\frac{2^n}{n}}$ time.

Randomization: For every permutation, all packets can be routed in at most $O(n)$ time steps with high probability.

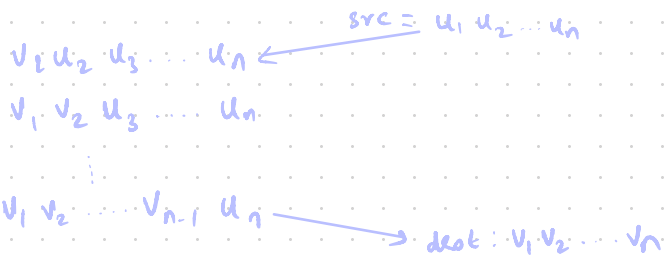
Lecture # 6

Theorem: There is a randomised protocol that routes all packets to their dest in at most $15n$ times w.h.p.

Claim: $\exists\ a \in [n]$ s.t.

$u_i = w_i \quad \forall\ i > a$

$v_i = w_i \quad \forall\ i \leq a$



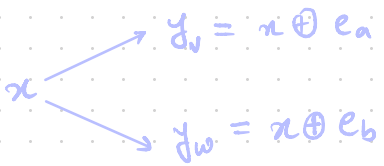
Path(v) = Path from $v \longrightarrow \pi(v)$ using BFS

Lemma 1: For $v \neq v'$, Path(v) and Path(v') do not meet again once they diverge.

Lemma 2: Let S be the set of vertices w s.t. Path(u) and Path(w) intersect. Then P_v takes at most $n + |S|$ time steps to reach $\pi(v)$

Lemma 3: # of w s.t. Path(w) intersects with path $v \to \sigma(v)$ is at most $6n$ w.p. $\geq 1 - 2^{-n}$

similarly, $\sigma(v) \to \pi(v)$ wp $\geq 1 - 2^{-n}$ intersections $\leq 6n$



$$Pr\left[\left|\sum_{i}^{X} x_i - \mu m\right| \geq \varepsilon \mu m\right]$$

$$\leq 2e^{-\mu m(\varepsilon^2)/3}$$

$$2e^{-\mu m(\varepsilon)^2/3} \leq \delta$$

$$\frac{-\mu m(\varepsilon)^2}{3} \leq \log\left(\frac{\delta}{2}\right)$$

$$\mu m \varepsilon^2 \geq 3\log\left(\frac{2}{\delta}\right)$$

$$m \geq \frac{3}{\mu\varepsilon^2}\log\left(\frac{2}{\delta}\right)$$

$x \geq 0$

$e^x \geq x + 1$

$x \to p_i(e^{x_i} - 1)$

Lecture # 7

Chernoff bound:

For some $t \in \mathbb{R}_{\geq 0}$,

$$\Pr[X > (1+\delta)\mu] = \Pr[Xt > (1+\delta)\mu t]$$

$$= \Pr[e^{Xt} > e^{(1+\delta)\mu t}]$$

$$\leq \frac{E(e^{Xt})}{e^{(1+\delta)\mu t}}$$

$$E(e^{Xt}) = E[e^{t\sum_{i=1}^{n} X_i}] = \prod_{i=1}^{n} E[e^{tX_i}]$$

$$E[e^{tX_i}] = e^{t\cdot 1} \cdot p_i + (1-p_i)\, e^{t\cdot 0}$$

$$= p_i(e^t - 1) + 1$$

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

$e^t = 1+\delta$

$$\therefore E[X > (1+\delta)\mu] \leq \left[\frac{e^{(e^t-1)}}{e^{(1+\delta)t}}\right]^{\mu}$$

minimize / derivative to get rid of t

$$= \left[e^{(e^t-1)-(1+\delta)t}\right]^{\mu}$$

set $e^t = 1+\delta$

$$\geq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

Home work : Chernoff bound for error reductions

$$X_i = \begin{cases} 1 & 1/2 \text{ prob} \\ -1 & 1/2 \text{ prob} \end{cases}$$

$$X = \sum X_i \qquad E(X) = 0$$

$$Pr[X > R]$$

$$e^t + e^{-t} \geq 2$$

$$1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} \cdots$$

$$E(e^{tX_i}) = \frac{1}{2} e^t + \frac{1}{2} e^{-t} = \sum_{i=0}^{\infty} \frac{t^{2i}}{(2i)!} \leq \sum \frac{t^{2i}}{2^i \cdot i!} = \sum \frac{(t^2/2)^i}{i!} = e^{t^2/2}$$

$$E(e^{tX}) = \prod_{i=1}^{n} \frac{1}{2} e^t + \frac{1}{2} e^{-t}$$

$$\leq \prod_{i=1}^{n} e^{t^2/2}$$

$$= e^{nt^2/2}$$

$nt - at$

$t = a/n$

$$\therefore E(X \geq a) \leq \frac{E(X)}{e^{at}} = \exp\left(\frac{nt^2}{2} - at\right)$$

$$\geq \frac{n}{2} \frac{a^2}{n^2} - \frac{a^2}{n}$$

$$\frac{a^2}{2n} - \frac{a^2}{n} = -\frac{a^2}{2n}$$

$$\therefore E(X \geq a) \leq \exp(-a^2/2n)$$

Name: Aaryan Ajay Sharma
Roll no: 2022121001
Date: 17/01/2024

Solution to Problem Set 1

1. Let the events $E_1, E_2, \ldots, E_n$ be mutually independent. Then show that the events $\overline{E}_1, \overline{E}_2, \ldots, \overline{E}_n$ are also mutually independent.

Let $P(n)$ be the statement:

The set of events $A = \{E_i \mid 1 \le i \le n\}$ are mutually independent $(\Rightarrow)$ implies

The set of events $B = \{\overline{E}_i \mid 1 \le i \le n\}$ are mutually independent.

We can use strong induction to prove $P(n)$.

Base case: $n = 2$

$P(2)$: E_1 & E_2 are mutually independent $\Rightarrow$ $\overline{E}_1$ and $\overline{E}_2$ are mutually independent

We have,

$$P(E_1 \cap E_2) = P(E_1)P(E_2)$$

$$\Rightarrow P(\overline{\overline{E}_1 \cup \overline{E}_2}) = P(\overline{E}_1)P(\overline{E}_2)$$

$$\Rightarrow 1 - P(\overline{E}_1 \cup \overline{E}_2) = (1 - P(\overline{E}_1))(1 - P(\overline{E}_2))$$

$$\Rightarrow 1 - (P(\overline{E}_1) + P(\overline{E}_2) - P(\overline{E}_1 \cap \overline{E}_2)) = 1 - (P(\overline{E}_1) + P(\overline{E}_2)) + P(\overline{E}_1)P(\overline{E}_2)$$

$$\Rightarrow 1 - (P(\overline{E}_1) + P(\overline{E}_2)) + P(\overline{E}_1 \cap \overline{E}_2) = 1 - (P(\overline{E}_1) + P(\overline{E}_2)) + P(\overline{E}_1)P(\overline{E}_2)$$

$$\Rightarrow P(\overline{E}_1 \cap \overline{E}_2) = P(\overline{E}_1)P(\overline{E}_2)$$

$\therefore$ $P(2)$ is true.

Now, let $P(2), P(3), \ldots, P(m)$ be true.

We need to prove $P(m+1)$.

$P(m+1)$:

The set of events $A = \{E_i \mid 1 \le i \le m+1\}$ are mutually independent $\Rightarrow$

The set of events $B = \{\overline{E}_i \mid 1 \le i \le m+1\}$ are mutually independent.

$$\therefore P(E_1 \cap E_2 \cdots E_{m+1}) = P(E_1)P(E_2)P(E_3)\cdots P(E_{m+1})$$

$$\Rightarrow P(\overline{\bar{E}_1 \cup \bar{E}_2 \cdots \bar{E}_{m+1}}) = P(\bar{E}_1)P(\bar{E}_2)\cdots P(\bar{E}_{m+1})$$

$$\Rightarrow 1 - P(\bar{E}_1 \cup \bar{E}_2 \cup \cdots \cup \bar{E}_{m+1}) = (1-P(\bar{E}_1))(1-P(\bar{E}_2))\cdots(1-P(\bar{E}_{m+1}))$$

(from inclusion-exclusion principle)

$$\Rightarrow 1 - \left\{ \sum_{i=1}^{m+1} P(\bar{E}_i) - \left[\sum_{1\le i<j\le m+1} P(\bar{E}_i \cap \bar{E}_j) \right] + \sum_{1\le i<j<k\le m+1} P(\bar{E}_i \cap \bar{E}_j \cap \bar{E}_k) - \cdots + (-1)^{m+1} P(\bar{E}_1 \cap \bar{E}_2 \cdots \bar{E}_{m+1}) \right\}$$

$$= 1 - \sum_{i=1}^{m+1} P(\bar{E}_i) + \left[\sum_{1\le i<j\le m+1} P(\bar{E}_i)P(\bar{E}_j) \right] - \sum_{1\le i<j<k\le m+1} P(\bar{E}_i)P(\bar{E}_j)P(\bar{E}_k) + \cdots + (-1)^{m+1} P(\bar{E}_1)P(\bar{E}_2)\cdots P(\bar{E}_{m+1}) \quad (Eq.1)$$

Now, since $P(K)$ is true for $\forall\ 2 \le K \le m$ from strong induction

$$\therefore \sum_{1\le a_1<a_2<\cdots<a_K\le m+1} P(\bar{E}_{a_1} \cap \bar{E}_{a_2} \cdots \bar{E}_{a_K}) = \sum_{1\le a_1<a_2<\cdots<a_K\le m+1} P(\bar{E}_{a_1})P(\bar{E}_{a_2})\cdots P(\bar{E}_{a_K}) \quad \forall\ 2 \le K \le m \quad (Eq.2)$$

Using (Eq.2) in (Eq.1), we get:

$$1 - \sum_{i=1}^{m+1} P(\bar{E}_i) + \left[\sum_{1\le i<j\le m+1} P(\bar{E}_i)P(\bar{E}_j) \right] - \sum_{1\le i<j<k\le m+1} P(\bar{E}_i)P(\bar{E}_j)P(\bar{E}_k) + \cdots + (-1)^{m+1} P(\bar{E}_1 \cap \bar{E}_2 \cdots \bar{E}_{m+1})$$

$$= 1 - \sum_{i=1}^{m+1} P(\bar{E}_i) + \left[\sum_{1\le i<j\le m+1} P(\bar{E}_i)P(\bar{E}_j) \right] - \sum_{1\le i<j<k\le m+1} P(\bar{E}_i)P(\bar{E}_j)P(\bar{E}_k) + \cdots + (-1)^{m+1} P(\bar{E}_1)P(\bar{E}_2)\cdots P(\bar{E}_{m+1})$$

$$\therefore P(\bar{E}_1 \cap \bar{E}_2 \cdots \bar{E}_{m+1}) = P(\bar{E}_1)P(\bar{E}_2)\cdots P(\bar{E}_{m+1})$$

$\therefore P(m+1)$ is true.

$\therefore$ The statement $P(n)$:

The set of events $A = \{E_i \mid 1 \le i \le n\}$ are mutually independent $\Rightarrow$ The set of events $B = \{\bar{E}_i \mid 1 \le i \le n\}$ are mutually independent.

is proved.

2. Consider a set of integers $\{1, 2, \ldots, n\}$ (denoted by $[n]$). We generate a subset X of $[n]$ using the following random process – a two-sided and unbiased coin is flipped independently for each element a of the set $[n]$ and we add a to the set X if and only if the coin lands HEADS.

   (a) What is the probability distribution over all subsets of $[n]$ under this process.

   (b) Suppose two sets X and Y are chosen independently and uniformly at random from all the subsets of $[n]$, then determine

      i. the probability that X is a subset of Y, and

      ii. the probability that $X \cup Y = \{1, 2, \ldots, n\}$.

(a) For each element a in $[n]$, there are two possibilities – either it is included in the subset (if coin lands HEADS) or it is not included in the set (if the coin lands TAILS). Since coin flips are independent, there are 2^n possible outcome, each corresponding to a different subset of $[n]$. The probability of each subset is $1/2^n$ because all the 2^n subsets are equally likely outcomes.

So, the probability distribution over all subsets of $[n]$ is a uniform distribution, where each subset has probability of $1/2^n$

(b) i) There are 4 distinct possibilities for an element a when X and Y are independently chosen: i) $a \notin X$, $a \notin Y$
ii) $a \notin X$, $a \in Y$
iii) $a \in X$, $a \notin Y$
iv) $a \in X$, $a \in Y$

For $X \subseteq Y$, only i), ii) and iv) should be true. Since all of these can happen with equal probability of $1/4$, the probability $X \subseteq Y$ will be $(\frac{1}{4}+\frac{1}{4}+\frac{1}{4})^n = (\frac{3}{4})^n$

ii) For $X \cup Y$ to be equal to $[n]$, every element in $[n]$ must be in either X or Y or both. The probability for each element to be in either X or Y (or both) is $1/2 + 1/2 - \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}$. Since this is true for each element in $[n]$, the overall probability that $X \cup Y = [n]$ is $\left(\frac{3}{4}\right)^n$, where n is the size of $[n]$.

3. Let Y be a random variable assuming only non-negative values. Then show that for all $t \in \mathbb{R}_{\geqslant 0}$,

$$\mathbb{P}[X \geqslant t] \leqslant \frac{\mathbb{E}[X]}{t}.$$

Let $\mathbb{I}$ be an indicator random variable s.t.

$$\mathbb{I} = \begin{cases} 1 & \text{if } X \geq t \\ 0 & \text{otherwise} \end{cases}$$

$\therefore E(\mathbb{I}) = 1 \cdot P(X \geq t) + 0 \cdot P(X < t)$

$= P(X \geq t)$

Also, $\mathbb{I} \leq \frac{X}{t}$ since for $X \in [t, \infty)$ $\mathbb{I} = 1$ and $\frac{X}{t} \in [1, \infty)$

and for $X \in [0, t)$, $\mathbb{I} = 0$ and $\frac{X}{t} \in [0, 1)$

$\therefore E(\mathbb{I}) \leq E(\frac{X}{t})$

$\Rightarrow P(X \geq t) \leq \frac{E(X)}{t}$

Reference: Mitzenmacher, upfal - Probabilistic Computing, Page 47.

4. Let X be a random variable with expectation μ_X and standard deviation σ_X. Then show that for any $t \in \mathbb{R}_{\geqslant 0}$,

$$\mathbb{P}\left[|X - \mu_X| \geqslant t \cdot \sigma_X\right] \leqslant \frac{1}{t^2}.$$

Similarly, show that for any $\nu \in \mathbb{R}_{\geqslant 0}$,

$$\mathbb{P}\left[|X - \mu_X| \geqslant \nu\right] \leqslant \frac{\mathrm{Var}[X]}{\nu^2}.$$

we will use the results we got in Q.3

We know that $P(X \geq t) \leq \frac{E(X)}{t}$

Now, substitute X with $|X - \mu_X|$ (since X is non-negative)

$\therefore P(|X-\mu_X| \geq t) \Rightarrow P(|X-\mu_X|^2 \geq t^2) \leq \frac{E(|X-\mu_X|^2)}{t^2}$

$\Rightarrow P(|X-\mu_X|^2 \geq t^2) \leq \frac{Var(X)}{t^2}$

this is valid for any $t = \nu \in \mathbb{R}_{\geq 0}$

$\therefore P(|X-\mu_X| \geq \nu) \leq \frac{Var(X)}{\nu^2}$

replacing $\nu = t \cdot \sigma_X$, where σ_X is the standard deviation, we get:

$$P(|X-\mu_X| \geq \sigma_X \cdot t) \leq \frac{Var(X)}{\sigma_X^2 \cdot t^2} = \frac{1}{t^2}.$$

$\therefore P(|X-\mu_X| \geq \sigma_X t) \leq \frac{1}{t^2}$

5. The weak law of large numbers states that if $X_1, X_2, X_3, \ldots$ are independent and identically distributed random variables with mean μ and standard deviation σ, then for any constant $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right] = 0.$$

We know the inequility :

$$P(|X - \mu_X| \geq t) \leq \frac{Var(X)}{t^2}$$

substituting X with $\frac{X_1 + X_2 + \ldots X_n}{n}$, we get

$$Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{n \cdot \sigma^2}{n^2} \leq \frac{\sigma^2}{n}$$

$$\therefore P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2}$$

$$\Rightarrow P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

As $\lim_{n\to\infty}$, $\frac{\sigma^2}{n\varepsilon^2} \to 0$

$$\therefore \lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) = 0$$

6. Suppose we run a experiment E (with randomness) that succeeds with a probability p and fails with a probability $1-p$. Suppose X is a random variable defined to be the number of trials needed for the experiment E to succeed. Then show that the following hold.

   (a) $\mathbb{P}[X = i] = (1-p)^{i} \cdot p$.

   (b) $\mathbb{P}[X \geqslant i] = (1-p)^{i-1}$.

   (c) $\mathbb{E}[X] = \frac{1}{p}$.

(a) $X = i \Rightarrow$ the experiment succeeds in i trails, meaning the i^{th} trial was a success and the $(i-1)$ trials were all fails.

$\therefore P(X=i) = (1-p)^{i-1} \cdot p.$

(b) $P(X \geq i) = \sum_{k=i}^{\infty} P(X=k)$

$= \sum_{k=i}^{\infty} (1-p)^{k-1} \cdot p$

$= p \cdot \frac{(1-p)^{i-1}}{1-(1-p)}$

$= \frac{p}{p} \cdot (1-p)^{i-1}$

$= (1-p)^{i-1}$

(c) $E(X) = \sum_{i=1}^{\infty} i P(X=i)$

$= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \cdot p$

$= \frac{1}{p}$.

$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$

$\sum_{i=1}^{\infty} i \cdot x^{i-1} = \frac{1}{(1-x)^2}$ (derivative w.r.t x)

$\sum_{i=1}^{\infty} i \cdot x^{i-1}(1-x) = \frac{1}{1-x}$

$\sum_{i=1}^{\infty} i(1-p)^{i-1} p = \frac{1}{p}$ (let $x = 1-p$)

7. A function $f : \mathbb{R} \mapsto \mathbb{R}$ is said to be convex if for any x_1, x_2 and $0 \leqslant \lambda \leqslant 1$, the following inequality is satisfied.

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2).$$

Let Z be a random variable that assumes values in the interval $[0, 1]$ and let $p = \mathbb{E}[Z]$. Define a Bernoulli random variable X such that $\mathbb{P}[X = 1] = p$ and $\mathbb{P}[X = 0] = 1 - p$. Then show that for any convex function f, $\mathbb{E}[f(Z)] \leqslant \mathbb{E}[f(X)]$. Please do not invoke Jensen's Inequality directly. You could however prove it and use it.

First, we will prove Jensen's Inequality which is as follows:

For a random variable X, and convex function f, the following inequality holds:

$$E(f(x)) \geq f(E(x))$$

Assuming that f is twice differentiable, we will use another well-known definition of convex function as follows:

For a twice differentiable function f, it is convex if and only if

$$f''(x) \geq 0$$

We will also use Taylor's theorem which is as follows: * Reference given at the end of the answer

Suppose that f is defined on some interval I around a and suppose $f^{N+1}(x)$ exists on this interval. Then for each $x \neq a$ in I there is a value z between x and a such that

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(z)}{(N+1)!}(x-a)^{N+1}$$

Now, since we assumed the second derivative of f exist, we can expand f using Taylor's Theorem as follows:

$$f(x) = f(u) + f'(u)(x-u) + \frac{f''(a)(x-u)^2}{2!}, \text{ for some } a \in (x, u)$$

since $f(x)$ is convex $\forall x \in \mathbb{R}$, $\therefore f''(a) \geq 0$

$\therefore f(x) = f(u) + f'(u)(x-u) + \underline{\frac{f''(a)(x-u)^2}{2!}} \geq f(u) + f'(u)(x-u)$

$\therefore f(X) \geq f(u) + f'(u)\cdot(X-u)$

$\therefore E(f(X)) \geq E(f(u)) + f'(u)(E(X)-u)$ (Taking expectation on both sides)

Now, since $E(X) = u$, we get:

$$E(f(X)) \geq f(E(X)) + f'(u)(u-u)$$

$$\Rightarrow E(f(X)) \geq f(E(X))$$ (Jensen's Inequality)

Now, since f is convex, the following inequality holds:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Given $p = E(z)$, we can write p as follows:

$p = \int_0^1 z \, dF(z)$, where $F(z)$ is the cdf of the continuous RV z.

Also, the expected value of $f(z)$ will be

$E(f(z)) = \int_0^1 f(z) \, dF(z)$

The expected value of $f(x)$ will be

$$E(f(x)) = pf(1) + (1-p)f(0)$$

Applying $x_1 = 0$, $x_2 = 1$, $\lambda = z$ (for $z \in [0,1]$), we get:

$f(z) = f(z\cdot 1 + (1-z)\cdot 0) \leq z \cdot f(1) + (1-z) f(0)$

$\Rightarrow f(z) \leq z \cdot f(1) + (1-z) f(0)$

$$\therefore \quad E(f(Z)) = \int_0^1 f(z) \cdot dF(z)$$

$$\leq \int_0^1 [z f(1) + (1-z) f(0)] \, dF(z)$$

$$= f(1)\left[\int_0^1 z \, dF(z)\right] + \left(1 - \left[\int_0^1 z \, dF(z)\right]\right) f(0)$$

$$= f(1) \cdot p + (1-p) f(0)$$

$$= E(f(X))$$

$$\therefore \quad E(f(Z)) \leq E(f(X))$$

---

* **Reference:** https://www.whitman.edu/mathematics/calculus_online/section11.11.html

↳ I ended up not using Jensen's Inequality at all in the end, but kept the proof just because I already wrote it before I came up with the proof. Sorry for the inconvenience. (T_T)

8. Suppose you pick a graph on n vertices by picking each edge with probability p independently of the others. What is the expected number of triangles in the graph.

Let X be the total triangles in a graph,

Let $X = X_1 + X_2 + \dots + X_N$

where N = maximum possible triangle in the graph and $X_i \rightarrow$ where the i^{th} triangle actually is within the graph.

$$\therefore X_i = \begin{cases} 1 & \text{if } i^{th} \text{ triangle present} \\ 0 & \text{otherwise} \end{cases}$$

Now, the probability that $X_i = 1$ is the probability that the i^{th} triangle is present in the graph, which is simply $p \times p \times p = p^3$ since edge is picked with probability p independently of each other.

$$\therefore P(X_i = 1) = p^3$$

$$\therefore E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0)$$

$$= P(X_i = 1) = p^3$$

$$\therefore E(X) = E(X_1 + X_2 + \dots + X_N)$$

$$= E(X_1) + E(X_2) + \dots + E(X_N) \quad \text{(linearity of Expectation)}$$

$$= p^3 + p^3 + \dots + p^3 = Np^3$$

Here, N = max. possible triangles in a graph = ${}^nC_3$, where n = no. of vertices in a graph.

$$\therefore E(X) = {}^nC_3 \cdot p^3$$

9. Consider the problem of deciding whether two integer multisets S_1 and S_2 are identical in the sense that each integer occurs the same number of times in both sets. The problem can be solved by sorting the two sets in $O(n \log n)$ time where n is the cardinality of the multisets. Suggest a way of representing this as a problem involving verification of polynomial identity and thereby obtain an efficient randomized algorithm.

We can represent an integer multiset as polynomial by representing an arbitrary integer of the multiset a and its frequency f as a term in the polynomial whose power is a and whose coefficient is f i.e. $f \cdot x^a$.

∴ The algorithm for constructing polynomial from multisets is as follows:

CONSTRUCT-POLYNOMIAL(S) : (S → multiset)

1. Initialize $P \leftarrow \Phi$
2. For each $a \in S$:
2.1. P.add(S.a.value, S.a.frequency)
3. Return P

Here we consider:

S → multiset containing objects as element

S.a → object representing an arbitrary element of the multiset

S.a.value → value of the integer that the object represents in the integer multiset

S.a.frequency → frequency of the integer represented by a.

P → Polynomial, initialized 0 at the beginning

P.add(m,n) → A method to add a term with power m and frequency n to P

Now that we can convert our two multisets S_1 and S_2 (say) to P_1 and P_2 (say), our problem has been reduced to verifying whether the two polynomials P_1 and P_2 are equal or not.

This can be done via Schwartz-Zippel lemma* which is as follows:

Let $Q(x_1, x_2, \ldots, x_n)$ be a non-zero multivariate polynomial of total degree d defined over the field $F[x_1, x_2, \ldots, x_n]$.

Fix any set $S \subseteq F$ and let $r_1, r_2 \ldots, r_n$ be chosen independently and uniformly at random from S. Then

$$Pr[\, Q(r_1, r_2, \ldots, r_n) = 0 \mid Q(x_1, x_2, \ldots x_n) \not\equiv 0 \,] \leq \frac{d}{|S|}$$

*Reference: Motwani, Raghvan, Prabhakar - Randomized Algorithms Page: 165

The lemma/theorem can be proved via induction.

∴ our final algorithm for verification is as follows:

VERIFY-EQUAL(S_1, S_2):

1. $P_1 \leftarrow$ CONSTRUCT-POLYNOMIAL(S_1)
2. $P_2 \leftarrow$ CONSTRUCT-POLYNOMIAL(S_2)
3. Sample $r_1, r_2, \ldots, r_n$ independently and uniformly at random from $S \subseteq F$, with $|S|$ being sufficiently large
4. Compute $P_1(r_1, r_2, \ldots, r_n)$ and $P_2(r_1, \ldots, r_n)$

5. If $P_1(r_1, r_2 \ldots, r_n) \neq P_2(r_1, r_2, \ldots r_n)$:
RETURN $P_1 \not\equiv P_2$

else:
RETURN $P_1 \equiv P_2$ with probability $> 1 - \frac{d}{|S|}$.

We can reduce our error probability by taking an arbitrarily large S or by repeating the algorithm K more times which would make our error probability $\left(\frac{d}{|S|}\right)^K$ since trials are independent from each other.

Complexity:

- The polynomial construction takes $O(n)$ time since each element in the multiset contributes one term to the polynomial.
- The evaluation of the polynomial at a point can be efficiently done in $O(n)$.
- Therefore, the overall complexity will be $O(n)$, which more efficient than the $O(n \log n)$ time required for sorting based approach.