APM461 Notes

bowen

February 13, 2026

Contents

1 Jan 7(Hall's THM, Bipartite, Det, Matching, SZ)

1. (Stirling Approximation)

Theorem 1.1

$n! = \Theta(\frac{1}{\sqrt{n}}\dot{(}\frac{n}{e})^n)$

Proof. $log(n!) = \sum_i log(i) \approx \int_1^n logxdx = nlogn - (n-1)$ □

2. Matching on Bipartite

- G is a bipartite if $V(G) = X \sqcup Y$ and $E(G) \subset X \times Y$
- A matching M in G is $MsubE(G)$ such that no two edges share a vertex
  In particular when $|X| = |Y| = |M| = n$, M is a perfect matching

3. When does G has a perfect matching?
   By Hall's Theorem
$$\exists S \subset X \ s.t. \ |N(S)| < |S|$$

4. (Hall's Theorem)

Theorem 1.2

In bipartite G, $V(G) = X \sqcup Y$. There exists a perfect matching iff

$$\forall S \subset X \, , \, |N(S)| \geqslant |S|$$

Proof. Induction on n
Case 1 $\forall S \subset X$ s.t. $0 < |S| < n$ and $|N(S)| \geqslant n+1$

Then take any edge e.

Consider $G' = G \backslash e$, apply induction hypothesis on G', we get perfect matching M' on it.

Then we have $M = M' \cup \{e\}$, which is a perfect matching on G

Remark:
This is the case when we have lots of candidate edges, when pick any matching forms a perfect matching.

Case 2 $\exists S^* \subset X$ s.t. $0 < |S^*| < n$ and $|N(S^*)| = |S^*|$

Consider $V(G') = S^* \sqcup N(S^*)$ where $|G'| < |G|$
Then by induction hypothesis, $\exists$ S^*-perfect matching onto $N(S^*)$, say M'

Now remains to match $X \backslash S^*$ onto $Y \backslash N(S^*)$

Let $G'' = (X \backslash S^*) \sqcup (Y \backslash N(S^*))$, let $S \subset X \backslash S^*$

Claim: $N_{G''}(S) \geqslant |S|$ (otherwise $N(S \cup S^*)$ is too small)

From assumption $|N_G(S \cup S^*)| \geqslant |S \cup S^*|$

$RHS = |N_G(S^*)| + |N_{G''}(S)| = |S^*| + |N_{G''}(S)|$
$LHS = |S| + |S^*|$

Hence $N_{G''}(S) \geqslant |S|$

In conclusion, in both cases induction hypothesis follows, which completes the induction. □

5. An algorithm to check if there exists a perfect matching on bipartite G.

   For each edge $e(= ij)$, pick a subset $\tilde{A} \subset [M]$.
   For non-edge (i, j), $(\tilde{A})_{ij} = 0$.

   If $det(\tilde{A_G}) = \begin{cases} 0 & \text{No perfect matching in G} \\ \neq 0 & \text{Yes perfect matching} \end{cases}$

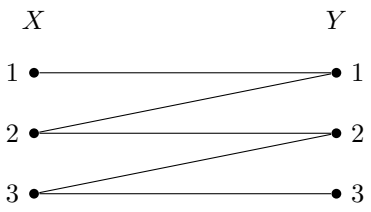
   (i) If G has a perfect matching, then

   $$Pr[\text{Algorithm outputs Yes}] \geqslant 1 - \frac{n}{M}$$

   (ii) If G has no perfect matching, then

   $$Pr[\text{Algorithm outputs No}] = 1$$

6. Take the following bipartite as an example

   Then we have the adjacency matrix A_G as

   $$A_G = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

   Then randomly replace the nonzero entries with real numbers, we have

   $$\tilde{A_G} = \begin{bmatrix} 7 & 0 & 0 \\ e & 13 & 0 \\ 0 & 1 & \pi \end{bmatrix}$$

   We compute $det(\tilde{A_G}) = 7 \cdot 13 \cdot \pi \neq 0$, hence G has a perfect matching.

7. Revisit determinant

   Let $A \in M_{n \times n}(\mathbb{R})$, we define the determinant as

   $$det(A) = \sum_{\sigma} sgn(\sigma) \cdot \prod_{i} a_{i,\sigma(i)}$$

   Remark:

   - $sign(\sigma)$ is number of flips to reach σ from identity e_0.
   - More precisely $sign(\sigma) = \begin{cases} +1 & \text{if permutation can be obtained by even number of steps} \\ -1 & \text{if permutation can be obtained by odd number of steps} \end{cases}$
   - $S_n = \{\text{all permutations on } [n]\}$. (i.e. the symmetric group)
   - $det : M_{n \times n} \to \mathbb{R}$ covers all the permutations on $[n]$ which is $2^{\text{constant}n}$ times, but we can easily reduce the amount of time by linear algebra tricks, such as similarity.

8. Existence of perfect matching via determinant

   If G has no perfect matching, then $\forall \sigma \in S_n$,

$$\exists i \ s.t. \ (i, \sigma(i)) \notin E(G) \ \text{ and } \ \prod A_{i,\sigma(i)} = 0$$

   Hence $det(\tilde{A_G}) = 0$.

   If G has a perfect matching. Define $M(x_{11}, x_{22}, \ldots, x_{nn})$ to be

$$M_{ij} = \begin{cases} x_{ij} & (i,j) \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

   Let $P(x_{11}, \ldots, x_{nn})$ be polynomial $det(M)$.

   If G has a perfect matching, then $P(x_{11}, \ldots, x_{nn})$ is nonnegative polynomial.

9. (Schwartz–Zippel Lemma)

Theorem 1.3

Let A be a finite set with $|A| = M$. If $Q(y_1, \ldots, y_k)$ is a nonzero polynomial of total degree at most d, then

$$\Pr\big[Q(\vec{y}) = 0\big] \leqslant \frac{d}{M}, \qquad \vec{y} \in A^k \text{ where } y \text{ is chosen uniformly at random.}$$

Proof. We proceed by induction on the number of variables k.

For $k = 1$, the polynomial Q has at most d roots, then we have

$$\Pr[Q(\vec{y}) = 0] \leqslant \frac{d}{M}$$

(Induction Step) Assume it holds for k variables.

Consider $Q(y_1, \ldots, y_k, z) = \sum_{i=0}^{t} Q_i(y_1, \ldots, y_k) z^i$, where $Q_t(y_1, \ldots, y_k)$ is non-negative polynomial and $deg(Q_t) \leqslant d - i$.

Then $\Pr[Q(y_1, \ldots, y_k, z) = 0] = \Pr_{\hat{y} \in A^k, z \in A}[Q(\hat{y}, z) = 0]$

Let $B \subset A^k$, where $B = \{\hat{y} : Q(\hat{y}, z) = 0\}$

Then we have

$$\begin{aligned} \Pr[Q(\hat{y}, z) = 0] &= \Pr[\hat{y} \in B] + \Pr[Q(\hat{y}, z) = 0 \mid \hat{y} \notin B] \cdot \Pr[\hat{y} \notin B] && (1) \\ &\leqslant \Pr[\hat{y} \in B] + \Pr[Q(\hat{y}, z) = 0 \mid \hat{y} \notin B] && (2) \\ &\leqslant \frac{d-t}{M} + \frac{t}{M} = \frac{d}{M} && (3) \end{aligned}$$

□

Remark:

- For nonzero polynomial $Q(y_1, \ldots, y_k)$ with degree $\leqslant$ d and $|A| = M$, then

$$|\{(y_1, \ldots, y_k) \in A^k : Q(y_1, \ldots, y_k) = 0\}| \leqslant d \cdot M^{k-1}$$

- Taking the set B eliminates the case before picking z, we already get the desired 0 value, which motivates the conditional probability.
- The first half of the inequality comes from IH on k-entries.

2 Jan 14(Pf for Hall's THM, MaxFlow-MinCut, Menger)

1. Basics of a flow network

   Let $G = (V, E)$ be a directed graph with $E \subseteq V \times V$ and $(a, a) \notin E$.

   - A flow in G is a map $f : E \to \mathbb{R}_{\geqslant 0}$ such that

     $$\forall v \in V,\ \text{Netflow}(f, v) = \sum_{\text{e pointing v}} f(e) - \sum_{\text{e leaving v}} f(e) = 0$$

     Notice that G can't have a positive flow unless it contains a cycle.
   - A s-t flow $(s, t \in V)$ in flow f satisfies

     $$\text{Netflow}(f, v) = 0 \quad \forall v \in V \backslash \{s, t\}$$

     Remark: s-t flow network can only have nonzero net flow on source and tank, all the rest points must obey the balance.
   - (Flow value) $Val(t) = \text{Netflow}(t) = -\text{Netflow}(s)$
   - (Netflow of a set) For $A \subseteq V$, $\text{Netflow}(A) =$

     $$\sum_{\substack{(u,v)\in E \\ u\notin A,\ v\in A}} f(u, v) \ - \sum_{\substack{(u,v)\in E \\ u\in A,\ v\notin A}} f(u, v)$$

   - (Maxflow) A capacity function on G is a map $cap : E \to \mathbb{R}_{\geqslant 0}$
     The maxflow value in G, cap is the maximum $val(f)$ overall s-t flow $f : E \to \mathbb{R}_{\geqslant 0}$ such that $\forall e,\ f(e) \leqslant cap(e)$
   - (Cut) An s-t cut is a subset of vertices such that $s \in A$ and $t \notin A$
     Given G,cap, cut-value $= \sum_{\substack{e=(u,v)\in E \\ u\in A \\ v\notin A}} \text{cap}(e)$

2. Some facts

   - $\text{Netflow}(A) = \sum_{v\in A} \text{Netflow}(v)$
   - $\text{Netflow}(V) = \sum_{v\in V} \text{Netflow}(v) = \text{Netflow}(s) + \text{Netflow}(t) = 0$
   - For any s-t flows, f obeying capacity constraints cap.
     For any set-cut A, $Val(f) \leqslant \text{cutvalue}(A)$
     Remark: To form a flow, every vertex except source and tank has to satisfies $in = out$.

3. (MaxFlow-MinCut Theorem)

Theorem 2.1

Let $G = (V, E)$ be a directed flow network with source s, sink t,
and capacity function $\text{cap} : E \to \mathbb{R}_{\geqslant 0}$. *Then*

$$\max\{Val(f) : f \text{ is a legal } s\text{–}t \text{ flow in } G\} \ = \ \min\{cutvalue(A) : (A, V\backslash A) \text{ is an } s\text{–}t \text{ cut}\},$$

where the value of a cut $(A, V\backslash A)$ is defined by

$$\text{cap}(A, V\backslash A) = \sum_{\substack{(u,v)\in E \\ u\in A,\ v\notin A}} \text{cap}(u, v).$$

Proof. (Ford-Fulkerson Algorithm)

Start with $f = 0$.
Repeatedly proceed the following steps:

- Graph G^* where $E^* = \{(u,v) \in V \times V \ : \ ((u,v) \in E,\ f(u,v) < cap(u,v)) \text{ OR } ((u,v) \in E,\ f(u,v) > 0)\}$
- Look for a path from s to t in G^*
- For some $\epsilon > 0$. For each $i = 0, \ldots, k-1$, either increase $f(v_i, v_{i+1})$ or decrease $f(v_{i+1}, v_i)$

Assume cap is $\mathbb{N}$-valued, always pick $\epsilon = 1$. Then the algorithm always creates $val(f)$ by 1 until there is no such s-t path in G^*.
If G^* has no such s-t path then let $A = \{v \in V \ : \text{ there exists a path from s to v in } G^*\}$

Claim: cutvalue(A) = Val(f)

Proof. From definition,

$$\text{cutvalue}(A) = \sum_{\text{e leaving A}} cap(e)$$

$$\begin{aligned}
\text{val}(f) &= -\,\text{Netflow}(s) = -\,\text{Netflow}(A) \\
&= -\left(\sum_{\substack{e=(u,v)\in E \\ u\notin A,\ v\in A}} f(e) - \sum_{\substack{e=(u,v)\in E \\ u\in A,\ v\notin A}} f(e) \right) \\
&= -\left(0 - \sum_{\substack{(u,v)\in E \\ u\in A,\ v\notin A}} \text{cap}(u,v) \right) \\
&= \sum_{\substack{(u,v)\in E \\ u\in A,\ v\notin A}} \text{cap}(u,v) = \text{cutvalue}(A).
\end{aligned}$$

□

□

Remark:

(1) There is no edge in G^* from A to A^c, then $\forall v \in A, u \in A^c$, $f(v,u) = cap(vu)$ and

$\forall u \in A^c, v \in A, f(u,v) = 0$

(2) The intuition of the proof is to use edges that still < capacity but do not reuse edges.

(3) Since in s-t flow network, on s,t have nonzero netflow which impacts the value, then we have $-\,\text{Netflow}(s) = -\,\text{Netflow}(A)$.

(4) To construct G^* from G, if an edge has extra capacity, then preserve its direction; if an edge reaches full capacity, flip the arrow and then flow on such new network.

4. Menger's Theorem (Edge version)

Theorem 2.2

Let G be an undirected graph. Let $s \neq t \in V$.

$$\exists\ k \text{ edge disjoint paths between } s \text{ and } t \iff \lambda(G) = k$$

Remark:

- Menger's Theorem also works on directed graphs.
- Edge-disjoint path means paths that can share vertex but not edges.
- $\lambda(G)$ is the edge connectivity which equals to the minimum number of edges whose removal disconnected G.

Proof. Let $\tilde{G}$ be the directed graph on G, which maps (u, v) to (u, v) and (v, u).

Let $cap(e) = 1 \quad \forall e \in E$. Given that Max-flow = Min-cut

We will show Menger's theorem by showing

(i) #{edge disjoint path} = #{cut-value}

(ii) #{flow in s-t system} = #{edge disjoint path}

Claim1. In a network where every edge has capacity 1, any integer s–t flow of value k can be decomposed into k edge-disjoint s–t paths.

Let f be an integer s–t flow with value k. Consider the directed subgraph

$$H = \{e \in E : \ f(e) = 1\}.$$

Since f satisfies flow, where every vertex $v \neq s, t$ has

$$\deg_H^+(v) = \deg_H^-(v),$$

while s has $\deg_H^+(s) - \deg_H^-(s) = k$ and t has $\deg_H^-(t) - \deg_H^+(t) = k$.

Step 1 (extract one path). Starting from s, follow an outgoing edge in H. Because all intermediate vertices have equal in- and out-degree in H, the walk cannot get stuck before reaching t; if a directed cycle appears, delete the cycle from the walk. Thus we obtain an s–t path P in H.

Step 2 (subtract the path). Define a new flow $f' = f - \mathbf{1}_P$, i.e., decrease $f(e)$ by 1 on edges of P and leave other edges unchanged. Then f' is still a feasible integer flow (no edge becomes negative, and conservation holds), and its value is $k - 1$.

Repeating this procedure k times yields k pairwise edge-disjoint s–t paths. □

Claim2. In a flow system with $\{0, 1\}$ capacity value, minimum cut value equals the local edge connectivity between s and t.

Assume $\lambda(s, t) = k$, then every set of edges with size $< k$ fails to disconnect s and t.

Suppose not the case, the minimum cut in s-t system $< k$. Then there exists a cut(A, A^c) where $s \in A, \ t \in A^c$ such that the cut value $\leqslant k - 1$.

Let S be the set containing the crossing edges from A to A^c. Notice that remove S disconnects s from t. But $|S| \leqslant k - 1$, then we found a set of edges with size $k - 1$ that disconnects s from t, contradict to $\lambda(s, t) = k$.

Therefore minimum cutvalue $\geqslant k$.

Claim3. If there exists a s-t edge cut F of size k, then there exists a s-t cut with cutvalue $\leqslant k$. The set of all crossing edges S from A to A^c is a subset of F.

Then $|S| \subset F$, hence $|S| \leqslant k$, which gives minimum cutvalue $\leqslant$ k.

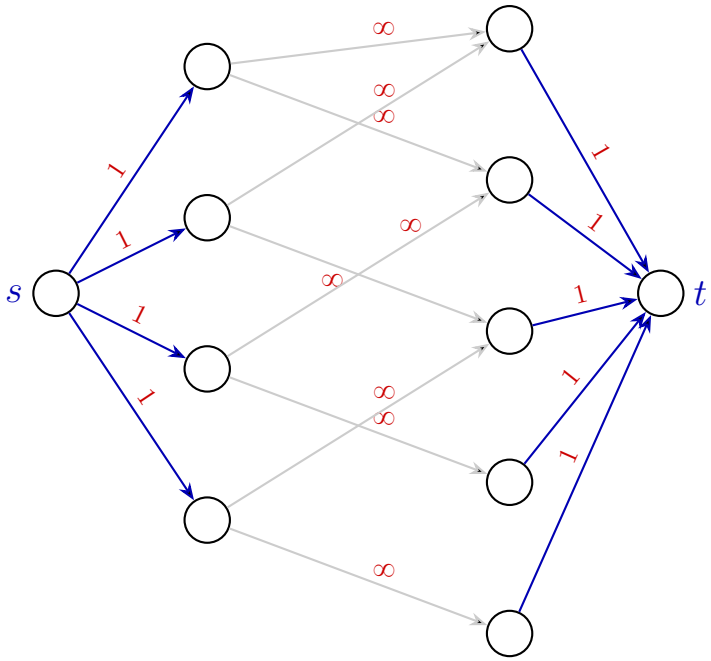
Therefore, we get minimum cutvalue $= k = \lambda(s, t)$. □ □

5. (An alternate proof for Hall's Theorem)

Theorem 2.3

Let $G = X \sqcup Y$, there exists perfect matching $\iff \forall S \subset X, \ |N(S)| \geqslant |S|$

Proof. We consider the following graph.



Claim1. minimum cutvalue $= |X|$.

First consider the cut $S' = \{s\}$, $T' = X \cup Y \cup \{t\}$. Then minimum cutvalue $\leqslant |X|$ for any cut S, T.

Let A and A^c be a cut in the flow network. There cannot be an edge $e = uv$ such that $u \in A \cap X$ and $v \in A \cap Y$. Otherwise the capacity of the flow will be infinite but t can only stand finite capacity. Then $N(A \cap X) \subseteq A \cap Y$

Then we can compute cap(A,A^c)

$$
\begin{aligned}
&= \sum_{\substack{u \in A \\ v \in A^c}} cap(u, v) \\
&= \sum_{v \in A^c \cap X} cap(s, v) + \sum_{u \in A \cap Y} cap(u, t) \\
&= |A^c \cap X| + |A \cap Y| \\
&\geqslant |X| - |X \cap A| + |N(A \cap X)| = |X|
\end{aligned}
$$

Therefore minimum cutvalue $= |X|$.

Claim2. There exists a X-perfect matching.

By maxflow-mincut theorem, there exists a flow with capacity $|X|$. From our construction, every vertex in X must be saturated, since $f(s, x) = 1 \quad \forall x \in X$ and $f(y, t) = 1 \quad \forall y \in Y$, there exists a matching M that saturates X.

Then minimum cutvalue $= |X| \iff |N(A)| \geqslant |A| \quad \forall A \subseteq X$

□

6. Konig Theorem

Theorem 2.4

Let $G = X \sqcup Y$, the size of maximum matching equals to the size of minimum vertex cover.

Proof. Construct a directed network G_0 as follows. Add a source s and sink t.

- For each $x \in X$, add an arc (s, x) of capacity 1.
- For each $y \in Y$, add an arc (y, t) of capacity 1.
- For each edge $\{x, y\} \in E$ (with $x \in X$, $y \in Y$), add an arc (x, y) of capacity U, where $U := |X| + 1$ (any number $> |X|$ suffices; one may also think of $U = \infty$).

(1) Flows and matchings. Given a matching $M \subseteq E$, define a feasible flow f of value $|M|$ by sending one unit of flow along each path $s \to x \to y \to t$ for every matched edge $xy \in M$ (and sending 0 on all other arcs). Hence $maxflow(G_0) \geqslant \nu(G)$.

Conversely, since all capacities are integers, there exists an integer maximum flow f.
Because $\mathrm{cap}(s, x) = 1$, each $x \in X$ has outflow at most 1; similarly each $y \in Y$ has inflow at most 1 since $\mathrm{cap}(y, t) = 1$. Therefore, even though $\mathrm{cap}(x, y) = U$, we necessarily have $f(x, y) \in \{0, 1\}$ for every (x, y). Define

$$M := \{\, xy \in E : x \in X,\ y \in Y,\ f(x, y) = 1 \,\}.$$

Then M is a matching: if $f(x, y) = f(x, y') = 1$ with $y \neq y'$, then x would have outflow $\geqslant 2$ but inflow at most 1 from s, contradiction. Similarly $f(x, y) = f(x', y) = 1$ would give inflow $\geqslant 2$ to y but outflow at most 1 to t. Finally, the value of f equals $|M|$ because

$$|f| = \sum_{x \in X} f(s, x) = \sum_{y \in Y} f(y, t) = \sum_{(x,y) \in E} f(x, y) = |M|.$$

Hence $maxflow(G_0) = \nu(G)$.

(2) Cuts and vertex covers. Let $W \subseteq X \cup Y$ be a vertex cover, and write $W(X) := W \cap X$, $W(Y) := W \cap Y$. Let $X' := X \backslash W(X)$ and $Y' := Y \backslash W(Y)$. Define a cut (S, T) by

$$S := \{s\} \cup X' \cup W(Y), \qquad T := W(X) \cup Y' \cup \{t\}.$$

Since W is a vertex cover, there are no edges of G between X' and Y'. Therefore no arc (x, y) of capacity U crosses from S to T. The only arcs crossing from S to T are:

- arcs (s, x) with $x \in W(X)$, contributing total capacity $|W(X)|$;
- arcs (y, t) with $y \in W(Y)$, contributing total capacity $|W(Y)|$.

Thus $\mathrm{cap}(S, T) = |W(X)| + |W(Y)| = |W|$. Hence $mincut(G_0) \leqslant \tau(G)$.

Conversely, let (S, T) be a minimum s–t cut of capacity $r := mincut(G_0)$. Since $r \leqslant |X|$ (e.g. cut $\{s\}$ has capacity $|X|$) and $U > |X|$, no arc (x, y) of capacity U can cross from S to T. Define

$$X' := S \cap X, \qquad W(Y) := S \cap Y, \qquad W(X) := T \cap X, \qquad Y' := T \cap Y.$$

Then $S = \{s\} \cup X' \cup W(Y)$ and $T = W(X) \cup Y' \cup \{t\}$. Because no arc (x, y) crosses from S to T, there are no edges between X' and Y' in G. Hence every edge of G is incident to $W(X) \cup W(Y)$, so

$$W := W(X) \cup W(Y)$$

is a vertex cover. Moreover, the cut capacity satisfies

$$\mathrm{cap}(S, T) = |W(X)| + |W(Y)| = |W|.$$

Thus $\tau(G) \leqslant mincut(G_0)$.

Combining, $\nu(G) = maxflow(G_0) = mincut(G_0) = \tau(G)$ by Max-Flow Min-Cut.

□

3 Jan 21(Dilworth's THM, Konig THM, Sperner's THM, LYM)

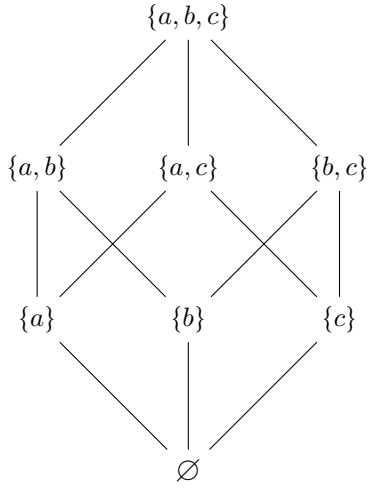
1. Basics

   - For a set to be partial order set $(S, \preceq)$ if it satisfies
     (i) (Reflexive) $\forall a \in S,\ a \preceq a$
     (ii) (Transitive) $\forall a, b, c \in S$, $a \preceq b$ and $b \preceq c$ then $a \preceq c$
     (iii) (Anti-symmetric) $\forall a, b \in S$, $a \preceq b$ and $b \preceq a$, then $a = b$
     (iv) $\forall a, b \in S$, $(a \preceq b)$ or $(b \preceq a)$

     Remark: If property 4 also holds, then S is totally ordered.
   - A **chain** in a poset S is a subset $C \subseteq S$ such that $\forall c, c' \in C$, either $c \preceq c'$ or $c' \preceq c$
     i.e. Any two elements in a chain are comparable.
   - An **antichain** in a poset S is a subset $A \subseteq S$ such that $\forall a, a' \in A$, $(a \npreceq a')$ and $(a' \npreceq a)$
     i.e. Any two elements in an antichain are incomparable.
     Remark: Singletons are both chain and antichain.

2. Examples

   - $(\mathbb{R}, \leqslant)$, whose largest size of antichain is 1.
   - $(\mathbb{R}^2, \text{coordinatewise}_{\leqslant})$, whose largest size of antichain is ∞. Consider the points reached by $(+1, -1)$.
   - $(\{a, b, \ldots, z\}, \leqslant)$ is totally order
   - (strings, lex-order) is totally order
   - $(P(X), \subseteq)$, power sets are posets
   - $(\mathbb{N}\backslash\{0\}, |)$, whose largest size of antichain is ∞, since there are infinitely many primes.
   - An example of drawing poset $(P(\{a, b, c\}), \subseteq)$



3. Max-Min argument on antichain

Theorem 3.1

In any finite poset $(S, \preceq)$,
The size of largest antichain equals the smallest number of chains that cover the poset

Proof. **Claim1.** $|A|_{\max} \leqslant k_{\min}$

Proof. Since for any antichain A, consider chains $c_1, \ldots, c_k$ such that $\bigcup c_i = S$. $|A \cap c_i| \leqslant 1$ for each i, then $|A| \leqslant k$. Hence $|A|_{\max} \leqslant k_{\min}$. □

Remark: c_i, c_j are not necessarily disjoint.

Claim2. $|A|_{\max} \geqslant k_{\min}$

Proof. Let $G = S^- \sqcup S^+$, $(x^-, y^+) \in E$ if $x \preceq y$ and $x \neq y$
Remark: Not equal eliminates the case when they form a perfect matching

Let $|M| = m$ be the size of maximum matching M, by konig theorem, $|U| = |M| = m$, where U is the smallest vertex cover.

WTS: $\exists$ an antichain A and chain covers $c_1, \ldots, c_k$ where $S = \bigcup c_i$ *s.t.* $|A| = k$

Start with n chains, which are all singletons.

Keep doing the following:

If there is an edge in the matching, $(x^-, y^+) \in M$, then merge the chain of x and Y

Invariants:
If $(x^-, y^+) \in M$ is an unused edge, then x^- is the largest element of its chain and y^+ is the smallest element of its chain

This ensures no cycle is created and each vertex has at most one successor and at most one predecessor.

Then eventually, # of chains $= n - m = k_{\min}$

Let U be the smallest vertex cover in the bipartite graph. Define $U^- = U \cap S^-$, $U^+ = U \cap S^+$.

Clearly there are no edges from U^{-^c} to U^{+^c}

Define $A = \{x \in S : \ x^- \notin U^-, \ x^+ \notin U^+\} = S \backslash (U^- \cup U^+)$.

Since there can be some unmatched vertices in U^-, then $|A| = n - (|U^-| + |U^+|) = n - m$

To see why A is an antichain, we notice A contains the vertices which are not in the vertex cover, hence can't be comparable, otherwise we find an edge (x^-, y^+) is not covered.

Since we have built an antichain of size $n - m$ from a chain cover, we have $|A|_{\max} \geqslant k_{\min}$

□

Remarks on the invariants:

- The invariant guarantees at every merge step, x has no successor and y has no predecessor, which makes merge legal.
- There can be two bad cases:
  - If x already has successor, it will turn into a tree (bifurcation)
  - If y already has predecessor, it will likely create a cycle

□

4. Konig Theorem see previous chapter
5. (Sperner's Theorem)

Theorem 3.2

Let $X = [n]$, poset $(P(X), \subseteq)$, the largest antichain in this poset has size $\binom{n}{\lfloor \frac{n}{2} \rfloor}$

Proof. Let $A \subseteq P([n])$ be an antichain, $\sigma \in S_n$ be **uniformly random** permutation.
For $B \subseteq [n]$, let E_B be the event that

$$B = \{\sigma(1), \sigma(2), \ldots, \sigma(|B|)\}$$

Since the permutation is uniformly random, let $|B| = k$, $\Pr[E_B] = \frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}}$

Remark: for example, $E_{\{}1\} \equiv' \sigma(1) = 1'$,
$E_{\{}1, 3\} \equiv' \sigma(1) = 1 \ \wedge \ \sigma(2) = 3' \ OR \ '\sigma(1) = 3 \ \wedge \ \sigma(2) = 1'$
If A is an antichain, consider $\{E_B : B \in A\}$

Claim2. If B, B' are incomparable, then $E_B \cap E'_B = \varnothing$
i.e. No two events share the same permutation.

Proof. Suppose not the case. $\sigma \in E_B, E'_B$. Then

$$B = \{\sigma(1), \sigma(2), \ldots, \sigma(|B|)\}$$
$$B' = \{\sigma(1), \sigma(2), \ldots, \sigma(|B'|)\}$$

That contradicts with B, B' are incomparable. □

Consider

$$\sum_{B \in A} \frac{1}{\binom{n}{|B|}} \leqslant 1.$$

For every $B \subseteq [n]$, we have

$$\binom{n}{|B|} \leqslant \binom{n}{\lfloor n/2 \rfloor},$$

then

$$\frac{1}{\binom{n}{|B|}} \geqslant \frac{1}{\binom{n}{\lfloor n/2 \rfloor}}.$$

Summing over all $B \in A$, we have

$$\sum_{B \in A} \frac{1}{\binom{n}{|B|}} \geqslant \sum_{B \in A} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} = \frac{|A|}{\binom{n}{\lfloor n/2 \rfloor}}.$$

Therefore,

$$\frac{|A|}{\binom{n}{\lfloor n/2 \rfloor}} \leqslant \sum_{B \in A} \frac{1}{\binom{n}{|B|}} \leqslant 1,$$

which gives

$$|A| \leqslant \binom{n}{\lfloor n/2 \rfloor}.$$

□

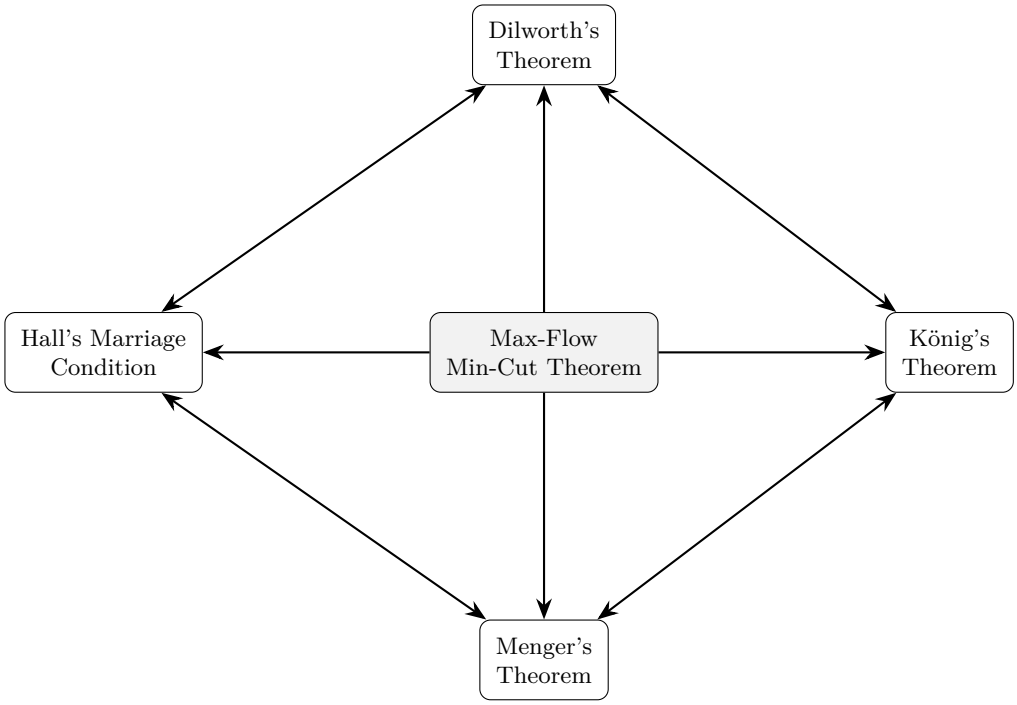
6. (LYM-inequality)

Theorem 3.3

Let A be an antichain in $P([n])$. Let $A_k = A \cap \binom{[n]}{k}$ which means k all subset of $[n]$ that lives in A.
Then $\sum_{k=0}^{n} \frac{|A_k|}{\binom{n}{k}} \leqslant 1$

Proof. From the previous one, $LHS = \sum_{B \in A} \frac{1}{\binom{n}{|B|}} \leqslant 1$ □

7. The relation among theorems



4 Jan 28(Matching Trick, Intersecting Family, FLT)

1. A magic trick on matching (**Easy Version**)

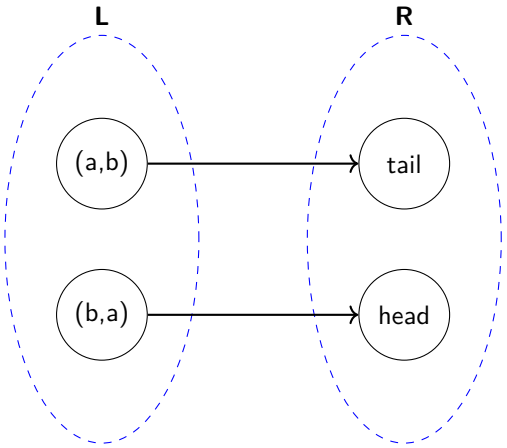
   Let M_1, M_2 be two magicians and V be a volunteer.

   - M_2 offers a deck of 52 cards to v
   - V picks 2 cards and flips a fair coin.
   - Then V gives those 2 cards and the coin to V_2
   - M_2 reads out 2 cards
   - M_1 announces the side of the coin

   Remark:
   Notice that V offers the card and the coin to M_2 without order, but in contrast M_2 can determine the order of reading the two cards.

   Trick:
   If it's tail, then M_2 announces the two cards in non-descending order. Otherwise, M_2 announces the card in descending order, when coin is head.

   To see the relation with matching, we denote $(a, b) =' a \geqslant b'$ and $(b, a) =' b > a'$



2. A magic on matching (**Full Version**)

   Let M_1, M_2 be two magicians and V be a volunteer.

   - M_2 offers a deck of 52 cards to v
   - V picks 5 cards from the deck and gives them to M_2
   - M_2 announces 4 of the cards
   - M_1 announces the fifth card

   Trick:

   - Consider the bipartite $G = X \sqcup Y$, where $X = \binom{[52]}{5}$, Y is the sequences of 4 cards from $[52]$. Clearly $|X| = \binom{52}{5}$ and $|Y| = 52 \cdot 51 \cdot 50 \cdot 49$
   - Join $A \in X$ to $\vec{b} \in Y$, if $\vec{b}$ is a sequence of 4 elements of A
   - Notice that X is $120(:= 5 \cdot 4!)$ regular and Y is $48(:= 52 - 4)$ regular
   - **Claim:** There is an X-saturated perfect matching

Proof. Let $S \subset X$, $N(S) \subset Y$
# of edges from S to $N(S)$ is $120 \cdot |S| \leqslant$ # of edges that can reach $N(S)$
where number of edges that can reach $N(S)$ is $|N(S)| \cdot 48$
Then we have $|N(S)| \geqslant \frac{120}{48}|S| \geqslant |S|$. By Hall's theorem, there exists an X-perfect matching. □

3. An **intersecting family** of sets in a family of sets $\mathcal{F}$ such that for any $A, B \in \mathcal{F}$, $|A \cap B| \neq 0$

   - It's natural to wonder if $\mathcal{F} \subset P([n])$, then how large can $|\mathcal{F}|$ be ?
   - Some educational guesses
     - Consider $\mathcal{F}_1 = \{A \cup \{1\} : A \subseteq P([n-1])\}$ Any two sets have intersection 1, which is admissible, hence the size is 2^{n-1}
     - Consider $\mathcal{F}_2 = \binom{[n]}{\geqslant \frac{n}{2}+1}$, which are all the subsets with size at least $\frac{n}{2}+1$. By pigeonhole, any two must intersects, also has size 2^{n-1}.
     - Consider $\mathcal{F}_3 = \{A \in P([n]) :$ either $\{1,2\} \subseteq A$ or $\{2,3\} \subseteq A$ or $\{1,3\} \subseteq A\}$, it's clear that it's admissible, hence the size is $|\mathcal{F}_3| = \frac{4}{8} \cdot 2^n = 2^{n-1}$
       To see why the count is this, there are four cases, $\{1,2,3,\dots\}$, $\{1,\dots\}$, $\{2,\dots\}$, $\{3,\dots\}$, each has probability $\frac{1}{8}$
   - **Claim:** Any admissible $\mathcal{F} \subseteq P([n])$, $|\mathcal{F}| \leqslant 2^{n-1}$

     Proof. Let $A \subseteq P([n])$ and partition into $\{A, A^c\}$. If $\mathcal{F} \geqslant 2^{n-1}+1$, then we would have $A, A^c \in \mathcal{F}$, but $A \cap A^c = \varnothing \rightarrow\leftarrow$ □
   - If $\mathcal{F} \subseteq \binom{[n]}{k}$
     - $\{A \in \binom{[n]}{k} : 1 \in A\}$, it has size $\binom{n-1}{k-1}$ We pick the rest $k-1$ empty entry from $n-1$ candidates.
     - $\{A \in \binom{[n]}{k} : k \geqslant \frac{n}{2}+1\}$, since all sets in such family are admissible, it has size $\binom{[n]}{k}$
     - $\{A \in \binom{[n]}{k} : A \cap \{1,2,3\} = \{1,2\}$ or $\{1,3\}$ or $\{2,3\}\}$, it has size $3 \cdot \binom{n-3}{k-2} = \Theta(n^{k-2})$, which is way smaller than $\binom{n-1}{k-1} = \Theta(n^{k-1})$ if $k \ll n$.

4. (Erdos-Ko-Rado)

Theorem 4.1 (Erdos-Ko-Rado Theorem)

If $k < \frac{n}{2}$, then any intersecting family $\mathcal{F} \subseteq \binom{[n]}{k}$, the size of such intersecting family

$$|\mathcal{F}| \leqslant \binom{n-1}{k-1}$$

Proof. Let $\mathcal{F}$ be an intersecting family. $N = [\sigma(1), \sigma(2), \dots, \sigma(n)]$ be the cyclic order.
Denote a **contiguous segment of length k** as

$$\{\sigma(i), \sigma(i+1), \dots, \sigma(i+(k-1))\}$$

which is exactly the image of cyclic interval under permutation.
Denote $\mathcal{F}_N = \{A \in \mathcal{F} : \exists i \ s.t. \ A = \{\sigma(i), \dots, \sigma(i+(k-1))\}\}$
Remark: For B_i, B_{i+k}, $B_i \cap B_{i+k} = \varnothing$,
where $B_i = \{\sigma(i), \dots, \sigma(i+k-1)\}$, $B_{i+1} = \{\sigma(i+k), \dots, \sigma(i+k+(k-1))\}$
So for tuple (B_i, B_{i+k}), at most one can be contained in $\mathcal{F}_N$
Define $X_A = \begin{cases} 1 & \text{if } A \in \mathcal{F}_N \\ 0 & \text{otherwise} \end{cases}$

Then $|X| = \sum_{A \in \mathcal{F}} X_A = |\mathcal{F}_N|$

Take the expect value on both sides,

$$E[X] = E[\sum_{A \in \mathcal{F}} X_A] = \sum_{A \in \mathcal{F}} E[X_A]$$

where $E[X_A] = \Pr[\text{A is contiguous in random permutation}] = \frac{\#\text{ number of interval position sets}}{\#\text{ number of total position sets}}$

To see why $E[X_A] = \frac{n}{\binom{n}{k}}$, total possible position sets is $\binom{n}{k}$ and there are exactly n cyclic intervals.

Hence $\Pr[X_A = 1] = \frac{n}{\binom{n}{k}}$

Therefore $E[X] = |\mathcal{F}| \cdot \frac{n}{\binom{n}{k}}$

Claim: $|\mathcal{F}_N| \leqslant k$

Proof. We can pick at most one from each tuple $(B_{-(k-1)}, B_1), (B_{-(k-2)}, B_2), \ldots, (B_{-1}, B_{k-1})$

So the maximum size is B_0 plus one from each pair, which gives $1 + (k-1) = k$ □

□

5. An intuition about the probability and the covering diagram

   Consider the total space $\binom{[n]}{k}$, we uniformly cover $\binom{[n]}{k}$ with sets of size n.

   'uniform' means each point in the target set is covered by same number of sets (covered same number of times).

   Moreover each set has size $\leqslant k$, at most k elements from $\mathcal{F}$

   Then $|\mathcal{F}| \leqslant \frac{k}{n} \cdot \binom{n}{k} = \binom{n-1}{k-1}$

6. (Fermat's Little Theorem)

Theorem 4.2 (Fermat's Little Theorem)

Let p be prime, $a \in \mathbb{Z}$, then

$$a^p \equiv a \pmod{p}$$

Proof. It's sufficient to show $\frac{a^p - a}{p} \in \mathbb{Z}$.

Consider the following scenario.

Color p distinct beads on a necklace with a colors, up to rotation symmetry. Then number of ways to color such necklace where no monochromatic necklace occurs is $\frac{a^p - a}{p}$. Clearly $\frac{a^p - a}{p} \in \mathbb{Z}$. □

Remark: A shortcut is to count number of ways without considering the group action, an then we consider the group action by dividing the size of such group.

In this case, without considering rotation symmetry, total number of ways $= a^p - a$. The order of the cyclic group is the size of multiplicative group, which has size p for prime.

7. (Orbit-Stabilizer Theorem)

Theorem 4.3 (Orbit-Stabilizer Theorem)

Let G be a well-defined group with $|G| = n$. Then

$$|\mathit{stabilizer}| \cdot |\mathit{orbit}| = n$$

8.

Theorem 4.4 (Burnside's Lemma)

Let G be a finite group acting on a finite set X. For each $g \in G$, define

$$\mathrm{Fix}(g) = \{x \in X \mid g \cdot x = x\}.$$

Then the number of distinct orbits of X under the action of G is

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |\mathrm{Fix}(g)|.$$

9. (A more generalize case)

When n is any positive integer, we recall burnside's Lemma, to count number of ways up to rotation symmetry. Hence we deduce the following.

Theorem 4.5

Let n be any positive integer. We color n distinct beads on a necklace with a colors. Then

$$\#\textit{ number of ways to color } n \textit{ beads with } a \textit{ colors} = \frac{1}{|G|} \cdot \sum_{k=0}^{n-1} a^{gcd(n,k)}$$

Remark: It's sufficient to find number of cycles of length k in n for each k.

Proof. Hence we are looking for the smallest t such that $tk \equiv 0 \pmod{n}$

Then for each $k \in 0, 1, \ldots, n-1$, length of the correspond cycle is $gcd(n, k)$

By Burnside's Lemma,

$$\#\text{ ways} = \frac{1}{n} \cdot \sum_{k=0}^{n-1} a^{gcd(n,k)}$$

□

We consider two special cases:

- $n = p$ where p is prime.
  Then # of ways $= \frac{1}{p} \cdot (a^p + a + \cdots + a) = \frac{1}{p} \cdot (a^p + (p-1) \cdot a) \in \mathbb{Z}$
  Then $p \mid (a^p - a)$ which shows Fermat's Little Theorem
- $n = p \cdot q$, where p, q are distinct primes.
  To count number of k with $gcd(n, k)$, we refer to $\varphi(\frac{n}{gcd(n,k)})$,
  where $\varphi(n) = |\{1 \leqslant k \leqslant n \ : \ gcd(n, k) = 1\}|$
  Then we compute

gcd(n, k)	Count of k	Contribution
pq	1	a^{pq}
p	$\varphi(q) = q - 1$	a^p
q	$\varphi(p) = p - 1$	a^q
1	$\varphi(pq) = (p-1)(q-1)$	a

  Hence # of ways $= \frac{1}{pq} \cdot (a^{pq} + (q-1)a^p + (p-1)a^q + (p-1)(q-1)a)$

10. $\frac{1-\sqrt{(1-4x)}}{2x}$

5 Feb 4th(Probabilistic Preliminary, Ballot, Indep & Lin Indep)

1. Burnside's Lemma (see last time)

2. Consider a $n \times n$ grid, starting at $(0,0)$ and ending at (n,n), in each step, we can either move right 1 unit or move up 1 unit.

   Then # ways $= \binom{2n}{n} = \Theta(\frac{1}{\sqrt{n}} 2^{2n})$

   If we add a constraint additionally, not exceeding the diagonal, then number of ways to reach (n,n) is the **Catalen Numbers**, $C_n := \frac{1}{n+1}\binom{2n}{n}$

   Also we can define it recursively,

   $$C_0 = 1 \text{ and } C_n = \sum_{i=1}^{n} C_i C_{n-i}$$

   Equivalently, we can also consider the generating function for

   $$g(x) = \frac{1 - \sqrt{1-4x}}{2x}$$

3. Consider a necklace with each bead assigned $+1$ or -1.
   A bead is **special** if all clockwise partial sums of the beads are strictly positive

4. (Bertrand's Ballot Theorem)

   Theorem 5.1 (Ballot Theorem)

   If we have a $+1$ and b -1 around the necklace, then

   $$|special\ beads| = max(a-b, 0)$$

   Proof. Notice that if $+1$ is followed immediately by -1, then removing such pair doesn't affect the total number of special beads.

   Keep removing the consecutive pairs until no such -1 exists

   Hence eventually only $+1$ or -1 remains, which gives $max(a-b,0)$ □

5. Application of Catalan Numbers

   Consider a necklace with $(n+1)$ '+1' and n '−1', each necklace has $2n+1$ representations as a string in $\{+1,-1\}^{2n+1}$

   By Ballot Theorem, there is exactly 1 of these strings whose all partial sums are positive

   Then number of catalan walks is $\frac{1}{2n+1}\binom{2n+1}{n+1}$, where $\frac{1}{2n+1}$ is the size of the necklace and number of such string is $\binom{2n+1}{n+1}$

6. Probobilistic Preliminary

   - Markov-inequality
     Let X be nonnegative random variable, then $\Pr[X \geqslant t] \leqslant \frac{\mathbb{E}[t]}{t}$

     Proof. $\mathbb{E}[X] = \sum_a \Pr[X=a] \cdot a = \sum_{t \leqslant a} \Pr[X=a] \cdot a + \sum_{0 \leqslant a \leqslant t} \Pr[X=a] \cdot a \geqslant t \cdot \Pr[X \geqslant a]$
     □

   - Chebyshev-inequality (pairwise indep case)

Proof. Let $\mathbb{E}[X] = \mu$ and $Var[X] = \mathbb{E}[(X-\mu)^2]$
Since $(X-\mu)^2$ is nonnegative, apply Markov Inequality, we have

$$\Pr[(X-\mu)^2 \geqslant t^2] \leqslant \frac{\mathbb{E}[(X-\mu)^2]}{t^2} = \frac{Var[X]}{t^2}$$

Hence we have $\Pr[|X-\mu| \geqslant t] \leqslant \frac{Var[X]}{t^2}$ □

- An example
  Let $X_1, X_2, \ldots, X_n$ be independent with $\Pr[X_i] = 0.1$ and $\Pr[X_i = 0] = 0.9$.
  Denote $X = \sum_i X_i$, then

$$\Pr[X > 0.2n] \leqslant \frac{1}{2}$$

  If we use Chebyshev's Inequality, we have the following:

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = 0.1n \quad \text{and} \quad \mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2]$$

  Using the fact that $(\sum a_i)^2 = \sum_i \sum_j a_i a_j$, we compute the variance:

$$\begin{aligned}
\mathbb{E}\left[\left(\sum X_i - 0.1n\right)^2\right] &= \mathbb{E}\left[\sum_i \sum_j (X_i - 0.1)(X_j - 0.1)\right] \\
&= \sum_i \mathbb{E}[(X_i - 0.1)^2] + \sum_{i \neq j} \mathbb{E}[X_i - 0.1]\mathbb{E}[X_j - 0.1] \\
&= \sum_i \mathrm{Var}[X_i] + 0 \\
&\leqslant n
\end{aligned}$$

  So $\mathrm{Var}[X] \leqslant n$. Then, setting the threshold $t = 0.1n$ for the tail probability:

$$\begin{aligned}
\Pr[|X - 0.1n| > 0.1n] &\leqslant \frac{\mathrm{Var}[X]}{(0.1n)^2} \\
&\leqslant \frac{n}{0.01n^2} \\
&= O\left(\frac{1}{n}\right)
\end{aligned}$$

- To make the previous estimate more precise, we consider 4-wise independent
  Let $\mathbb{E}[X] = \mu$, consider

$$\mathbb{E}[(X-\mu)^4], \Pr[(X-\mu)^4 \geqslant t^2] \leqslant \frac{\mathbb{E}[(X-\mu)^4]}{t^4}$$

  We evaluate $\mathbb{E}[(X-\mu)^4]$ as the following:

$$\begin{aligned}
\mathbb{E}[(X-\mu)^4] &= \mathbb{E}\left[\sum_{i,j,k,l} (X_i - 0.1)(X_j - 0.1)(X_k - 0.1)(X_l - 0.1)\right] \\
&= \sum_i \mathbb{E}[(X_i - 0.1)^4] + \binom{4}{2} \sum_{i<j} \mathbb{E}[(X_i - 0.1)^2 (X_j - 0.1)^2] \\
&= \sum_i \mathbb{E}[(X_i - 0.1)^4] + 6 \sum_{i<j} \mathbb{E}[(X_i - 0.1)^2] \cdot \mathbb{E}[(X_j - 0.1)^2] \\
&\leqslant n + 6n^2
\end{aligned}$$

  Hence $\Pr[|X-\mu| > 0.1n] \leqslant \frac{6n^2}{(0.1n)^4} = O(\frac{1}{n^2})$

7. $X_1, X_2, \ldots, X_k$ are indep if

$$\Pr[X_1 = a_1 \wedge \cdots \wedge X_n = a_n] = \Pi_{i=1}^{n} \Pr[X_i = a_i]$$

Claim: If $X_1, \ldots, X_k$ are independent, then $\mathbb{E}[X_1 \cdot \cdots \cdot X_k] = \Pi_{i=1}^{k} \mathbb{E}[X_i]$

Proof.

$$\begin{aligned}
\mathbb{E}[X_1 \ldots X_k] &= \sum_b \Pr[X_1 \cdot \cdots \cdot X_k = b] \cdot b \\
&= \sum_b \left(\sum_{\substack{a_1,\ldots,a_k \\ \prod a_i = b}} \Pr[X_1 = a_1 \wedge \cdots \wedge X_k = a_k] \right) \cdot b \\
&= \sum_b \sum_{\substack{a_1,\ldots,a_k \\ \prod a_i = b}} \Pr[X_1 = a_1 \wedge \cdots \wedge X_k = a_k] \cdot \prod a_i \\
&= \prod_{i=1}^{k} \left(\sum_{a_i} \Pr[X_i = a_i] \cdot a_i \right) \\
&= \prod_{i=1}^{k} \mathbb{E}[X_i]
\end{aligned}$$

□

8. Linear-Independence and Independence of Random variables

- Random Variables $X_1, \ldots, X_n$ are k-wise independent if only k of the X_is are independent
- Consider $\mathbb{F}_p$ and $v_1, \ldots, v_n \in \mathbb{F}_p^m$, pick $y \in \mathbb{F}_p^m$ uniformly at random.
  Define $X_i = \langle y, v_i \rangle = \sum_{j=1}^{m} y_j \cdot v_{i,j}$
  Claim: If $v_1, \ldots, v_n$ are linearly independent, then $X_1, \ldots, X_n$ are independent.

  Proof. Let V be the $n \times m$ matrix with rows $v_1, \ldots, v_n$. Denote the set of solutions for a given vector a as:

  $$S_a = \{y \in \mathbb{F}_p^m : Vy = a\}$$

  where $V = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$

  When $a = 0$, $S_0 = \ker(V)$. By the Rank-Nullity Theorem:

  $$\dim(S_0) = m - \operatorname{rank}(V) = m - n$$

  Since $v_1, \ldots, v_n$ are linearly independent, $\operatorname{rank}(V) = n$. Because V maps onto $\mathbb{F}_p^n$, for any $a \in \mathbb{F}_p^n$, the set S_a is a coset of the kernel:

  $$S_a = y^* + S_0 \quad \text{where } Vy^* = a$$

  It follows that the cardinality of each set is $|S_a| = |S_0| = p^{m-n}$.
  To show independence, we calculate the joint probability:

  $$\begin{aligned}
  \Pr[X_1 = a_1 \wedge \cdots \wedge X_n = a_n] &= \Pr[Vy = a] \\
  &= \frac{|\{y : Vy = a\}|}{p^m} \\
  &= \frac{p^{m-n}}{p^m} = \frac{1}{p^n}
  \end{aligned}$$

Since this equals $\prod_{i=1}^{n} \Pr[X_i = a_i] = (\frac{1}{p})^n$, the variables are independent.
(i.e. The overall logic is show X_i are uniform in $\mathbb{F}_p^m$, then it shows X_i are mutual independent) $\square$

- A collection of vectors $v_1, \dots v_n$ is k-wise linear independent if any k of them are linear independent, so if $v_1, \dots, v_n$ are linear independent, then $X_i =< y, v_i >$ are k-wise independent
- How many pairwise linear independent elements of $\mathbb{F}_2^m$?
  Recall that on $\mathbb{F}_2$ can only have coefficient 0 or 1, but clearly if the coefficient is 0, then they are linear dependent
  Then we consider all distinct nonzero vectors in $\mathbb{F}_2$

6 Feb 11(Chernoff's Bound, Bernstein's Trick, Hoeffding Inequality, Counting Triangles)

1. Recall The Estimations
   - $f = O(g) \rightsquigarrow \leqslant$
     $\exists A < \infty$ such that $f(n) \leqslant A \cdot g(n)$
   - $f = \Omega(g) \rightsquigarrow \geqslant$
     $\exists A < \infty$ such that $f(n) \geqslant A \cdot g(n)$
   - $f = o(g) \rightsquigarrow \ll$
     $f < 0.001 \cdot g$ or $f < 0.00001 \cdot g$
     f is less than sufficiently small portion of g.
   - $f = \omega(g) \rightsquigarrow \ll$
     $f > 10^{10} \cdot g$ or $f > 10^{100} \cdot g$
     f is greater than sufficiently large amount of g.
   - $f = \Theta(g) \rightsquigarrow \approx$
2. Let $X_1, \dots, X_n$ are 0,1-valued mutually independent random variables.
   Let $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$, denote $X = \sum_{i=1}^{n} X_i$
   Then $\Pr[|X - np| > t] \leqslant exp(\Omega(\frac{-t^2}{n})) \iff \Pr[|X - np| > \epsilon \cdot n] \leqslant exp(\frac{-\epsilon^2 \cdot n}{3})$
   When $t^2 = \omega(n)$,
   $$exp(\frac{-t^2}{3n}) \to Const$$
   Hence
   $$\Pr[|X - np| > \omega(\sqrt{n})] = o(1)$$
   and
   $$\Pr[|X - np| > \Omega(n)] = exp(-\Omega(n))$$
3. Bernstein's Trick

   Overview

   Bernstein's Trick estimates the upper bound of the probability by using monotone of exponential functions, and then apply Markov-Inequality along with trivial optimization to estimate the smooth function of random variables.

Procedure

Let $z_i = X_i - p, Z = \sum z_i = X - np$, $c > 0$

We compute,

$$\begin{aligned}
\Pr[Z > t] &= \Pr[cZ > ct] \\
&= \Pr[e^{cZ} > e^{ct}] \\
&= \Pr\left[\prod_{i=1}^{n} e^{cz_i} > e^{ct}\right] \\
&\leqslant \frac{\mathbb{E}\left[\prod_{i=1}^{n} e^{cz_i}\right]}{e^{ct}} \\
&= \frac{\prod_{i=1}^{n} \mathbb{E}[e^{cz_i}]}{e^{ct}} \\
&= \frac{\left(pe^{c(1-p)} + (1-p)e^{-cp}\right)^n}{e^{ct}}
\end{aligned}$$

Since the inequality works for arbitrary $c > 0$, we should optimize the choice of c by minimizing the righthand side.

Consider

$$pe^{c(1-p)} + (1-p)e^{-cp} \tag{†}$$

We first consider a special case when $p = \frac{1}{2}$

We have

$$e^{-cp}(pe^c + (1-p)) = \frac{1}{2} \cdot (e^{\frac{c}{2}} + e^{-\frac{c}{2}}) \leqslant e^{\frac{c^2}{2}}$$

Since $e^{\frac{c}{2}} = \sum \frac{(\frac{c}{2})^k}{k!}$ and $e^{-\frac{c}{2}} = \sum \frac{(-\frac{c}{2})^k}{k!}$

Add them together we get

$$\sum_{i=1}^{n} \frac{(\frac{c}{2})2i}{(2i)!} \leqslant \sum_{i=1}^{n} \frac{(\frac{c^2}{2})^i}{i!}$$

Then we have

$$\Pr[Z > t] \leqslant \frac{e^{\frac{c^2}{2}} n}{e^{ct}} = exp\Big(\frac{c^2}{2} \cdot n - ct\Big)$$

We try to maximize $-\frac{c^2}{2} \cdot n + ct$

Take $c = \frac{t}{n}$

Then we have

$$\Pr[Z > t] \leqslant \frac{e^{\frac{t^2}{2n}}}{e^{\frac{t^2}{n}}} = e^{\frac{-t^2}{2n}}$$

Now revisit †, we estimate it by computing the Taylor Expansion for $e^{c(1-p)}$ and e^{-cp}

Then we have

$$p \cdot (1 + c(1-p) + \frac{c^2(1-p)^2}{2!} + \dots)$$

+

$$(1-p)(1 + c(-p)^2 + \cdots + \frac{c^i(-p)^i}{i!} + \dots)$$

We get

$$1 + 0 + \frac{c^2 p(1-p)}{2!} + \frac{c^3(p(1-p))(p^2 - (1-p)^2)}{3!} + \dots$$

Sadly, the Taylor Expansion Method doesn't work out but motivates our estimation

For $c \in [1, 2]$, we have

$$e^c \leqslant 1 + c + c^2$$

If $c \leqslant 2$, then

$$\begin{aligned} e^{-cp}(1 + p(e^c - 1)) &\leqslant e^{-cp}(1 + cp + c^2 p) \\ &\leqslant e^{-cp} \cdot e^{cp + c^2 p} \\ &= e^{c^2 p} \end{aligned}$$

Then

$$\Pr[Z > t] \leqslant \frac{e^{c^2 pn}}{e^{ct}}$$

Take $c = \frac{t}{2pn}$, if $t \leqslant 2pn$, we have

$$\Pr[Z > t] \leqslant e^{-t^2} 4pn \leqslant e^{-t^2} 2n$$

If $t > 2pn$, then

$$\begin{aligned} \Pr[Z > t] &\leqslant \Pr[Z > 2pn] \\ &\leqslant \exp\left(\frac{-4p^2 n^2}{4pn}\right) \\ &\leqslant e^{-pn} \\ &\leqslant e^{-\frac{t^2}{2n}} \end{aligned}$$

4. **Erdős–Rényi Random Graph**

Theorem 6.1 (Erdős–Rényi)

*The **Erdős–Rényi** random graph, denoted by $G(n, p)$,*
is a probability space over graphs with n vertices where each possible edge occurs independently with probability p.
The probability of obtaining a specific graph $\mathcal{G}$ with $|E|$ edges is given by:

$$P(\mathcal{G}) = p^{|E|}(1 - p)^{\binom{n}{2} - |E|}$$

5. Hoeffding Inequality

   Let $X_1, \ldots, X_n$ be mutually independent 0,1-valued random variables

   Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$

   For any $t > 0$, we have

$$\Pr[\sum_{i=1}^{n} a_i \cdot X_i] \leqslant exp(-\frac{t^2}{2||a||_2^2})$$

6. Counting Triangles on Random Graph

   We might consider the following questions:

   (1) What's the distribution of the number of triangles in $G(n, p)$ approximately

   (2) Estimate what's the number of triangles with (high) probability P_i

   Set-Up

For each $\{i,j\}$, we have $z_{ij} = \begin{cases} 1 & \text{input is p} \\ 0 & \text{input is } (1-p) \end{cases}$

For each $\{i,j,k\} \in \binom{[n]}{3}$, we have

$$X_{i,j,k} = \begin{cases} 1 & \text{ij,ik,jk all are edges} \\ 0 & \text{otherwise} \end{cases}$$

Then

$$X = \sum_{\{i,j,k\} \in \binom{[n]}{3}} X_{ijk} = \#\text{ of triangles}$$

By linearity,

$$\mathbb{E}[X] = \sum_{ijk} \mathbb{E}[X_{ijk}] = p^3 \cdot \binom{n}{3}$$

To find **Variance**, we directly compute via definition,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[\left(\sum_{\{i,j,k\} \in \binom{[n]}{3}} (X_{ijk} - p^3)\right)^2\right] \\
&= \sum_{\substack{\{i,j,k\} \\ \{i',j',k'\}}} \mathbb{E}\left[(X_{ijk} - p^3)(X_{i'j'k'} - p^3)\right] \\
&= \sum_{\substack{|\{i,j,k\} \cap \\ \{i',j',k'\}| \leqslant 1}} \mathbb{E}[X_{ijk} - p^3]\mathbb{E}[X_{i'j'k'} - p^3] \\
&+ \sum_{\substack{|\{i,j,k\} \cap \\ \{i',j',k'\}| = 2}} \mathbb{E}[X_{ijk} - p^3]\mathbb{E}[X_{i'j'k'} - p^3] \\
&+ \sum_{\substack{|\{i,j,k\} \cap \\ \{i',j',k'\}| = 3}} \mathbb{E}[X_{ijk} - p^3]\mathbb{E}[X_{i'j'k'} - p^3] \\
&= \sum_{\substack{|\{i,j,k\} \cap \\ \{i',j',k'\}| = 2}} \mathbb{E}[X_{ijk} - p^3]\mathbb{E}[X_{i'j'k'} - p^3] \\
&+ \sum_{\substack{|\{i,j,k\} \cap \\ \{i',j',k'\}| = 3}} \mathbb{E}[X_{ijk} - p^3]\mathbb{E}[X_{i'j'k'} - p^3] \\
&= \Theta(n^4)(p^5 - p^6) + \Theta(n^3)(p^3 - p^6) \qquad (\dagger)
\end{aligned}$$

Remark: When the two triangles do not share any edges, they are linearly independent, which gives the covariance is 0, hence the variance is 0

To see why † is the case, we compute the expect value,

$$\begin{aligned}
\mathbb{E}[X_{ijk} - p^3]\mathbb{E}[X_{i'j'k'} - p^3] &= \mathbb{E}[X_{ijk} \cdot X_{ijk}] - \mathbb{E}[p^3 X_{ijk}] - \mathbb{E}[p^3 X_{ijk}] + \mathbb{E}[p^6] \\
&= p^5 - p^6 - p^6 + p^6 \\
&= p^5 - p^6
\end{aligned}$$

For the size of the summation, we attain it via direct counting.

- When the intersection has size 1, there are $\binom{n}{5}$ such pairs, which gives

$$\#\text{ pairs} = \Theta(n^5)$$

- When the intersection has size 2, there are $6 \cdot \binom{n}{4}$ such pairs, which gives

$$\#\text{ pairs} = \Theta(n^4)$$

- When the intersection has size 3, there are $\binom{n}{3}$ such pairs, which gives

$$\#\text{ pairs} = \Theta(n^3)$$

Then

$$\Pr\Big[|X - \binom{n}{3}| > t\Big] \leqslant \frac{\text{Var(X)}}{t^2} = O\Big(\frac{n^4p^5 + n^3p^3}{t^2}\Big)$$

In particular, when $p = \frac{1}{2}$, we have

$$\Pr\Big[|X - \binom{n}{3}| > t\Big] \leqslant O\Big(\frac{n^4}{t^2}\Big)$$

Notice that for $t = \omega(n^2)$, $O(\frac{n^4}{t^2}) = o(1)$

So

$$X \in \Big[\binom{n}{3} \cdot \frac{1}{8} - \omega(n^2)\ ,\ \binom{n}{3} \cdot \frac{1}{8} + \omega(n^2)\Big]$$

has probability $1 - o(1)$, which is very likely to happen

7. Compare with the adjacency matrix approach
Recall that let $(A_G)_{n\times n}$ be adjacency matrix of graph G, then

$$\begin{aligned}\#\text{ of length 3 closed walk} &= \#\text{ of triangles} \\ &= \frac{1}{6} \cdot \text{trace}(A^3)\end{aligned}$$

We apply the randomness on the entries of the matrix