Entropy

I/ Introduction

1) Definition and properties

The entropy measures the amount of randomness in a distribution.

Let X be a discrete random variable. Its entropy is

$$\boxed{H(X) = \sum_{x\in\Omega} -\mathbb{P}(X=x)\log_2\left(\mathbb{P}(X=x)\right).}$$

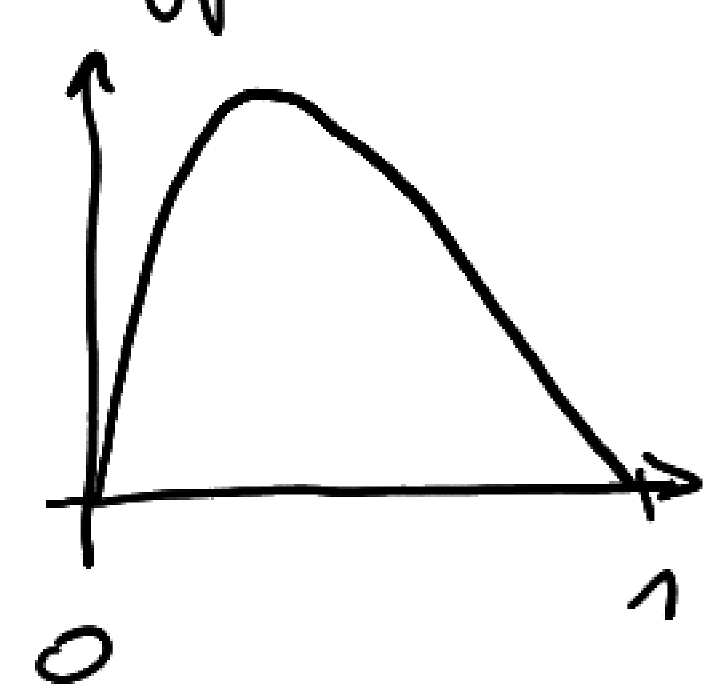
where Ω is the support of X i.e. $\{x : \mathbb{P}(X=x)>0\}$.

- $H(X)$ measures the amount of surprises in the randomness of X.
- Alternatively, $H(X)$ measures the amount of information learned when seeing X. (Shannon source coding theorem).

Lemma 1: $H(X) \leq |\log_2(\Omega)|$ with equality iff X is uniform.

Proof: Let $f(p) = -p\log_2 p$

f is concave for $p\in[0,1]$.



$$H(X) = \sum_{x\in\Omega} f(\mathbb{P}(X=x))$$

$$\underset{\text{concavity}}{\leq} |\Omega|\, f\left(\frac{1}{|\Omega|}\sum_{x\in\Omega}\mathbb{P}(X=x)\right)$$

$$\leq |\Omega|\, f\left(\frac{1}{|\Omega|}\right) = \log_2|\Omega|. \quad \blacksquare$$

Let X, Y be random variables. Let $Z=(X,Y)$. The joint entropy of X and Y is

$$H(X,Y) := H(Z) = \sum_{x,y} -\mathbb{P}(X=x, Y=y) \log_2 (\mathbb{P}(X=x, Y=y)).$$

We write similarly $H(X_1, \dots X_n)$.

Lemma 2 (Independence). Let X and Y be independent random variables. then $H(X,Y) = H(X) + H(Y)$.

Proof:
$$H(X,Y) = \sum_{xy} -\mathbb{P}(X=x)\mathbb{P}(Y=y)\left[\log_2(\mathbb{P}(X=x)) + \log_2(\mathbb{P}(Y=y))\right]$$
$$= \underbrace{\sum_y \mathbb{P}(Y=y)}_{1} \cdot \sum_x -\mathbb{P}(X=x)\log_2(\mathbb{P}(X=x))$$
$$+ \sum_x \mathbb{P}(X=x) \cdot \sum_y -\mathbb{P}(Y=y)\log_2(\mathbb{P}(Y=y))$$
$$= H(X) + H(Y) \quad \blacksquare$$

Let X and Y be two random variables.

$$H(X|Y) := \mathbb{E}_y[H(X|Y=y)]$$
$$= \sum_y P(Y=y)\, H(X|Y=y)$$
$$= \sum_y P(Y=y) \sum_x -P(X=x)\log_2(P(X=x)).$$

$H(X|Y)$ measures the amount of information of X that is not already contained in Y.

Lemma 3 (Chain rule) $H(X,Y) = H(X) + H(Y|X)$.

Proof: Exercise session.

Observation : • If X and Y are independent $H(Y|X) = H(Y)$

• If $X = f(Y)$ then $H(X|Y) = 0$

Lemma 4 (Subadditivity): $H(X,Y) \leq H(X) + H(Y)$

Proof: Exercise session

Mutual information $I(X,Y) = H(X) + H(Y) - H(X,Y)$
measures the amount of information common to X and Y.

Lemma 5 : $H(X|Y,Z) \leq H(X|Z)$

Proof: $\underbrace{H(X|Y) = H(X,Y) - H(Y)}_{\text{chain rule}} \underbrace{\leq H(X)}_{\text{subadditivity}}$

For every z, we have $H(X|Y, Z=z) \leq H(X|Z=z)$

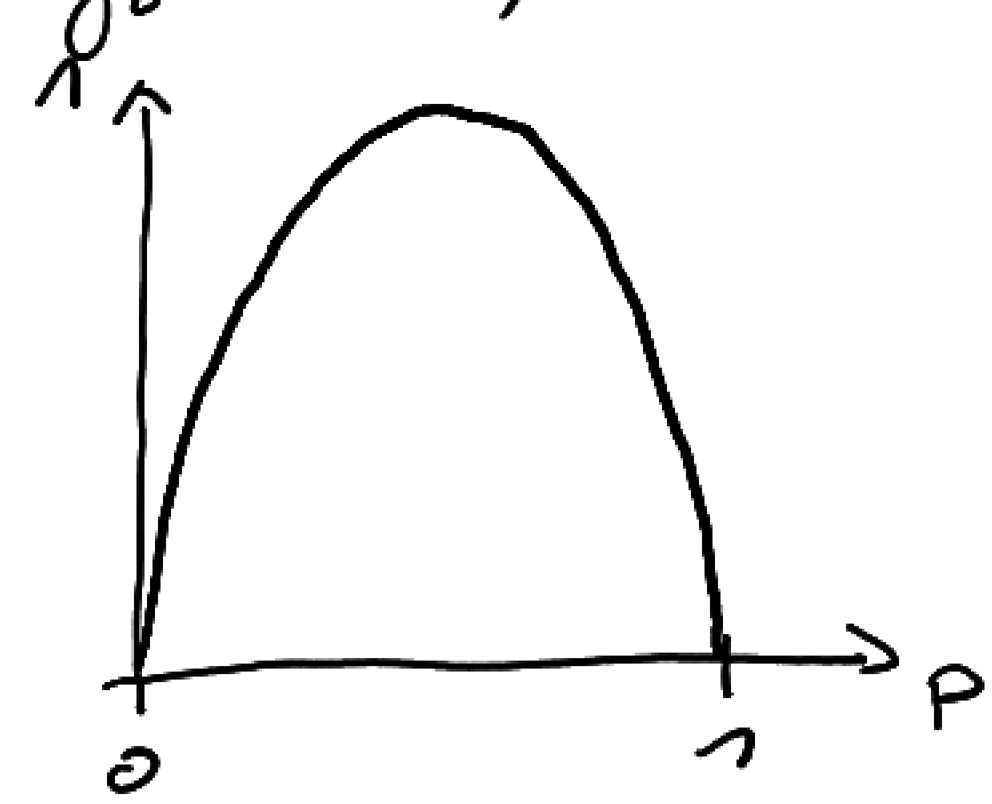
so taking expectation on z, we have $H(X|Y,Z) \leq H(X|Z)$ ∎

2) Usual distribution

Uniform : $X \sim \mathcal{U}(\Omega)$ $H(X) = \log_2(|\Omega|)$

Bernoulli variable : $X \sim B(p)$

$H(p) = H(X) = -p\log p - (1-p)\log(1-p)$



Binomial distribution: $X \sim Bin(n,p)$

$$H(X) = -\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \log_2 \left[\binom{n}{k} p^k (1-p)^{n-k}\right]$$

$$\sim -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \times \log_2 \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] dx$$

(where $\mu = np$ and $\sigma^2 = np(1-p)$)

$$\sim \log_2 (\sqrt{2\pi}\sigma) + \frac{\sigma^2}{2\sigma^2} \log_2 e$$

$$= \frac{1}{2} \log_2 (2\pi e \sigma^2)$$

de Moivre Laplace theorem (special case of the central limit theorem):

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-\mu)^2}{2\sigma^2}}$$

3) Entropy and binomial coefficients

Theorem Let $p \in [0,1]$, $n \in \mathbb{N}^+$ s.t. pn is an integer.

$$\frac{2^{nH(p)}}{n+1} \overset{(1)}{\leq} \binom{n}{np} \overset{(2)}{\leq} 2^{nH(p)}$$

Proof: Immediate if $p=0$ or $p=1$. $\longrightarrow$ We assume $p \in (0,1)$

- $2^{nH(p)} = 2^{-pn\log_2 p - (1-p)n\log_2(1-p)} = p^{-pn} (1-p)^{-(1-p)n}$

We have $\binom{n}{np} p^{pn} (1-p)^{(1-p)n} \leq \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$

$$= (p + (1-p))^n = 1 \quad \text{So } (2).$$

- $\binom{n}{np} 2^{-nH(p)}$ is one of the terms of $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$

we show it is the largest.

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

$$= \binom{n}{k} p^k (1-p)^{n-k} \underbrace{\left[1 - \frac{n-k}{k+1} \cdot \frac{p}{1-p}\right]}_{\geq 0 \text{ iff } k \geq pn-1+p}$$

As there are $n+1$ terms in $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$ this proves (1) ■

This is useful when considering sequences of biased coin tosses i.e. $X \sim Bin(n,p)$ where $p > 1/2$.

By Chernoff bound, $X \sim np$ almost always.

Hence the sequence of coin toss will almost always be one of $\approx \binom{n}{np} \approx 2^{nH(p)}$ sequences, each occurring with probability $\approx p^{np} (1-p)^{n(1-p)} = 2^{-nH(p)}$

Theorem: For $p \in (0, \frac{1}{2}]$ and $n \in \mathbb{N}^+$, where np is an integer,

$$\sum_{i=0}^{pn} \binom{n}{i} \leq 2^{nH(p)} = p^{pn} (1-p)^{(1-p)n}$$

<u>Proof</u>: denote $k = pn$

Let $(X_1, \dots X_n) \in \{0,1\}^n$ chosen uniformly conditioned on $\sum x_i \leq k$.

$$H(X_1, \dots X_n) \underset{\substack{\uparrow \\ \text{uniformity}}}{=} \log_2 \left(\sum_{i=0}^{k} \binom{n}{i} \right) \underset{\text{by subadditivity}}{\leq} H(X_1) + \dots H(X_n)$$

Each $X_i \sim \mathcal{B}(q)$ where $q = \mathbb{P}(X_i = 1 \mid \sum X_i = m \leq k)$

for each $m \leq k$, $\mathbb{P}(X_i = 1 \mid \sum X_i = m) = \frac{m}{n}$, so

$q \leq \frac{k}{n} = p$ Hence $H(X_i) \leq H(p)$.

Hence, $\log_2 \left(\sum_{i=0}^{k} \binom{n}{i} \right) \leq n H(p)$. $\blacksquare$

This gives tails bound on binomial variables: $X \sim \mathcal{B}(n,p)$

For every $q \leq p$, $\mathbb{P}(X \leq qn) = 2^{-n} \sum_{i=0}^{\lfloor qn \rfloor} \binom{n}{i}$

so $\dfrac{\log_2 \mathbb{P}(X \leq qn)}{n} \leq H(q) = -q \log \dfrac{q}{p} - (1-q) \log \dfrac{1-q}{1-p}$

$< 1 \rightsquigarrow \mathbb{P}(X \leq qn) \leq e^{-\lambda(q) n}$

And likewise for $p \leq q \leq 1$:

$$\frac{\log_2 \mathbb{P}(X \geq qn)}{n} \leq H(q)$$

II/ Upper bounds on size of solution space.

1) Maximum number of perfect matchings.

Let G be a balanced bipartite graph on $2n$ vertices.

Denote $\phi(G)$ the number of perfect matchings of G.

We have seen in the exercise session on absorption method that for every $\varepsilon > 0$ and every large enough n,

$$\delta(G) \geq \left(\frac{1}{2} + \varepsilon\right) n \implies \phi(G) \geq \Omega(n)^{n/2}$$

In fact, the stronger result holds:

[Cuckler, Kahn 2009] : $\delta(G) \geq \frac{n}{2} \implies \phi(G) \geq \left(\frac{\delta(G)}{e + o(1)}\right)^{n/2}$

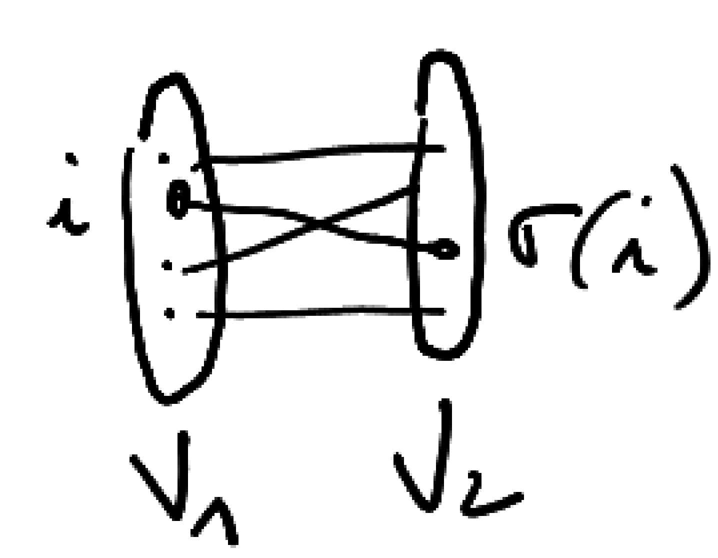
What about upper bounds on $\phi(G)$?

$\phi(G)$ is counted by the permanent of the adjacency matrix of G:

Let $V_1 \sqcup V_2$ be the bipartition of G. Let $A = (a_{ij})$ s.t.

$$a_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ in } V_1 \text{ is adjacent to vertex } j \in V_2 \\ 0 & \text{otherwise.} \end{cases}$$

$$\operatorname{perm}(A) = \sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^{n} a_{i,\sigma(i)}$$



This formula is close to the determinant: $\det(A) = \sum_{\sigma \in \mathfrak{S}_n} (-1)^{\varepsilon(\sigma)} \prod_{i=1}^{n} a_{i,\sigma(i)}$

but without the signature term.

[Conjecture by Minc 1963, Proof by Brégman 1973:]

Let $A = (a_{ij}) \in \{0,1\}^{n \times n}$. Then $\mathrm{perm}(A) \leq \prod_{i=1}^{n} (d_i!)^{1/d_i}$

where $d_i = \sum_{j=1}^{n} a_{ij}$.

Observation: • equality is attained if A is of the following form:

$$\begin{pmatrix} \boxed{1} \updownarrow d_1 & & & & 0 \\ & \boxed{1} \updownarrow d_2 & & & \\ & & \boxed{1} & \ddots & \\ 0 & & & & \boxed{1} \end{pmatrix}$$

i.e. $G = K_{d_1,d_1} \sqcup \dots \sqcup K_{d_t d_t}$

• If G is δ-regular, where $\delta \sim cn$, then

$$\phi(G) \leq \prod_{i=1}^{n} (\delta!)^{1/\delta}$$
$$\leq (\delta!)^{n/\delta}$$

$$\log(\phi(G)) \lesssim \frac{n}{\delta}\left[\log\left(\sqrt{2\pi\delta}\right) + \delta \log\left(\frac{\delta}{e}\right)\right]$$

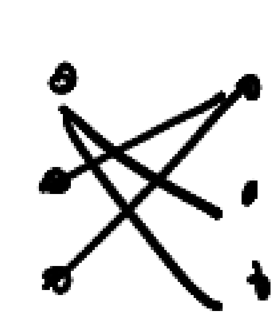
$$\lesssim n \log\left(\frac{\delta}{e}\right) + \frac{n}{\delta}\log\left(\sqrt{2\pi\delta}\right)$$
$$= O\left(n \log\left(\frac{\delta}{e}\right)\right)$$

• A similar bound exists if G is not bipartite:

[Kahn Lovász] $$\phi(G) \leq \prod_{v \in V(G)} (\deg(v))^{1/(2\deg(v))} = \prod_{v \in V(G)} \phi(K_{\deg v, \deg v})^{1/2\deg(v)}$$

Proof: $$\phi(G) = \sqrt{\phi(G \sqcup G)} \leq \sqrt{\phi(G \times K_2)} \quad \blacksquare$$

$G \quad G$



Proof of Brégman (by Radhakrishnan 1997):

Let σ be a random uniform permutation, conditioned on $a_{i\sigma(i)} = 1$ for $i \in [n]$.

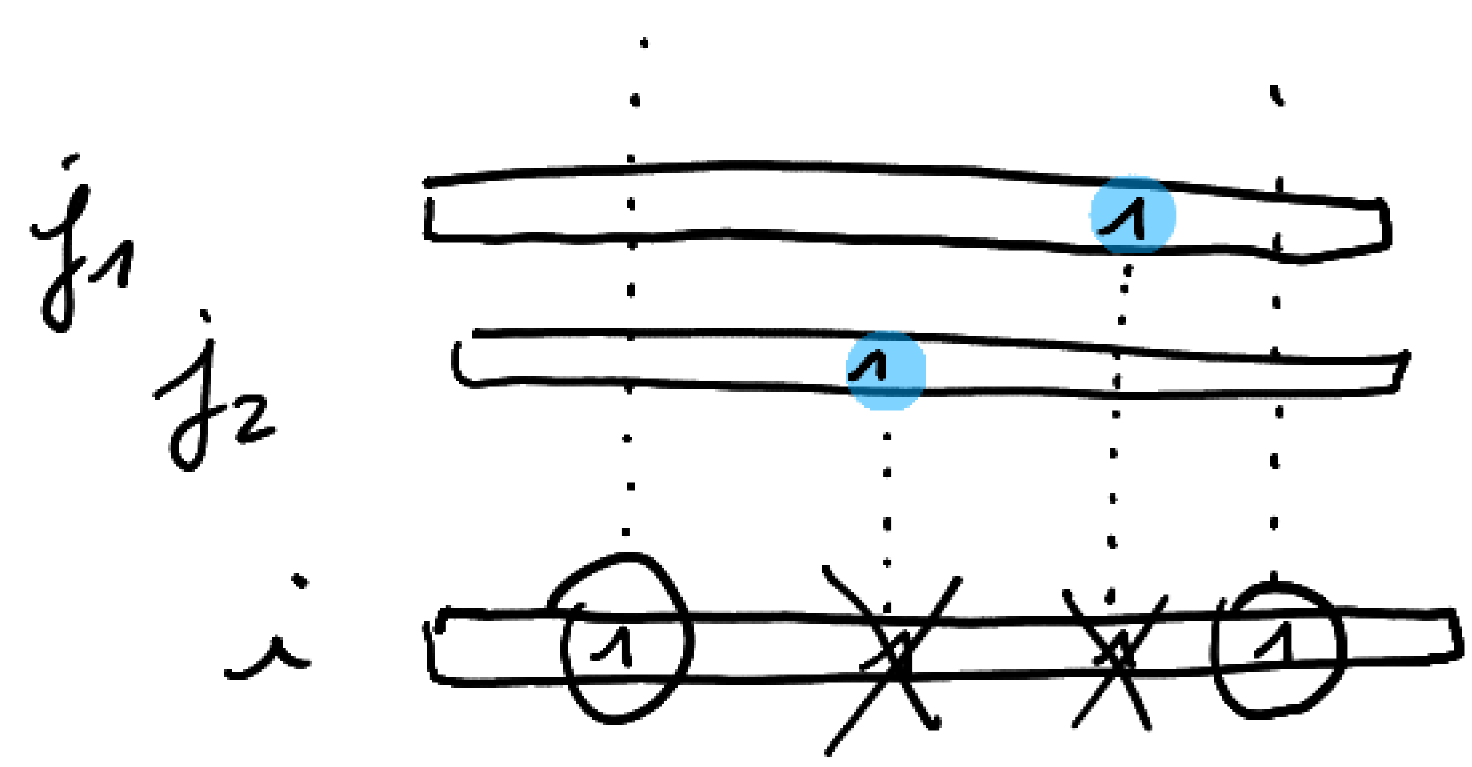
$$H(\sigma) = \log_2(\mathrm{per}(A))$$
$$= H(\sigma(1), \sigma(2), \dots \sigma(n))$$
$$= H(\sigma(1)) + H(\sigma(2) \mid \sigma(1)) + \dots H(\sigma(n) \mid \sigma(1), \dots \sigma(n-1))$$

We have $H(\sigma(i) \mid \sigma(1), \dots \sigma(i-1)) \le H(\sigma(i)) \le \log_2 |\mathrm{support}\ \sigma(i)|$
$\le \log_2 (d_i)$

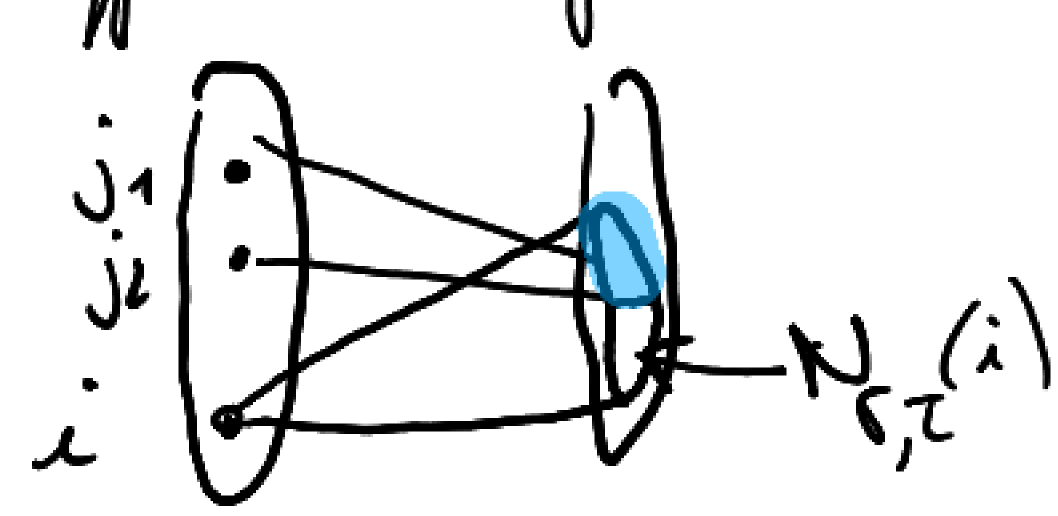
This is not good enough. Key idea: consider the $\sigma(i)$ in a random order.

$\rightsquigarrow$ Let τ be a random permutation of the columns of A.
We say j appears before i if $\tau(j) < \tau(i)$.

- Let $N_{\sigma,\tau}(i) = \#\left\{ k : \begin{array}{l} a_{ik} = 1 \text{ and} \\ \forall j \text{ appearing before } i,\ \sigma(j) \neq k \end{array} \right\}$



j_1, j_2 appear before i



in other words, $N_{\sigma,\tau}(i)$ is the number of available choices to match i knowing $\{\sigma(j) : j \text{ appears before } i\}$.

$\rightarrow$ For every fixed σ, $1 \le N_{\sigma,\tau}(i) \le d_i$ and

Let S be the set of indices s.t. $s \in S \iff a_{is} = 1$

$N_{\sigma,\tau}(i)$ depends on how the rows $\sigma^{-1}(S)$ are ordered compared to i by τ. If there are exactly l rows of $\sigma^{-1}(S)$ before i, then $|N_{\sigma,\tau}(i)| = d_i - l$. Thus

$$\mathbb{P}_\tau\left(|N_{\sigma,\tau}(i)| = k\right) = \frac{1}{d_i} \qquad \forall k \in [d_i]$$

- For any fixed τ, $H(\sigma) = \sum_{i=1}^{m} H(\sigma(i) \mid \{\sigma(j) : j \text{ appears before } i\})$

Chain rule

$$\leq \sum_{i=1}^{m} \mathbb{E}_\sigma \log_2 |N_{\sigma,\tau}(i)|.$$

uniform bound

If τ is a uniform permutation,

$$H(\sigma) \leq \mathbb{E}_\tau\left[\sum_{i=1}^{m} \mathbb{E}_\sigma\left[\log_2 |N_{\sigma,\tau}(i)|\right]\right]$$

$$\leq \mathbb{E}_\sigma\left[\sum_{i=1}^{m} \mathbb{E}_\tau\left[\log_2 |N_{\sigma,\tau}(i)|\right]\right]$$

discrete

$$\leq \mathbb{E}_\sigma\left[\sum_{i=1}^{m} \frac{\log_2(1) + \dots + \log(d_i)}{d_i}\right]$$

$$\leq \sum_{i=1}^{m} \frac{\log_2(d_i!)}{d_i} \qquad \blacksquare$$

2) Application: counting Hamiltonian paths in tournaments

Szele 1943 : the random uniform tournament on n vertices has in expectation $n!/2^{n-1}$ Hamiltonian paths.

Proof: Let $X_{u,n}$ be the number of Hamiltonian paths starting at u. Let's show by induction on n that

$$\mathbb{E}[X_{u,n}] = \frac{(n-1)!}{2^{n-1}}.$$

- this is true for $n=1$; $\mathbb{E}[X_{u,1}] = 1 = \frac{0!}{2^0}$

- $$\mathbb{E}[X_{u,n}] = \sum_{v \neq u} P(u \to v) \cdot \mathbb{E}[X_{v,n-1} \mid u \to v]$$

$$\underset{\text{by induction}}{=} \sum_{v \neq u} \frac{1}{2} \cdot \frac{(n-2)!}{2^{n-2}} = \frac{(n-1)!}{2^{n-1}}$$

the number of Hamiltonian paths of a random transitive tournament is $\sum_u X_{u,n} = \frac{n!}{2^{n-1}}$. ∎

In particular there exists a tournament with this many Hamiltonian paths! How tight is it?

Szele 1943: $O(n!/2^{3n/4})$

Alon 1990: $O(n^{3/2}\, n!/2^n)$

Friedgut & Kahn 2005: $O(n^{3/2-\varepsilon}\, n!/2^n)$ for some small $\varepsilon > 0$

+ Improvements on the constant factors for the lower bound.

Alon 1990: Every n-vertex tournament has at most $O(n^{3/2} n!/2^{n-1})$ Hamiltonian paths.

Proof: We show that every n-vertex tournament has at most $O(\sqrt{n} \cdot n!/2^n)$ Hamiltonian cycles.

Alon's result then follows by observing that:

Lemma: Given an n-vertex tournament with P Hamiltonian paths, there exists an $(n+1)$-vertex tournament with at least $P/4$ Hamiltonian cycles.

Proof of the Lemma: Let T be an n-vertex tournament with P Hamiltonian paths.

Add a vertex v with a random adjacency. Let T' be the $(n+1)$-vertex tournament we obtain this way and X be the number of Hamitonian cycles of T'.

$$\mathbb{E}[X] = \sum_{u,w \neq v} \text{Prob}(u \to v \to w) \cdot \#(\text{Hamiltonian paths of } T \text{ between } u \text{ and } w)$$

$$= \sum_{u,w \neq v} \frac{1}{4} \cdot \#(\text{Hamiltonian paths of } T \text{ between } u \text{ and } w)$$

$$= \frac{P}{4}. \qquad \square$$

- Let $A = (a_{ij})_{1 \le i,j \le n}$ be the $n \times n$ matrix s.t.

$$a_{ij} = \begin{cases} 1 & \text{if } u_i \to u_j. \\ 0 & \text{otherwise} \end{cases}$$

The permanent of A counts the number of partitions into cycles (1-factor).

Hence $\#$ hamiltonian cycles $\leq \text{perm}(A) \underset{\text{Brégman}}{\leq} \prod_{i=1}^{n} (d_i!)^{1/d_i}$ $(*)$

where d_i is $\deg^+(u_i)$

- $g : n \in \mathbb{N} \mapsto (n!)^{1/n}$ is log-concave

i.e. $g(n)\, g(n+2) \leq g(n+1)^2 \quad \forall n \geq 0$

$$\left(\frac{\log g(n) + \log(g(n+2))}{2} \leq \log g(n+1) \right)$$

$\leadsto$ as $\sum_{i=1}^{n} d_i = \binom{n}{2}$, $\prod_{i=1}^{n} (d_i!)^{1/d_i}$ is maximal when all d_i differ by at most 1.

$$\text{perm}(A) \leq \prod_{i=1}^{n} \left(\left\lceil \frac{n-1}{2} \right\rceil ! \right)^{1/\lceil \frac{n-1}{2} \rceil} \underset{\text{Stirling}}{=} O\left(\sqrt{n}\; n!/2^n\right)$$

$\blacksquare$

Last time: $H(X) = \sum_{x \in \Omega} \mathbb{P}(X=x) \log_2 (\mathbb{P}(X=x))$

- $H(X) \leq \log_2 |\Omega|$ with equality iff X is uniform
- Chain rule: $H(X,Y) = H(X) + H(Y|X)$

$$\underset{\text{subadditivity}}{\leq} H(X) + H(Y)$$

(with equality if $X \perp\!\!\!\perp Y$).

Application $\quad \mathrm{per}(A) = \sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^{n} a_{i\sigma(i)}$

Bregman–Minc inequality: $\mathrm{per}(A) \leq \prod_{i=1}^{n} (d_i!)^{1/d_i}$

where d_i is the number of 1's in the i^{th} row.

$\rightsquigarrow$ counting perfect matchings and Hamiltonian paths.

Sketch of proof of Bregman–Minc:

1. Take σ a random uniform permutation conditionned on $a_{i\sigma(i)} = 1 \ \forall i \in [n]$. $\quad \rightsquigarrow \mathrm{per}(A) = 2^{H(\sigma)}$

2. Chain rule: For every ordering τ of the rows of A:

$$H(\sigma) = \sum_{i=1}^{n} \underbrace{H\left(\sigma(i) \,|\, \sigma(j) \text{ for } j <_\tau i\right)}_{\leq \log_2 N_{\sigma,\tau}(i)}$$

3. $H(\sigma) \leq \mathbb{E}_\tau \mathbb{E}_\sigma \sum_i \log_2 N_{\sigma,\tau}(i) = \mathbb{E}_\sigma \left[\mathbb{E}_\tau \sum_i \log_2 N_{\sigma,\tau}(i) \right]$

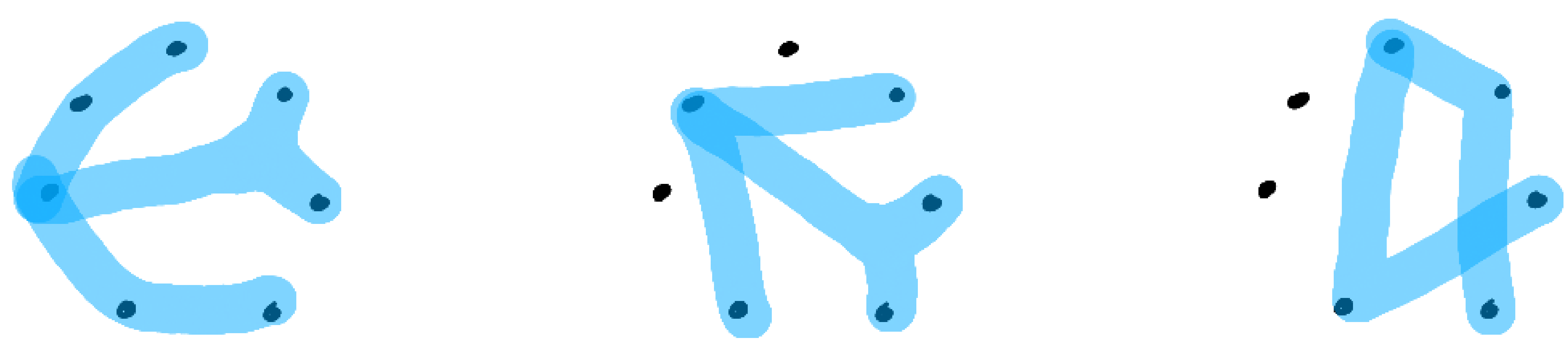
3) Steiner Triple systems

Lecture on Rödl - Nibble method:

(n,k,r) block design: Partition of the r-uniform complete hypergraph on n vertices into r-uniform complete hypergraphs on k vertices.

We saw asymptotics on the packing/covering problems.

Steiner triple systems: $k=3$, $r=2$:
decompose K_n into edge-disjoint triangles.



Steiner triple systems exist iff $n \equiv 1$ or $3 \mod 6$.

How many Steiner triple system are there on n vertices?

Linial Luria 2013: The number of Steiner triple system on n labelled vertices is at most

$$\left(\frac{n}{e^2+o(1)}\right)^{n^2/6}$$

(Keevash proved a matching lower bound when $n \equiv 1,3 \mod 6$).

Proof Let X be a random uniform Steiner triple system on n vertices.

We encode X as follows: $X \in [n]^{\binom{n}{2}}$ and for every

$ij \in E(K_n)$, $X_{ij} = k$ if k is the third vertex of the triangle containing ij.

1. $\log(\#STS(n)) = H(X)$

$$= \sum_{ij \in \binom{n}{2}} H(X_{ij} \mid X_{kl} : kl \leq_\tau ij) \quad (1)$$

for every ordering τ of the edges of K_n

To sample τ, we consider a random uniform real $y_{ij} \in [0,1]$ for every edge and order the edges by increasing value of y_{ij}.

2. Let $N_{X,y}(ij)$ be the number of possibilities for X_{ij}, when the X_{kl} for $kl <_{\tau(y)} ij$ are already revealed.

By uniform bound, for every y,

$$H(X_{ij} \mid X_{kl} : kl <_{\tau(y)} ij) \leq \log_2 N_{X,y}(ij) \quad (2)$$

So $H(X) \leq \mathbb{E}_X \left[\sum_{ij} \log_2 N_{X,y}(ij) \right]$ for every fixed y (3)

By summing (3) over all possible y,

$$H(X) \leq \mathbb{E}_y \left[\mathbb{E}_X \left[\sum_{ij} \log_2 N_{X,y}(ij) \right] \right]$$

$$= \mathbb{E}_X \left[\mathbb{E}_y \left[\sum_{ij} \log_2 N_{X,y}(ij) \right] \right]$$ because X is discrete and finite space.

Write y_{-ij} to denote y with the ij-coordinate removed.

$$H(X) \le \mathbb{E}_X\left[\sum_{ij} \mathbb{E}_{y_{ij}}\left[\mathbb{E}_{y-ij}\left(\log_2 N_{X,y}(ij)\right)\right]\right].$$

3. We compute $\mathbb{E}_{y-ij}\left(\log_2 N_{X,y}(ij)\right)$ where X and y_{ij} are fixed.

Let $k = X_{ij}$ if ki or $kj <_{\tau(y)} ij$,

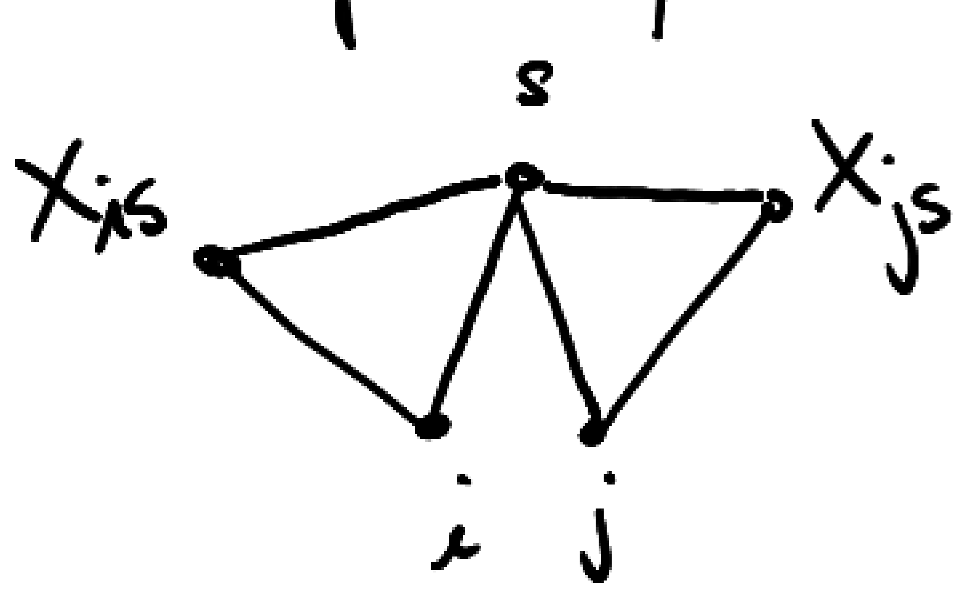
we say that ij does <u>not</u> appear first in its triple and we have $N_{X,y}(ij) = 1$ so $\log_2(N_{X,y}(ij)) = 0$.

$$P(ij \text{ appears first in its triple}) = P(y_{ki} > y_{ij} \text{ and } y_{kj} > y_{ij})$$
$$= (1-y_{ij})^2.$$

so $$\mathbb{E}_{y-ij}\left(\log_2 N_{X,y}(ij)\right) = (1-y_{ij})^2\, \mathbb{E}_{y-ij}\left[\log_2(N_{X,y}(ij)) \,\middle|\, ij \text{ appears first in its triple}\right]$$

$$\underset{\text{convexity of log}}{\le} (1-y_{ij})^2 \log_2\left(\mathbb{E}_{y-ij}\left[N_{X,y}(ij) \,\middle|\, \ldots\right]\right)$$

For every $s \in [n] \setminus \{i,j,k\}$, s is available for ij if none of the following edges is revealed before ij:



so $$\mathbb{E}_{y-ij}\left[N_{X,y}(ij) \,\middle|\, ij \text{ appears first in its triple}\right)$$

$$= \underbrace{1}_{\text{for } k} + \underbrace{(n-3)}_{\text{for } s \in [n]\setminus\{i,j,k\}} (1-y_{ij})^6$$

So $\mathbb{E}_{y_{ij}} \mathbb{E}_{y_{-ij}} \log_2 N_{x,y}(ij) \leq \int_0^1 (1-y_{ij})^2 \log_2 \left(1+(n-3)(1-y_{ij})^6\right) dy_{ij}$

setting $x_{ij} = 1 - y_{ij}$ $\quad \leq \int_0^1 x_{ij}^2 \log_2\left(1+(n-3)\, x_{ij}^6\right) dx_{ij}$

setting $t = x_{ij}^3$ $\quad \leq \frac{1}{3} \int_0^1 \log_2\left(1+(n-3)t^2\right) dt$

$$= \frac{\log_2(n-3)}{3} + \frac{1}{3}\int_0^1 \log_2\left(\frac{1}{n-3} + t^2\right) dt$$

$$\int_0^1 \log_2\left(\frac{1}{n-3} + t^2\right) dt \longrightarrow \int_0^1 \log_2 (t^2)\, dt = -2 \log e.$$

by monotone convergence theorem

so $\mathbb{E}_{y_{ij}} \mathbb{E}_{y_{-ij}} \log_2 N_{x,y}(ij) \leq \dfrac{\log(n/e^2) + o(1)}{3}$

and $H(x) \leq \mathbb{E}_x \sum_{ij} \mathbb{E}_y \log_2 N_{x,y}(ij)$

$$\leq \mathbb{E}_x \binom{n}{2} \frac{\log_2(n/e^2) + o(1)}{3}$$

$$\leq \frac{n^2}{6} \log_2\left(\frac{n}{e^2+o(1)}\right) \qquad \blacksquare$$

Observations: Some intuition on this formula.

- We remove the triangles one at a time. After removing k triangles, there are $\binom{n}{2} - 3k$ edges remaining.

Pretend that the remaining edges are uniformly distributed. Then there are in expectation

$$\binom{n}{3}\left(1-\frac{3k}{\binom{n}{2}}\right)^2 \sim \frac{36}{n^3}\left(\frac{1}{3}\binom{n}{2}-k\right)^3 \text{ triangles to choose from.}$$

If we multiply the above for k ranging in $\left[\frac{1}{3}\binom{n}{2}\right]$, and divide by $\left(\frac{1}{3}\binom{n}{2}\right)!$ to avoid double counting + orders of the triangles, we get roughly the same thing.

(this is the approach followed by Keevash 2018 to prove the existence of STSs)

- Alternatively, suppose we select independently at random $\frac{1}{3}\binom{n}{2}$ triangles in K_n.

  $\mathbb{E}[\#\text{ triangles containing } uv] = 1.$

  So by the Poisson paradigm,

  $$P(uv \text{ belongs to exactly one triangle}) = \frac{1}{e} + o(1).$$

If all these events were independent then the probability that every edge is in exactly triangle $= \left(\frac{1}{e}+o(1)\right)^{\binom{n}{2}}$

So the number of STSs should be

$$\approx \frac{1}{\left(\frac{1}{3}\binom{n}{2}\right)!}\binom{n}{3}^{\frac{1}{3}\binom{n}{2}}\left(\frac{1}{e}+o(1)\right)^{\binom{n}{2}}$$

$$\approx \left(\left(\frac{n^2}{6e}\right)^{-n^2/6}\left(\frac{n^3}{6}\right)^{n^2/6}\left(\frac{1}{e}\right)^{n^2/2}\right)^{1+o(1)}$$

$$\approx \left(\frac{n}{e^2+o(1)}\right)^{n^2/6}$$

4) Shearer's lemma

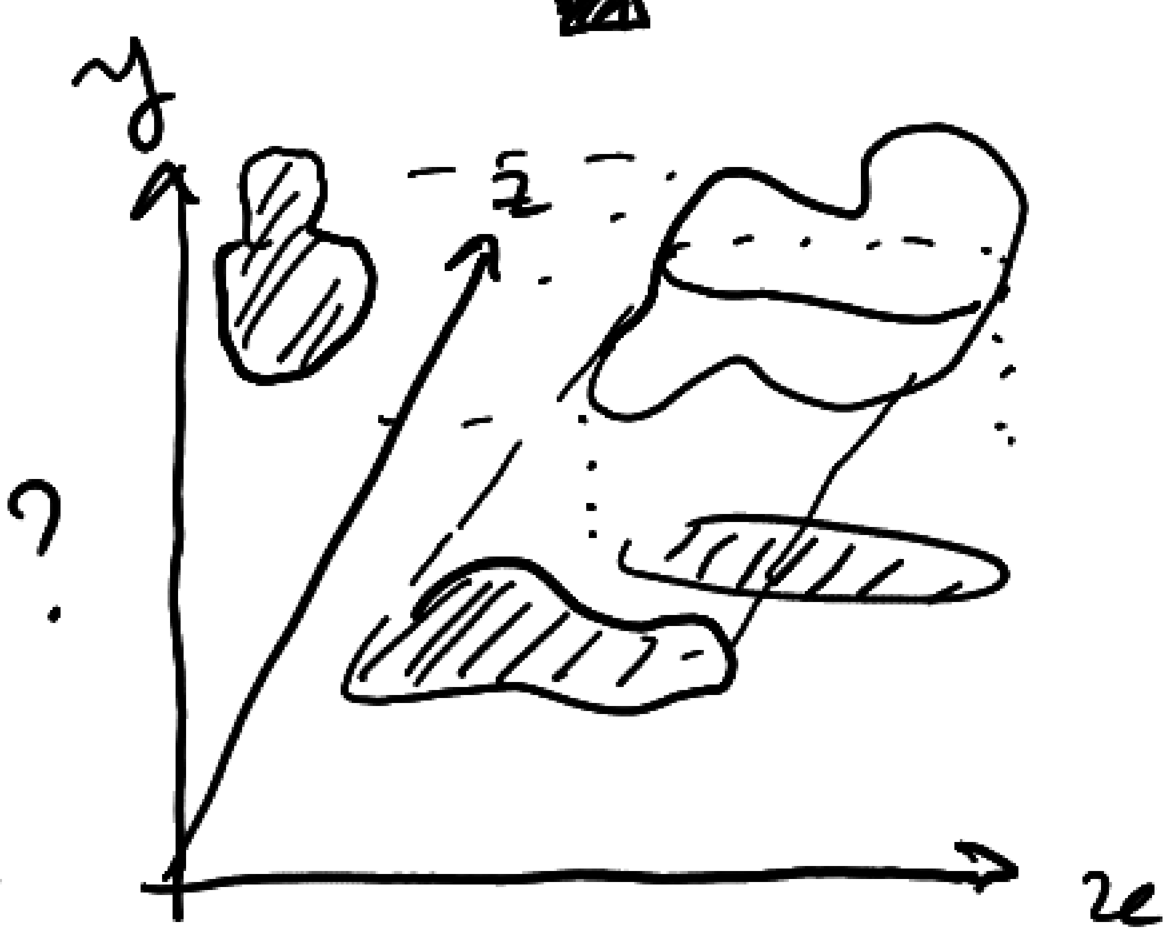
$$H(X,Y) \leq H(X) + H(Y)$$

Generalisation of the subadditivity.

Special case: $2H(X,Y,Z) \leq H(X,Y) + H(X,Z) + H(Y,Z)$

Proof $H(X,Y,Z) = H(X) + H(Y|X) + H(Z|X,Y)$ by chain rule

- $H(X) + H(Y|X) = H(X,Y)$
- $H(X) + H(Z|X,Y) \leq H(X) + H(Z|X) \leq H(X,Z)$ dropping the cond chain rule
- $H(Y|X) + H(Z|X,Y) \leq H(Y) + H(Z|Y) \leq H(Y,Z)$



Question: What is the maximum volume of a body in $\mathbb{R}^3$ s.t. its projection to each of the 3 coordinate plane has area at most 1?

The cube $[0,1]^3$ gives a lower bound of 1.

Theorem: Let $S \subseteq \mathbb{R}^3$ a finite set and $\pi_{xy}(S)$ be its projection on the xy-plane.

$$|S|^2 \leq |\pi_{xy}(S)| \cdot |\pi_{xz}(S)| \cdot |\pi_{yz}(S)| .$$

Proof: Let (X,Y,Z) be a random uniform point of S.

$$2\log_2 |S| = 2H(X,Y,Z) \leq H(X) + H(Y) + H(Z)$$

$$\leq \log_2 \pi_{xy}(S) + \log_2 \pi_{yz}(S) + \log_2 \pi_{xz}(S)$$

By approximating the volume of our body by small cubes of the same size, we get:

Corollary: Let S be a body in $\mathbb{R}^3$.

$$|\mathrm{vol}(S)|^2 \leq \text{area } \pi_{xy}(S) \cdot \text{area } \pi_{xz}(S) \cdot \text{area } \pi_{yz}(S).$$

Shearer's lemma: Let $X_1, \dots X_d$ be discrete random variables and $S_1, \dots S_m \subseteq [d]$ such that each $i \in [d]$ belongs to at least k sets S_j. Then

$$H(X_1, \dots X_d) \leq \frac{1}{k} \sum_{j=1}^{m} H((X_i)_{i \in S_j}).$$

Observation:
- The subadditivity corresponds to the case $k=1$, $S_i = \{i\}$.
- The special case corresponds to the case $S_1 = \{1,2\}$, $S_2 = \{1,3\}$, $S_3 = \{2,3\}$, $k=2$

Proof: Similar to the special case. $\blacksquare$

Loomis & Whitney 1949: Denote π_i the projection along the i^{th} coordinate. For every finite $S \subseteq \mathbb{R}^n$,

$$|S|^{n-1} \leq \prod_{i=1}^{n} |\pi_i(S)|.$$

Corollary: Let $S_1, \dots S_s \subseteq \Omega$ where each $i \in \Omega$ appears in at least k sets S_j. Then for every family $\mathcal{F}$ of subsets of Ω,

$$|\mathcal{F}|^k \leq \prod_{j \in [s]} |\mathcal{F}_{|A_j}| \quad \text{where } \mathcal{F}_{|A_j} = \{F \cap A_j : F \in \mathcal{F}\}.$$

Application 1: Triangle-intersecting families:

Let $\mathcal{G}$ be a collection of graphs on the same vertex set $[n]$.
We say $\mathcal{G}$ is triangle-intersecting if $\forall G_1, G_2 \in \mathcal{G}$, $G_1 \cap G_2$ contains a triangle.

Question: What is the largest triangle intersecting family on n vertices?

Observations:

- Similar to Erdős-Ko-Rado which bounds the maximum size of a family of k-elements subsets of $[n]$ that are pairwise intersecting
- The set of all graphs containing a triangle: $\dfrac{2^{\binom{n}{2}}}{8}$

  $\rightsquigarrow$ Conjecture (Simonovits-Sós 1976): this is optimal
- Any family of size $> \dfrac{2^{\binom{n}{2}}}{2}$ contains a graph and its complement.
- We will get an upper bound of $\dfrac{2^{\binom{n}{2}}}{4}$ using Shearer's lemma.
- Simonovits-Sós conjecture is true [Ellis, Filmus, Friedgut 2012].

Chung, Graham, Frankl and Shearer 1986: Every triangle intersecting family on n vertices has size $< 2^{\binom{n}{2}-2}$.

Proof: Let $\mathcal{G}$ be a triangle intersecting family. We view each $G \in \mathcal{G}$ as a subset of edges of K_n.

For $S \subseteq [n]$ with $|S| = \lfloor n/2 \rfloor$, let A_S be the graph composed of a clique on S and a disjoint clique on $\bar{S}$.

S $\bar{S}$

For every S, every triangle of K_n has an edge in S or in $\overline{S}$.

So $\mathcal{G}_{|A_S}$ is an intersecting family, so $|\mathcal{G}_{|A_S}| \leq 2^{|A_S|-1}$

Let $r = |A_S| = \binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} \leq \frac{1}{2}\binom{n}{2}$

Each edge of K_n belongs to at least $k = \frac{r}{\binom{n}{2}} \binom{n}{\lfloor n/2 \rfloor}$ different A_S.

So by Shearer's lemma, $|\mathcal{G}|^k \leq \prod_{A_S} |\mathcal{G}_{|A_S}| \leq \left(2^{r-1}\right)^{\binom{n}{\lfloor n/2 \rfloor}}$

So $|\mathcal{G}| \leq 2^{\binom{n}{2} - \frac{\binom{n}{2}}{r}} \leq 2^{\binom{n}{2} - 2}$ ■