1.5 Conditional versions of basic properties

Def.
- X, Y, Z, X_i disc. r.v.
- E some event
- f a deterministic function
- range$(X|E)$ = set of values $X|E$ takes

- **Maximality** $H(X|E) \leq \log|\text{range}(X|E)|$

  "=" holds iff $X|E \sim \text{range}(X|E)$

- **Chain rule** $H(X_1, \cdots, X_n|Y) = H(X_1|Y) + H(X_2|X_1,Y) + \cdots + H(X_n|X_1, \cdots, X_{n-1}, Y)$

- **Monotonicity**
  - if $X = f(Y)$, then $H(X) \leq H(Y)$
  - if $X = f(Y,Z) \Rightarrow H(X|Z) \leq H(Y|Z)$
  - if $Y = f(Z) \Rightarrow H(X|Z) \leq H(X|Y)$

- **Dropping Conditioning** $H(Z|Y,X) \leq H(Z|Y)$

- **Subadditivity** $H(X_1, \cdots, X_n | Y)$

  $\leq H(X_1|Y) + H(X_2|Y) + \cdots + H(X_n|Y)$

- **Shearer's lemma** Given random $F \subseteq [n]$ w/ $\Pr(i \in F) \geq \mu$, for $\forall i \in [n]$. then

$$H(X_1, \dots, X_n | Y) \leq \frac{1}{\mu} \mathbb{E}_F H(X_F | Y).$$

- **Gibbs inequality** X, Y take value over same set A, write $p_a = \Pr(X=a)$, $q_a = \Pr(Y=a)$

$$\Rightarrow \quad H(X) = \sum_{a \in A} p_a \cdot \log \frac{1}{p_a} \leq \sum_{a \in A} p_a \cdot \log \frac{1}{q_a}$$

where "=" holds iff $p_a = q_a$ $\forall a \in A$.

Equivalently, Kullback-Leibler divergence or relative entropy is non-negative:

$$D_{KL}(X \| Y) = \sum_{a \in A} p_a \log \frac{p_a}{q_a} \geq 0$$

§ 2 Applications.

2.1 Volume of Hamming balls.

Consider vectors in $\{0,1\}^n$, Hamming distance between $\underline{x}, \underline{y}$ is # coord. where $x_i \neq y_i$

Hamming ball of radius r around $\underline{x}$ $= \{$ all $\underline{y} \in \{0,1\}^n : \Delta(\underline{x}, \underline{y}) \leq r \}$

radius r

Hamming ball around $\underline{0}$ = $\{\underline{x} : |\underline{x}| \le r\}$

$\Uparrow$

$\#$ non-zero coord.

$$\mathrm{vol}(B(r)) = \sum_{i=0}^{r} \binom{n}{i}$$

Def The binary entropy function $h : [0,1] \to \mathbb{R}$

$$h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$

- $h(p) = H(X)$, $X \sim \mathrm{Binom}(1, p)$
- $h(p) \nearrow (0, \tfrac{1}{2})$, $h(\tfrac{1}{2}) = \max = 1$, $\searrow (\tfrac{1}{2}, 1)$

Stirling's formula: $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

$$\Rightarrow \binom{n}{pn} \sim \frac{2^{h(p)\cdot n}}{\sqrt{2\pi n\, p(1-p)}}$$

Thm 2.2 Let $0 < p \le \tfrac{1}{2}$. Then $\forall n \in \mathbb{N}$,

$$\sum_{i \le pn} \binom{n}{i} \le 2^{h(p)\cdot n}$$

quite tight

Pf: • Let A = collectⁿ of all subsets of $[n]$ of size $\le pn$, $|A| = \sum_{i \le pn} \binom{n}{i}$

- Take $X \sim A \overset{\text{maximality}}{\Rightarrow} H(X) = \log |A|$

Suffices to show

$$\underline{H(X) \leq h(p) \cdot n}$$

- Now think of $X = (X_1, \dots, X_n)$ where

$X_i = \mathbb{1}_{\{i \in X\}}$

NTS $H(X_1) \leq h(p)$

$$H(X) = H(X_1, \dots, X_n) \overset{\text{subadd.}}{\leq} \sum_{i \in [n]} H(X_i) = n \underline{\underline{H(X_1)}}$$

Let $\alpha_1 = \Pr(1 \in X) \leq p \leq \frac{1}{2}$

$X_1 \sim \text{Binom}(1, \alpha_1)$

$H(X_1) = h(\alpha_1) \leq h(p)$ ∎

Exer 2.3 (Chernoff using Thm 2.2)

Let $X \sim \text{Binom}(n, \frac{1}{2})$ w/ standard deviation $\sigma = \frac{\sqrt{n}}{2}$.

$$\Rightarrow \forall c \geq 0, \quad \Pr\left(\left|X - \frac{n}{2}\right| \geq c\sigma\right) \leq 2^{-\frac{c^2}{2}+1}$$

2.3 Triangle maximisation

Q : Given that a graph G has m edges,

What is max # Δs it can have ?

Intuitively more compact $\Rightarrow$ more Δs.

Say $e(G) = m = \binom{k}{2}$, $k \in \mathbb{N}$

$$\#\Delta s = \binom{k}{3} \approx \frac{(2m)^{3/2}}{6}$$

Thm 2.5 An m-edge graph has at most $\frac{(2m)^{3/2}}{6}$ triangles.

Rmk. Thm 2.5 is an easy consequence of Kruskal-Katona.

Pf: • Let $T = \{$ all labeled Δs in $G \}$

$t = |T|$, NTS $t \leq (2m)^{3/2}$

Take $X = (X_1, X_2, X_3) \sim T$, where X_i's are vxs of the random Δ X.

$$\xRightarrow{\text{max.}} \quad H(X) = \log t \overset{\text{NTS}}{\underline{\underline{\leq}}} \frac{3}{2} \log(2m)$$

• Let $F \sim \binom{[3]}{2} \Rightarrow \Pr(i \in F) = \frac{2}{3} = \mu$
$i \in \{1, 2, 3\}$

$$\xRightarrow{\text{Shearer's lem}} \quad H(X) \leq \frac{3}{2} \mathbb{E}_F H(X_F)$$

$\forall$ choice of value for F, X_F takes values in the set of all labeled (ordered) edges

$$\overset{\max}{\Longrightarrow} \quad H(X_F) \le \log(2m) \qquad ▨$$

§2.4 Δ-intersecting family.

Def A family of graphs $\mathcal{G}$ on $[n]$ is Δ-intersecting

if $\forall\, G, G' \in \mathcal{F}$, G & G' shares ≥ 1 Δ.

- Ellis-Filmus-Friedgut: Δ-int. fam. on $[n]$ size $\le 2^{\binom{n}{2}-3}$.

  tight: lower bd constr. fix $\underset{1\quad 2}{\overset{3}{\Delta}}$, take all graphs containing this triangle.

Thm 2.6 $\forall$ Δ-int. fam $\mathcal{F}$ on $[n]$ size $\le 2^{\binom{n}{2}-2}$

Pf: • Let $X \sim \mathcal{F}$,

NTS $H(X) = \log|\mathcal{F}| \le \binom{n}{2} - 2$

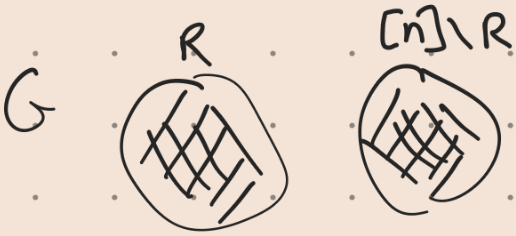
- Think of X as the random vector

  $$\left(X_e : e \in \binom{[n]}{2}\right), \quad \text{where } X_e = \mathbb{1}\{e \in X\}$$

  the set of coordinates

- Consider $R \sim 2^{[n]}$ (unif random subset of $[n]$)

- Let G be the random graph (the random set of coord., w/ which we apply Shearer's lem.)



$$\mathbb{E}_G(e(G)) = \frac{1}{2}\binom{n}{2} \quad \& \quad \Pr(e\in G)=\mu=\frac{1}{2} \quad \forall e \in \binom{[n]}{2}$$

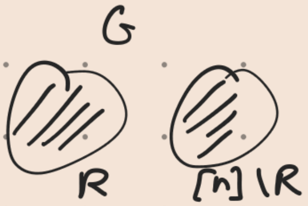
Shearer's lem.

$$\Longrightarrow H(X)=H\left(X_e : e\in\binom{[n]}{2}\right)$$

$$\le \frac{1}{1/2}\,\mathbb{E}_G\,\underline{H(X_G)} \ ?$$

- $\mathcal{F}|_G$ is an intersecting family.

$\forall F, F' \in \mathcal{F}$, $E(F)\cap E(F')$ contains a Δ.
& this Δ intersects G.



- X_G takes values in $\mathcal{F}|_G$

$$\left|\mathcal{F}|_G\right| \underset{\text{intersecting}}{\le} 2^{e(G)-1}$$

$$\xRightarrow{\text{max.}} \quad H(X_G) \le \log\left|\mathcal{F}|_G\right| = e(G)-1$$

- Finally $H(X) \le 2\,\mathbb{E}_G\, H(X_G)$

$$\le 2\,\mathbb{E}_G\,(e(G)-1) = \binom{n}{2}-2 \quad \blacksquare$$

§ 1.4 Axioms determine entropy funct. uniquely

We show that

Lem 1.17 Let X be a r.v. defined over A.

If $H(X)$ satisfies all six axioms

$$\Rightarrow H(X) = \sum_{a \in A} p_a \cdot \log \frac{1}{p_a} \quad , \quad \text{where } p_a = \Pr(X=a).$$

Lem 1.16 If $X \sim A \Rightarrow H(X) = \log |A|$.

Pf:
- Consider the special case $n = |A| = 2^k$

Take $Y \sim [2] \xRightarrow{\text{normalisation}} H(Y) = 1$

Take i.i.d $Y_1, \dots, Y_k$ of Y.

$$\xRightarrow{\text{indep}} H(Y_1, \dots, Y_k) = k H(Y) = k$$

$$(Y_1, \dots, Y_k) \sim [2]^k \cong [n] \sim X$$

$$\xRightarrow{\text{invariance}} H(X) = H(Y_1, \dots, Y_k) = k = \log n = \log |A|.$$

- General case, let $\delta = H(X) - \log n$

$X \sim A$

$|A| = n$

WTS $\delta = 0$.

Take i.i.d. copies $X_1, \dots, X_r$ of X.

$\overset{\text{indep}}{\Longrightarrow} \quad \underline{H(X_1, \dots, X_r)} = r\,H(X)$

- if $\underline{2^k \le n^r} \le 2^{k+1}$

& the special 2^k-case $\overset{\text{Lem 1.3}}{\Longrightarrow} \quad k \le r \cdot H(X) \le k+1$

$$\Longrightarrow \quad \overset{\ominus}{\underline{\frac{k}{r} - \log n}} \le \delta \le \overset{\oplus}{\underline{\frac{k+1}{r} - \log n}}$$

$$\Longrightarrow \quad |\delta| \le \oplus - \ominus = \frac{1}{r} \longrightarrow 0 \quad \text{as } r \to \infty$$

▨