Recap

Jensen's: For convex f, $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$

Entropy: $H(X) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

$$0 \leq H(X) \leq \log n$$

Prefix-free codes:

$$H(X) \underbrace{\leq}_{\text{for all } C} \mathbb{E}|C(X)| \underbrace{\leq}_{\text{for Shannon code}} H(X) + 1$$

Joint entropy

$$Z = (X, Y)$$

$$H(Z) = H(X, Y)$$

$$= \sum_{x,y} p(x,y) \cdot \log \frac{1}{p(x,y)}$$

$$= \sum_{x,y} p(x) \cdot p(y|x) \cdot \log \frac{1}{p(x) \cdot p(y|x)}$$

$$= \sum_{x} p(x) \underbrace{\left(\sum_{y} p(y|x) \right)}_{1} \cdot \log \frac{1}{p(x)} + \sum_{x} p(x) \cdot \underbrace{\sum_{y} p(y|x) \cdot \log \frac{1}{p(y|x)}}_{\underset{x}{E}\, H(Y|X=x)}$$

$$= H(X) + H(Y|X)$$

$$\underset{x}{E}\, H(Y|X=x) = H(Y|X)$$

Chain rule for entropy

$$H(X,Y) = H(X) + H(Y|X)$$

$$= \mathbb{E}_x\left[H(Y|X=x)\right]$$

Ex: $H(X_1, X_2 | Y) = H(X_1 | Y) + H(X_2 | Y, X_1)$

$$H(X_1, X_2, \ldots, X_m) = H(X_1) + H(X_2 \ldots X_m | X_1)$$

$$= H(X_1) + H(X_2 | X_1) + H(X_3 \ldots X_m | X_1, X_2)$$

$$\vdots$$

$$= H(X_1) + H(X_2 | X_1) + \cdots + H(X_m | X_1 \ldots X_{m-1})$$

An example

$$H(X,Y) = H(X) + H(Y|X)$$

$$= H(Y) + H(X|Y)$$

$$(X,Y) = \begin{cases} 01 & \text{w.p. } 1/3 \\ 10 & \text{w.p. } 1/3 \\ 11 & \text{w.p. } 1/3 \end{cases}$$

$$H(X,Y) = \log_2 3$$

$$H(X) = H(X,Y) - H(Y|X) = H(X)$$

$$\underline{H(Y)} = \frac{1}{3}\log 3 + \frac{2}{3}(\log 3 - 1) = \underline{\log 3} - \frac{2}{3}$$

$$H(X,Y) - H(X|Y)$$

$$\underline{H(Y|X=0)} = 0$$

$$H(Y|X=1) = \underline{1}$$

$$\underline{H(Y|X)} = \frac{1}{3}\cdot 0 + \frac{2}{3}\cdot 1 = \frac{2}{3}$$

Conditioning reduces entropy on average

- $\underbrace{H(Y \mid X)}_{\text{conditional entropy}} \leq H(Y)$

Proof: $\mathbb{E}_{x} H(Y \mid X = x) \leq H(Y)$

$$\sum_{x} p(x) \sum_{y} p(y|x) \log \frac{1}{p(y|x)} - \sum_{y} p(y) \log \frac{1}{p(y)}$$

$$\sum_{y,x} p(x,y) \log \frac{1}{p(y)}$$

$$= \sum_{x,y} p(x,y) \log \frac{p(y)\, p(x)}{p(x,y)}$$

$$Z = \frac{p(y) \cdot p(x)}{p(x,y)}$$

$$= \mathbb{E} \log Z \leq \log \left(\sum \cancel{p(x,y)} \cdot \frac{p(y)\, p(x)}{\cancel{p(x,y)}} \right) \leq 0$$

w.p. $p(x,y)$

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

- Use that log is strictly concave to show

  $H(Y|X) = H(Y)$ if and only if X, Y are independent

$$H(X_1, X_2 \ldots \ldots X_m) = H(X_1) + H(X_2|X_1) + \cdots + H(X_m|X_1 \ldots X_{m-1})$$

$$\leq H(X_2) \qquad \leq H(X_3) \qquad \ldots \leq H(X_m)$$

$$\leq H(X_1) + \cdots + H(X_m)$$

equality when $X_1 \ldots X_m$ are independent

Improved Source Coding

$$\exists\ C : \mathcal{X}^m \longrightarrow \{0,1\}^* \text{ s.t. } \mathbb{E}_{x_1 \ldots x_m}\left[C(x_1 \ldots x_m) \right] \leq m \cdot H(X) + 1$$

Proof: Use Shannon code for $Z = (X_1, \ldots, X_m)$

$$\mathbb{E}\left| C(X_1 \ldots X_m) \right| \leq H(X_1 \ldots X_m) + 1$$

$$= H(X_1) + \cdots + H(X_m) + 1$$

$$= m \cdot H(X) + 1$$

Binary entropy

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases} \qquad H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} = H_2(p)$$

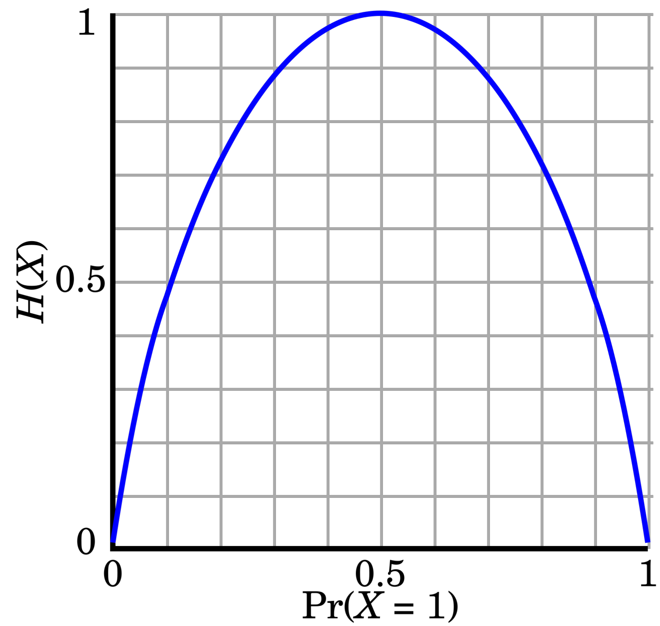
- $H_2(p) \leq 1 = H_2(1/2)$

- $H_2(p)$ is concave

To show: $\alpha \cdot H_2(p_1) + (1-\alpha) \cdot H_2(p_2) \leq H_2(\alpha \cdot p_1 + (1-\alpha) \cdot p_2)$

$$X = \begin{cases} 1 & \text{wp } \alpha \\ 0 & \text{wp } 1-\alpha \end{cases} \qquad \underline{Y} \mid X=1 = \begin{cases} 1 & \text{wp } p_1 \\ 0 & \text{wp } 1-p_1 \end{cases}$$

$$Y \mid X=0 = \begin{cases} 1 & \text{wp } p_2 \\ 0 & \text{wp } 1-p_2 \end{cases}$$



(Source: Wikipedia)

Ex - $H_2(p)$ is increasing for $p \leq 1/2$

Application: Estimating binomial sums

Estimate: $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k}$ (for $k \le \frac{n}{2}$)

equivalently, for $S = \{(x_1, \ldots, x_n) \in \{0,1\}^n \mid \sum x_i \le k\}$

estimate $|S|$

$S = \{ (x_1 \ldots x_n) \in \{0,1\}^n \mid \sum x_i \leq k \}$

Let $Z = (X_1 \ldots X_n)$ be uniform on S

$$\log|S| = H(X_1 \ldots X_n)$$

$$\leq H(X_1) + \cdots + H(X_n)$$

$$= n \cdot H(X_1)$$

$$\leq n \cdot H_2\left(\frac{k}{n}\right)$$

$$\therefore |S| \leq 2^{n \cdot H_2(k/n)}$$

$n \cdot \mathbb{E}[X_1]$

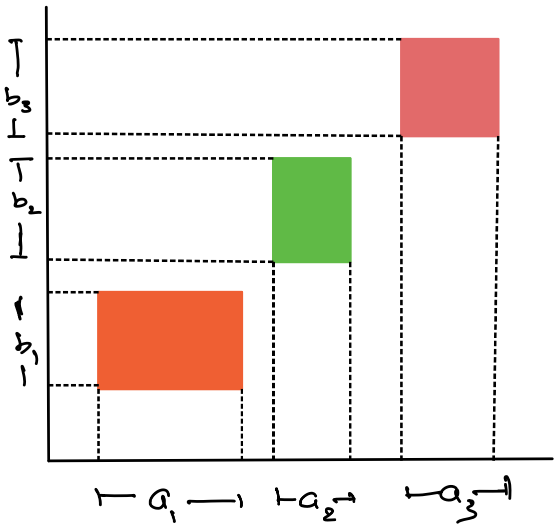
$= \mathbb{E}[X_1 + \cdots + X_n] \leq k$

$\Rightarrow \mathbb{E}[X_1] \leq \frac{k}{n}$

$\therefore H(X_1) \leq H_2\left(\frac{k}{n}\right)$

Counting Points

$a_1, \ldots\ldots a_n \in \mathbb{N}$
$b_1 \ldots\ldots b_n \in \mathbb{N}$



- Consider n disjoint rectangles of dimensions $a_1 \times b_1, \ldots\ldots, a_n \times b_n$
- Pick R w.p. proportional to area
$$\mathbb{P}[R = r_i] = \frac{a_i \cdot b_i}{\sum_j a_j \cdot b_j}$$
- Pick two independent points (X_1, Y_1), $(X_2, Y_2) \in R$

$$\left(\sum a_i b_i\right)^2 \le \left(\sum a_i^2\right)\left(\sum b_i^2\right)$$

$$2 \log\left(\sum a_i b_i\right) = H(X_1, Y_1, R) + H(X_2, Y_2, R)$$

$$= 2H(R) + H(X_1, Y_1 | R) + H(X_2, Y_2 | R)$$

$$= 2H(R) + H(X_1|R) + H(Y_1|R) + H(X_2|R) + H(Y_2|R)$$

$$= \underbrace{H(X_1, X_2, R)}_{\le \log\left(\sum a_i^2\right)} + \underbrace{H(Y_1, Y_2, R)}_{\le \log\left(\sum b_i^2\right)}$$

Hypergraphs, entropy and inequalities [Friedgut 04]

e.g. $\sum a_i b_i \leq \left(\sum a_i^{1/\theta}\right)^{\theta} \left(\sum b_i^{1/1-\theta}\right)^{1-\theta}$ [Hölder's ineq]

$$\left(\mathrm{Tr}(ABC)\right)^2 \leq \mathrm{Tr}(AA^T) \cdot \mathrm{Tr}(BB^T) \cdot \mathrm{Tr}(CC^T)$$

$\vdots$

Generalizing sub-additivity

$$H(X_1, X_2, X_3) \le H(X_1) + H(X_2) + H(X_3)$$

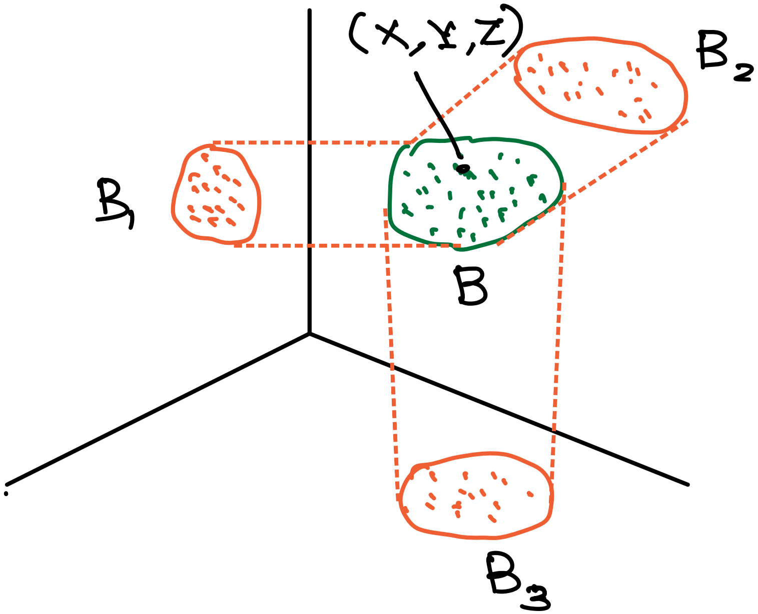
Better bounds using pairs ?

$$H(X_1, X_2, X_3) = H(X_1, X_2) + \overbrace{H(X_3 \mid X_1, X_2)}^{\le H(X_3)}$$

$$H(X_1, X_2, X_3) = H(X_2, X_3) + \underbrace{H(X_1 \mid X_2, X_3)}_{\le H(X_1 \mid X_3)}$$

$$H(X_1, X_2, X_3) \le \frac{1}{2}\left(H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1)\right)$$

Counting 3D points



$$2H(X,Y,Z) \le H(X,Y) + H(Y,Z) + H(Z,X)$$

$$2\log|B| \le \log|B_1| + \log|B_2| + \log|B_3|$$

$$\mathrm{Vol}(B) \le \left(\mathrm{area}(B_1)\cdot \mathrm{area}(B_2)\cdot \mathrm{area}(B_3)\right)^{1/2}$$

Loomis - Whitney

Generalizing subadditivity : Shearer's lemma

- Random vars $X_1, \dots, X_n$. $X_S = \{X_i\}_{i\in S}$

Let $\mathcal{F}$ be a collection of subsets of $[n]$ s.t. each $i \in [n]$ appears in at least t subsets in $\mathcal{F}$

Then. $$t \cdot H(X_1, \dots, X_n) \leq \sum_{S\in\mathcal{F}} H(X_S)$$

- (Distributional version): Distribution D on $2^{[n]}$ s.t. $\forall i \in [n]$ $\mathbb{P}_{S\sim D}[i \in S] \geq \mu$

Then, $$\mu \cdot H(X_1, \dots, X_n) \leq \mathbb{E}_{S\sim D}[H(X_S)]$$

- Ex: Check that second version implies first

Let $S = \{i_1, i_2, \ldots, i_k\}$

$$H(X_S) = H(X_{i_1}) + H(X_{i_2} \mid X_{i_1}) + \cdots + H(X_{i_k} \mid X_{i_1} \ldots X_{i_{k-1}})$$

$$= \sum_{i \in S} H(X_i \mid X_{S \cap [i-1]})$$

$$\geq \sum_{i \in S} H(X_i \mid X_{[i-1]})$$

$$= \sum_{i \in [n]} \mathbb{1}_{\{i \in S\}} \cdot H(X_i \mid X_{[i-1]})$$

To prove: $$\mathop{\mathbb{E}}_{S\sim D}\left[H(X_S)\right] \geq \mu \cdot H(X_1 \dots X_n)$$

$$\mathop{\mathbb{E}}_{S\sim D}\left[H(X_S)\right] \geq \mathop{\mathbb{E}}_{S\sim D}\left[\sum_{i\in[n]} \mathbb{1}_{\{i\in S\}} \cdot H(X_i \mid X_{[i-1]})\right]$$

$$\mathop{\mathbb{E}}_{S\sim D}\left[\mathbb{1}_{\{i\in S\}}\right] \geq \mu$$

$$\geq \mu \cdot \sum_i H(X_i \mid X_{[i-1]})$$

$$= \mu \cdot H(X_1 \dots X_n)$$