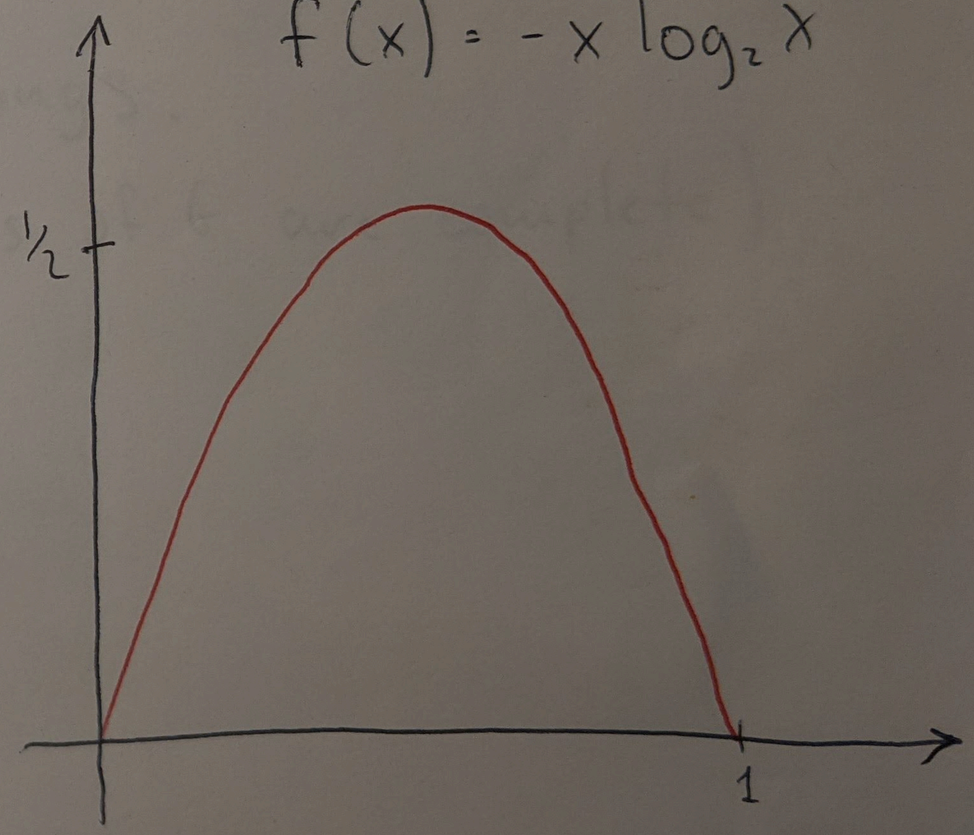
Lecture 16

Entropy

$$H(X) = \sum_{s \in S} -p_s \log_2 p_s$$



9. Entropy.

Let X be a random variable taking values in some Finite set S.

The entropy of X is defined as

$$H(X) = \sum_{x \in S} -p(X=x) \log_2 p(X=x)$$

Suppose X is uniFormly distributed on S, $|S| = k$.

$$H(X) = \log_2 k.$$

You can encode n independent copies of X using

$$n \cdot \log_2 k = n H(X) \text{ bits}$$

(and you can not do much better).

The same is true for general X

Shannon noiseless encoding theorem.

$f(x) = -x \log_2 x = \frac{-x \log x}{\log 2}$

$\frac{\partial}{\partial x} f = -\log x - 1$ f is increasing upto $x = \frac{1}{e}$.

$\frac{\partial^2}{\partial x^2} f = -\frac{1}{x}$. f is concave for $x \geq 0$

Lemma 9.1 (uniform bound): Let X be a random variable supported on S then

$$H(X) \leq \log_2 |S|.$$

with equality iff X is uniform.

Proof:

$$H(X) = \sum_{x \in S} -p(x) \log_2 p(x) = \sum_{x \in S} f(p(x)) \overset{\text{by concavity}}{\leq} |S| \cdot f\left(\sum_{x \in S} \frac{p(x)}{|S|}\right)$$

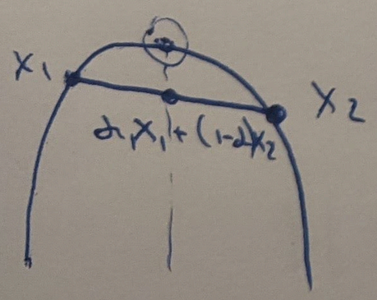
$p(x) = P(X = x)$

$$= |S| f\left(\frac{1}{|S|}\right) = |S| \cdot \frac{1}{|S|} \cdot (-\log_2 |S|) = \log_2 |S|.$$

Holder

- f is concave $\sum \alpha_i = 1$ $\alpha_i \geq 0$

$$\sum \alpha_i f(x_i) \leq f\left(\sum \alpha_i x_i\right)$$ Jensen



If X & Y are two random variables

we write $H(X,Y)$ for entropy of (X,Y)

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

If X & Y are independent then

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

Conditional entropy ([redacted]) $H(Y|X)$

(amount of information you gain from Y given that you already know X.

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

$$= \sum_x \mathbb{P}(X=x) \sum_y -\mathbb{P}(Y=y|X=x) \log_2 \mathbb{P}(Y=y|X=x)$$

$H(Y|X) = 0$ if and only if. $Y = f(X)$.

Lemma 9.2: $H(Y|X) = H(X,Y) - H(X)$

(Chain rule)

Proof: $H(Y|X) = \sum_x p(x) \sum_y -p(y|x) \cdot \log_2 p(y|x)$

$$= \sum_{x,y} - p(y|x)\,p(x) \cdot \log_2 p(y|x)$$

$$= \sum_{x,y} - p(y,x) \cdot \log_2 \frac{p(y,x)}{p(x)}$$

$= H(X,Y)$

$p(Y=y, X=x)$ ←

$$= \Big(\sum_{x,y} -p(x,y) \log_2 p(x,y) \Big)$$

$$- \sum_{x,y} \big(-p(x,y) \log_2 p(x) \big)$$

$= H(X)$

$$\sum_y p(x,y) = p(x)$$

Corollary: $H(X,Y) \geq H(X)$ for any Y.

Lemma 9.3: $H(X,Y) \leq H(X) + H(Y)$

(Subadditivity)

Proof:

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

$$= \sum_{x,y} \left(-\underbrace{p(x,y)\log_2 p(x)}_{= H(x)} - \underbrace{p(x,y)\log_2 p(y)}_{= H(Y)} + \underbrace{p(x,y)\log_2 p(x,y)}_{= H(X,Y)} \right)$$

$$= \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_{x,y} p(x)\cdot p(y) \cdot \frac{p(x,y)}{p(x)\cdot p(y)} \log_2 \frac{p(x,y)}{p(x)p(y)}$$

$f(x) = -x\log_2 x$

$$= \sum_{x,y} p(x)\cdot p(y) \cdot \left(-f\left(\frac{p(x,y)}{p(x)\cdot p(y)} \right) \right)$$

$$\geq -f\left(\sum_{x,y} \underbrace{p(x)\cdot p(y)}_{=1} \cdot \frac{p(x,y)}{p(x)\cdot p(y)} \right)$$

$$= 0.$$

Corollary 9.4: $H(X_1, X_2, \dots, X_n) \leq H(X_1) + H(X_2) + \dots + H(X_n)$

(by induction from 9.3

Corollary 9.5: $H(Y|X) \leq H(Y)$ → equivalent to 9.3

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

" $\mathbb{E}_z(H(X|Y, Z=z))$ " $\mathbb{E}_z(H(X|Z=z))$

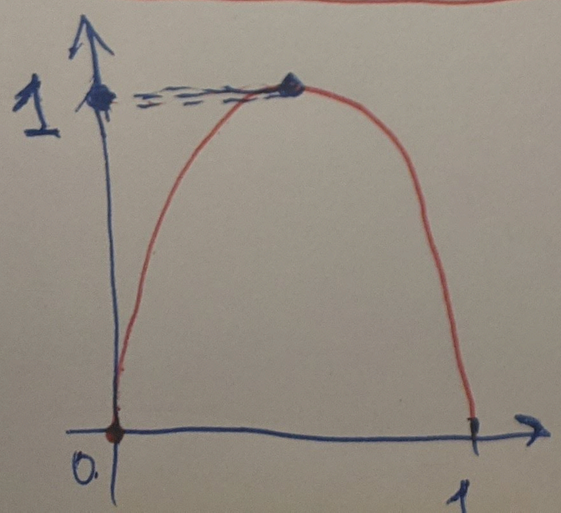
Let X be a Bernoulli random variable:

$\mathbb{P}(X=0) = 1-p \qquad \mathbb{P}(X=1) = p$

$H(X) = -p\log_2 p - (1-p)\log_2(1-p)$

$$h(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

binary entropy function.



Theorem 9.6: If $k \le \frac{n}{2}$ then

$$\sum_{0 \le i \le k} \binom{n}{i} \le 2^{h(k/n) \cdot n}.$$

$$\sum_{0 \le i \le n/3} \binom{n}{i} \le 2^{h(1/3)n}$$

Proof: Let $(X_1, \ldots, X_n) \in \{0,1\}^n$ be chosen conditioned on $X_1 + \ldots + X_n \le k$. uniformly at random.

$$\log_2 \sum_{0 \le i \le k} \binom{n}{i} = H(X_1, \ldots, X_n) \le H(X_1) + H(X_2) + \ldots + H(X_n)$$

$$\le n \cdot h(k/n).$$

$$\mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n] \le k$$

$$\mathbb{E}[X_i] \le k/n.$$

So X_i is Bernoulli random variable with expectation $p \le \frac{k}{n}$.

$$H(X_i) = h(p) \le h(k/n)$$

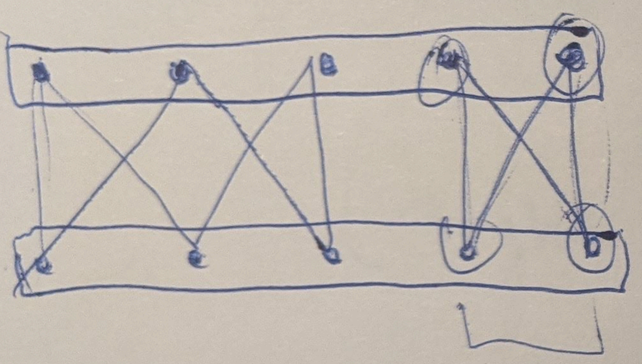
$h(p)$ is increasing for $p \le 1/2$

Another application:

Number of perfect matchings

Let G be a bipartite graph with bipartition (A, B)
$|A| = |B|$.

A perfect matching in G is a set of edges s.t. every vertex is incident to exactly one of them.



$|A| = |B| = \boxed{n}$.

We want upper bounds on the number of perfect matchings in G

$\boxed{n!}$ if G is complete bipartite.

What if all vertices of A have degree $\boxed{d}$? (and in B)

$d = 1$ $\boxed{1}$

$d = 2$ $2^{n/2}$

general d $(d!)^{\frac{n}{d}}$?

Theorem (Brégman): If G is a bipartite graph with bipartition (A, B) and the vertices of A have degrees $d_1, d_2, \ldots, d_n$ then

G has at most $\prod_{i=1}^{n} (d_i!)^{(1/d_i)}$

perfect matchings.

(tight if all of components of G are complete).