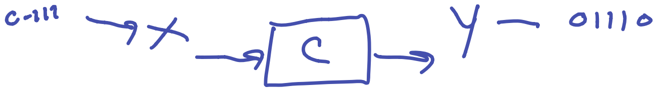
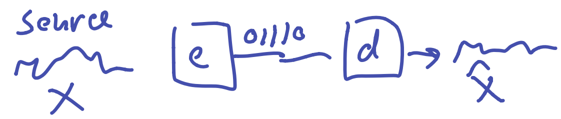
# 12.3 A Bit of information theory




channel coding




source coding

$R(D)$ - bit rate needed for distortion level.

Fano's Inequality

If $X$ can be reconstructed from $Y$ then $I(X;Y)$ must be close to $H(X)$

---

Given a discrete rv $X$ $\in \{1, \dots, m\}$

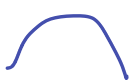
entropy $H(X) = \sum_{i=1}^{m} p_X(i) \log \frac{1}{p_X(i)}$ $\quad u \log \frac{1}{u}$

surprise at seeing $X=i$

$p_X(i) = \underline{\underline{\frac{1}{m}}}$ for all $i$

$H(X) = \log m$ $\quad$ ( $\log_2 2^{10} = 10$ $\quad$ 0111011101 $\quad$ $1, \dots, 1024$ )

$H(X)$ is a concave function of $p_X$



$-u \log u$

$$H(Y|X) \triangleq \sum_i H(Y|X=i)\, p_X(i)$$

$$= \sum_i \left[ \sum_j -p_{Y|X}(j|i) \log p_{Y|X}(j|i) \right] p_X(i)$$

$$= \sum_j \sum_i \quad " \quad\quad "$$

$$\leq -\sum_j \underbrace{\left( \sum_i p_{Y|X}(j|i)\, p_X(i) \right)}_{p_Y(j)} \log \left( \quad \right) = H(Y)$$

Note $\quad H(X,Y) = H(X) + H(Y|X)$ "chain rule" of entropy

$$P_{X,Y}(i,j) = P_X(i)\, P_{Y|X}(j|i)$$

If $X - Y - Z$ is a Markov chain

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

Shannon multual information

$$I(X;Y) = H(X) - H(X|Y)$$

Data processing inequality

If $X - Y - Z$ is a Markov chain

$$I(X;Y) \geq I(X;Z)$$

$$H(X) - \underset{H(X|Y,Z)}{\underset{\shortparallel}{H(X|Y)}} \geq H(X) - H(X|Z)$$

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

---

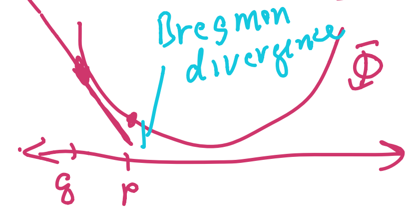
— $D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}$ $\quad \left( \sum_i \left( \frac{p_i}{q_i} \log \frac{p_i}{q_i} \right) q_i \right)$

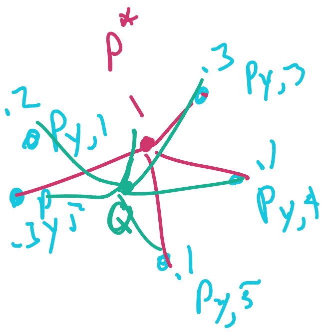
$\sum_i |p_i - q_i|^2$ $\quad \geq 0$ with equality if $p_i \equiv q_i$ $\quad u \log u$

— f-divergence $\; d_f(p\|q) = \sum_i q_i\, f\!\left(\frac{p_i}{q_i}\right)$ $\quad f(u) = u \log u$

~ Bregman divergence $d_B(p\|q) = \Phi(p) - [\Phi(q) + \nabla\Phi(q)(p-q)]$

$\Phi$ convex



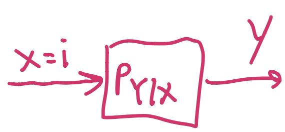

Find $p^*$ to minimize $\sum_i d(p_{Y;i}\|p^*)p_X(i)$

$$p_Y(j) = p^*_j = \sum_i p_X(i)\,\underline{\underline{p_{Y|X}(j|i)}}$$

---

$$I(X;Y) = \sum_i p_X(i)\, D(p_{Y|X}(\cdot|i)\|p_Y)$$

$$= D(p_{Y|X}\|p_Y|p_X)$$




$$= \sum_i\sum_j p_{Y|X}(j|i)\log\left(\frac{p_{Y|X}(j|i)}{p_Y(j)}\right)p_X(i)$$

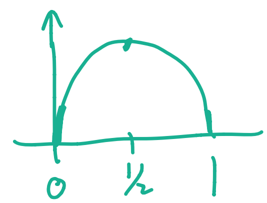
$$= \sum_i\sum_j p_{Y|X}(j|i)\log\left(\frac{p_{Y|X}(j|i)}{Q(j)}\right)p_X(i) + \underbrace{\sum_i\sum_j p_{Y|X}(j|i)\log\left(\frac{Q(j)}{p_Y(j)}\right)p_X(i)}_{\sum_j p_Y(j)\log\frac{Q(j)}{p_Y(j)}}$$

$$= D(p_{Y|X}\|Q|X) - \underbrace{D(p_Y\|Q)}_{\geq 0}$$

so $$I(X;Y) \le D(P_{Y|X} \| Q | X)$$

for any $Q$

$$h(p) = -p\log p - (1-p)\log(1-p)$$
$$= H(Be(p))$$




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

**Lemma 12.2**

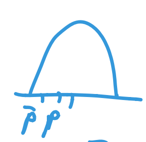
Let $B = (B_1, \cdots, B_N)$ be a random binary vector $(B_i \in \{0,1\})$ such that

$E\left[\sum_{i=1}^{N} B_i\right] \le Np$ for some $p$ with $0 \le p \le 1/2$

Then $H(B) \le Nh(p)$ (i.e. independent + identically distributed bits)

**Proof** Let $p_i = \mathbb{P}(B_i = 1)$

$$H(B) = H(B_1, \cdots, B_n)$$
$$= H(B_1) + H(B_2|B_1) + \cdots + H(B_{n-1}|B_1 \cdots B_{n-2}) + H(B_n|B_1 \cdots B_{n-1})$$
$$\le H(B_1) + H(B_2) + \cdots + H(B_n)$$
$$= h(p_1) + h(p_2) + \cdots + h(p_N)$$
$$\le Nh\left(\frac{p_1 + \cdots + p_N}{N}\right) = Nh(\bar{p}) \le Nh(p)$$

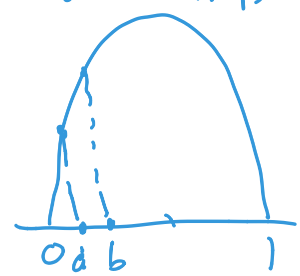



$$\frac{1}{N}\sum_i p_i = \bar{p} \le p$$

Lemma 12.3 $\quad 0 \le a \le 0.25$

$-p\log p - (1-p)\log(1-p) \quad h(p)$

$p\left(\log \frac{1}{p}\right)$




special case:

If $h(a) \ge \frac{1}{2} h(b)$

then $\boxed{a \ge \frac{1}{8} b}$

implies $h(a) \ge \frac{1}{2} b \log \frac{1}{b}$

almost like $a \log \frac{1}{a} \ge \frac{1}{2} b \log \frac{1}{b}$

here $\Rightarrow$ $a \log \frac{1}{a} \ge \frac{1}{4} b \log \frac{1}{b}$

---

$\mathcal{F}$ has VC dimension $\ge V$ if for any $b_1, \cdots b_V$

there exists $x_1, \cdots x_V$

so $(f(x_1), \cdots, f(x_V)) = (b_1, \cdots b_n)$

$\mathcal{F}$ is $(N, D)$ rich if there exists $x_1, \cdots x_N$

so that for any $b_1, \cdots b_N$ with $b_1 + \cdots + b_N = D$

$(f(x_1), \cdots f(x_N)) = (b_1, \cdots\cdots b_N)$

**Claim** $\mathcal{F}$ is $(N, D)$ rich for any

$$0 \le D \le N \le V$$

If $\mathcal{F}$ is $(N, D)$ rich it is $(N-1, D-1)$ rich and $(N-1, D)$ rich (if $N-1 \ge D$)

$D = 3$

$$\begin{array}{cccccccccc|c} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \vdots & & & & & & & & & & \vdots \\ x_1 & & \dots & & & & & & & & x_N \end{array}$$