RWTH AACHEN UNIVERSITY

Recap:

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

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

Mutual information

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

KL divergence

$$D(p\|q) = \sum_{i=1}^{m} p_i \log \frac{p_i}{q_i}$$

$$I(X;Y) = D\big(p(x,y) \,\|\, p(x)\cdot p(y)\big)$$

Consider a channel with input distribution $p = (p_1, \dots, p_m)$

$$p \longrightarrow \boxed{p(y_j|x_i)} \longrightarrow r$$

$X \sim$ input distr. channel output distr. $\sim Y$

Let matrix $W = (w_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,d}} \in \mathbb{R}^{m\times d}$,

W is a stochastic matrix, i.e., $\sum_{j=1}^{d} w_{ij} = 1, \ \forall i = 1,\dots,m$.

Output distribution $r = (r_1, \dots, r_d)$ is obtained as

$$r = pW$$

Thinking the Future
Zukunft denken

Lemma 2.1.13. For any distr. p, q with support $\mathcal{X} = \{x_1, \dots, x_m\}$ and stochastic matrix

$$W = \left(p(y_j | x_i)\right)_{i,j} \in \mathbb{R}^{m \times d},$$

$$D(p \| q) \geq D(pW \| qW). \quad \lrcorner$$

Proof. Use the log-sum inequality (La. 2.1.7)

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j}$$

$$D(p \| q) = \sum_{i=1}^{m} p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{d} \underbrace{p(x_i)\, p(y_j | x_i)}_{a_i} \log \frac{\overbrace{p(x_i)\, p(y_j | x_i)}^{a_i}}{\underbrace{q(x_i)\, p(y_j | x_i)}_{b_i}}$$

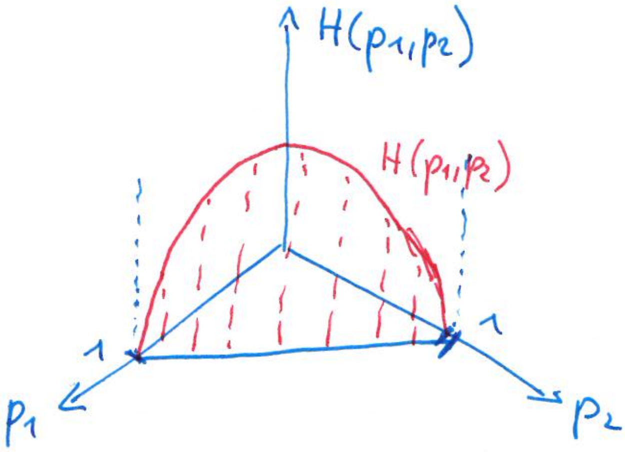
$$\geq \sum_{j=1}^{d} p \underline{w}_j \log \frac{p \underline{w}_j}{q \underline{w}_j} \qquad W = (\underline{w}_1, \dots, \underline{w}_d)$$

$$= D(pW \| qW). \quad \blacksquare$$

Theorem 2.1.14.
$H(p)$ is a concave function of $p = (p_1, \dots, p_m)$.

Figure w.r.t. Th. 2.1.14

2-dim. $p = (p_1, p_2)$, $p_1, p_2 \geq 0$, $p_1 + p_2 = 1$



Proof. Let $u = \left(\frac{1}{m}, \dots, \frac{1}{m}\right)$ be uniform distribution.

$$D(p \| u) = \sum_{i=1}^{m} p_i \log \frac{p_i}{1/m} = \log m - H(p)$$

Hence $H(p) = \log m - D(p\|u)$

By Th. 2.1.12 b)

$$D\left(\lambda p + (1-\lambda) q \,\|\, \lambda u + (1-\lambda) u\right)$$
$$\leq \lambda D(p\|u) + (1-\lambda) D(q\|u), \text{ i.e.,}$$

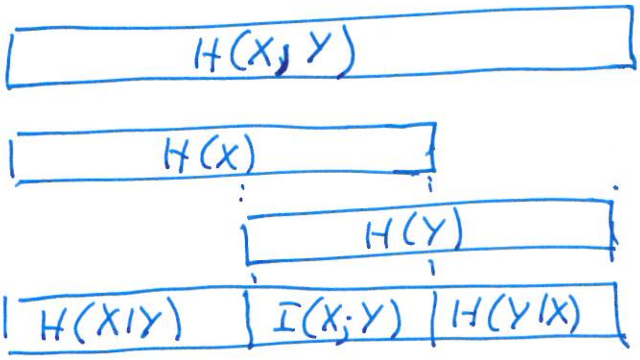
convexity in the first argument.

Thus, $H(p)$ is a concave fct. of p. $\blacksquare$

Thinking the Future
Zukunft denken

RWTH AACHEN UNIVERSITY

Collecting the quantities in a single picture.



Connections are easy to derive, e.g.,

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

2.2. Inequalities

Def. 2.2.1. Random variables X, Y, Z are said to have the Markovian property if the joint pmf satisfies

$$p(x,y,z) = p(x)\, p(y|x)\, p(z|y). \lrcorner$$

~~For~~ Notation: $X \to Y \to Z$.

For $X \to Y \to Z$ the conditional distr. of Z depends only on Y and is conditionally independent of X.

Thinking the Future
Zukunft denken

Lemma 2.2.2.

a) If $X \to Y \to Z$, then
$$p(x,z|y) = p(x|y) \cdot p(z|y).$$

b) If $X \to Y \to Z$, then $Z \to Y \to X$.

c) If $Z = f(Y)$, then $X \to Y \to Z$.

Proof.

a)
$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)\,p(y|x)\,p(z|y)}{p(y)}$$
$$= \frac{p(x,y)\;p(z|y)}{p(y)} = p(x|y)\,p(z|y).$$

b)
$$p(x,y,z) = p(x)\,p(y|x)\,p(z|y)$$
$$= p(x,y)\;\frac{p(z,y)}{p(y)}\;\frac{p(z)}{p(z)}$$
$$= p(z)\,p(y|z)\,p(x|y), \text{ i.e., } Z \to Y \to X.$$

c) $Z = f(Y)$, then
$$p(x,y,z) = \begin{cases} p(x,y), & \text{if } z = f(y) \\ 0, & \text{otherwise} \end{cases}$$

Hence $p(x,y,z) = p(x,y)\,\mathbb{1}(z = f(y)) = p(x)\,p(y|x)\,p(z|y)$,

since $p(z|y) = \begin{cases} 1, & z = f(y) \\ 0, & \text{otherwise} \end{cases}$

Thinking the Future
Zukunft denken

RWTH AACHEN UNIVERSITY

Theorem 2.2.3. (Data-processing inequality)

If $X \to Y \to Z$, then $I(X;Z) \leq \min\{I(X;Y), I(Y;Z)\}$ $\lrcorner$

"No processing of Y can increase the information that Y contains about X."

Proof By the chain rule

$$I(X;Y,Z) = I(X;Z) + \overbrace{I(X;Y|Z)}^{\geq 0}$$
$$= I(X;Y) + \underbrace{I(X;Z|Y)}_{=0}$$

Since X and Z are cond. independent given Y:

$I(X;Z|Y)=0$. Since $I(X;Y|Z) \geq 0$, we have

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

Equality holds iff $I(X;Y|Z)=0$, i.e., $X \to Z \to Y$.

$I(X;Z) \leq I(Y;Z)$ is shown analogously. ∎

Assume X, Y r.v. with support $\mathcal{X} = \{x_1, \dots, x_m\}$.

Define $p_e = P(X \neq Y)$, the "error probability".

Theorem 2.2.4. (Fano inequality)

$$H(X|Y) \leq H(p_e) + p_e \log(m-1).$$

This implies that $p_e \geq \dfrac{H(X|Y) - \log 2}{\log(m-1)}$ ∎

Thinking the Future
Zukunft denken

$$H(p_e) = -p_e \log p_e - (1-p_e)\log(1-p_e)$$

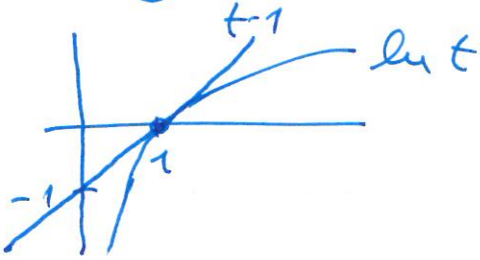
~~Bew.~~ Proof.

(i) $H(X|Y) = \sum_{x \neq y} p(x,y) \log \frac{1}{p(x|y)} + \sum_{x} p(x,x) \log \frac{1}{p(x|x)}$

(ii) $p_e \log(M-1) = \sum_{x \neq y} \log(M-1)\, p(x,y)$

(iii) $H(p_e) = -p_e \log p_e - (1-p_e) \log (1-p_e)$

(iv) $\ln t \leq t-1 \,,\; t \geq 0$



$$H(X|Y) - p_e \log(M-1) - H(p_e)$$

$$\overset{(i),(ii)}{=} \sum_{x \neq y} p(x,y) \log \frac{p_e}{p(x|y)(M-1)} + \sum_{x} p(x,x) \log \frac{(1-p_e)}{p(x|x)}$$

$$\leq (\log e) \left[\sum_{x \neq y} p(x,y) \left(\frac{p_e}{p(x|y)(M-1)} - 1 \right) + \sum_{x} p(x,x) \left(\frac{1-p_e}{p(x|x)} - 1 \right) \right]$$

$$= (\log e) \left[\frac{p_e}{M-1} \sum_{x \neq y} p(y) - \sum_{x \neq y} p(x,y) + (1-p_e) \sum_{x} p(x) - \sum_{x} p(x,x) \right]$$

$$= (\log e) \left[p_e - p_e + (1-p_e) - (1-p_e) \right] = 0 \quad \blacksquare$$