Consequences of Information-inequality-theorem:
$I(X;Y) \geq 0$, $H(X|Y) \leq H(X)$ and

$$H(X_1, \dots, X_n) \leq \sum_{k=1}^{n} H(X_k)$$

and

$$H(X_1, \dots, X_n) = \sum_{k=1}^{n} H(X_k) \iff X_1, \dots, X_n \text{ are independent}$$

since $H(X_1, \dots, X_n) = \sum_{k=1}^{n} H(X_k|X_{k-1}, \dots, X_1)$.

Lemma 0.1 (Log-Sum-inequality)**.** *Let* $a_1, \dots, a_n, b_1, \dots, b_n \geq 0$. *Then*

$$\sum_{k=1}^{n} \left(a_k \log_2 \frac{a_i}{b_i} \right) \geq \left(\sum_{k=1}^{n} a_k \right) \log_2 \frac{\sum_{k=1}^{n} a_k}{\sum_{k=1}^{n} b_k}$$

Equality holds if and only if there exists a c such that $\frac{a_k}{b_k} = c$ *for all* $1 \leq k \leq n$.

Proof. We may assume $a_k > 0$ and $b_k > 0$ for all k. Otherwise a_k would not contribute to the left and right side. Let $A = \sum_{k=1}^{n} a_k$, $B = \sum_{k=1}^{n} b_k$, $p_k = \frac{b_k}{B}$ and $t_k = \frac{a_k}{b_k}$ and $f(t) = t \log_2(t)$. And consider Jensen's inequality

add pic

$$\sum_{k=1}^{n} p_k f(t_k) \geq f(\sum_{k=1}^{n} p_k t_k.$$

Then

$$\sum_{k=1}^{n} \frac{b_k}{B} \frac{a_k}{b_k} \log_2 \frac{a_k}{b_k} \geq \underbrace{\left(\sum_{k=1}^{n} \frac{b_k}{B} \frac{a_k}{b_k} \right)}_{\frac{A}{B}} \log_2 \frac{A}{B}$$

Thus,

$$\sum_{k} a_k \log_2 \frac{a_k}{b_k} \geq A \log_2 \frac{A}{B}$$

□

Corollary 0.2. *Let* $p^{(1)}, p^{(2)}, q^{(1)}, q^{(2)}$ *be probabilities on* $\mathcal{X}$ *and* $0 < \lambda < 1$.

$$D(\lambda p^{(1)} + (1-\lambda)p^{(2)} || \lambda q^{(1)} + (1-\lambda)q^{(2)}) \leq \lambda D(p^{(1)}||q^{(1)}) + 1 - \lambda)D(p^{(2)}||q^{(2)})$$

Proof. Let $x \in \mathcal{X}$, $a_1 = \lambda p^{(1)}(x)$, $a_2 = (1-\lambda)p^{(2)}(x)$, $b_1 = \lambda q^{(1)}(x)$, $b_2 = (1-\lambda)q^{(2)}j(x)$.

$$\begin{aligned} &\lambda p^{(1)}(x) \log_2 \frac{p^{(1)}(x)}{q^{(1)}(x)} + (1-\lambda)p^{(2)}(x) \log_2 \frac{p^{(2)}(x)}{q^{(2)}(x)} \\ &\qquad \geq \left(\lambda p^{(1)}(x) + (1-\lambda)p^{(2)}(x) \right) \log_2 \frac{\lambda p^{(1)}(x) + (1-\lambda)p^{(2)}(x)}{\lambda q^{(1)}(x) + (1-\lambda)q^{(2)}(x)} \end{aligned}$$

Sum over x and we get statement from corollary. We have equality if $\frac{p^{(1)}(x)}{q^{(1)}(x)} = \frac{p^{(2)}(x)}{q^{(2)}(x)}$ for all x. □

Corollary 0.3. $H(\lambda p^{(1)} + (1-\lambda)p^{(2)}) \geq \lambda H(p^{(1)}) + (1-\lambda)H(p^{(2)})$

Proof. We use uniform distribution.

$$\begin{aligned}
0 \leq D(p||U_{\mathcal{X}}) &= \sum_x p(x) \log_2 \frac{p(x)}{\frac{1}{|\mathcal{X}|}} \\
&= \sum_x p(x)(\log_2 p(x) + \log_2 |\mathcal{X}|) \\
&= -H(p) + \log_2 |\mathcal{X}|
\end{aligned}$$

Equality holds if and only if $p = U_{\mathcal{X}}$.

$$\begin{aligned}
-H(\lambda p^{(1)} + (1-\lambda)p^{(2)}) + \log_2|\mathcal{X}| &= D(\lambda p^{(1)} + (1-\lambda)p^{(2)}||\lambda U_{\mathcal{X}} + (1-\lambda)U_{\mathcal{X}}) \\
&\leq \lambda D(p^{(1)}||U_{\mathcal{X}}) + (1-\lambda)D(p^{(2)}||U_{\mathcal{X}}) \\
&= \lambda(-H(p^{(1)}) + \log_2|\mathcal{X}|) + (1-\lambda)(-H(p^{(2)}) + \log_2|\mathcal{X}|)
\end{aligned}$$

□

Definition. *Let X, Y, Z be random variable with values in $\mathcal{X}$. They form a* Markov triple*, written as $X \to Y \to Z$, if*

$$\mathbb{P}[Z = z|Y = y, X = x] = \mathbb{P}[Z = z|Y = x]$$

for all x, y, z such that $\mathbb{P}[Y = y, X = x] > 0$.

Note: This is equivalent to X, Z are independent conditionally upon Y:

$$\mathbb{P}[X = x, Z = z|Y = y] = \mathbb{P}[X = x|Y = y]\,\mathbb{P}[Z = z|Y = y]$$

for all y.

$p(x, y, z)$ is the joint distribution of X, Y, Z.

$$\mathbb{P}[Z = z|Y = y, X = x] = \frac{p(x, y, z)}{\sum_{\tilde{z}} p(x, y, \tilde{z})}$$

$$\mathbb{P}[Z = z|Y = y] = \frac{\sum_{\tilde{x}} p(\tilde{x}, y, z)}{\sum_{\tilde{x}\tilde{z}} p(\tilde{x}, y, \tilde{z})}$$

Theorem 0.4 (Data processing inequality)**.** *If $X \to Y \to Z$ is a Markov triple, then*

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

Proof.

$$\begin{aligned}
I((X_1, X_2); X) &= I(X_1; X) + I(X_2; X|X_1) \\
I(Y, Z; X) &= I(Y; X) + I(Z; X|Y) \\
I(Z, Y; X) &= I(Z; X) + I(Y; X|Z)
\end{aligned}$$

and $I(Y, Z; X) = I(Z, Y; X)$. Recall that $\sum_y \mathbb{P}[Y = y] I(Z; X|Y = y) = 0$. Then

$$I(Z; X|Y) = 0$$

Thus,

$$I(Y;X) = I(Z;X) + I(Y;X|Z)$$

$\square$

$X \to Y \to \hat{X}$ is called the "estimate" of X.

Lemma 0.5. *Let $X, \hat{X}$ be two random variables on $\mathcal{X}$ and $p_{err} = \mathbb{P}[\hat{X} \neq X]$. Then*

$$\underbrace{H(p_{err}, 1-p_{err})}_{\leq 1} + p_{err}\log_2|X| \geq H(X|\hat{X})$$

Proof.

$$E = \mathbb{1}_{[\hat{X}\neq X]}$$

We know $H(X, E|\hat{X}) = H(E, X|\hat{X})$.

$$H(X,E|\hat{X}) = H(X|\hat{X}) + H(E|X,\hat{X})$$
$$H(E,X|\hat{X}) = H(E|\hat{X}) + H(X|E,\hat{X})$$

Since $H(E|X,\hat{X}) = 0$, it follows

$$H(X|\hat{X}) = \underbrace{H(E|\hat{X})}_{\bigstar} + \underbrace{H(X|E,\hat{X})}_{\blacktriangle}$$

$$\bigstar \leq H(E) = H(p_{\text{err}}, 1-p_{\text{err}})$$
$$\blacktriangle = \mathbb{P}[E=0]\underbrace{H(X|E=0,\hat{X})}_{=0} + \mathbb{P}[E=1]H(X|E=1,\hat{X})$$
$$\leq p_{\text{err}} H(X|E=1) \leq \begin{cases} p_{\text{err}}\log_2|\mathcal{X}| \\ p_{\text{err}}\log_2(|\mathcal{X}|) \end{cases}$$

$\blacktriangle \leq p_{\text{err}}\log_2|\mathcal{X}|$

Then

$$\underbrace{H(p_{\text{err}}, 1-p_{\text{err}})}_{\leq 1} + p_{\text{err}}\log_2|X| \geq H(X|\hat{X})$$

$\square$

Theorem 0.6 (Fana's inequality)**.** *If $X \to Y \to \hat{X}$ is a Markov triple and $p_{err} = \mathbb{P}[\hat{X} \neq X]$ then*

$$H(p_{err}, 1-p_{err}) + p_{err}\log_2|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y).$$

In particular,

$$p_{err} \geq \frac{H(X|Y) - 1}{\log_2|\mathcal{X}|}$$

Proof. The first inequality follows by the lemma before. The second inequality follows from the Data processing inequality:

$$H(X) - H(X|Y) = I(X;Y) \geq I(X;\hat{X}) = H(X) - H(X|\hat{X})$$

$\square$