Lec. 2

Shannon's Code (attributed to Fano)

$$l_i = \left\lceil \log \frac{1}{p_i} \right\rceil$$

Check Kraft's inequality:

$$\sum_{i=1}^{n} \frac{1}{2^{l_i}} = \sum_{i=1}^{n} \frac{1}{2^{\lceil \log \frac{1}{p_i} \rceil}} \leq \sum p_i = 1.$$

Expected length of encoding of a symbol:

$$= \sum_{i=1}^{n} p_i \cdot \left\lceil \log \frac{1}{p_i} \right\rceil \geq \sum p_i \log \frac{1}{p_i} = H(X).$$

$$< \sum p_i \left(\log \frac{1}{p_i} + 1 \right) = H(X) + 1.$$

The worst-case length can be bad:

e.g., for $X = \begin{cases} 0 & \text{w.p. } 1-10^{-5} \\ 1 & \text{w.p. } 10^{-5} \end{cases}$

But Huffman is better.

Huffman vs. Shannon:

Huffman was better in the last example

but:

a	$1/3$	2	1
b	$1/3$	2	2
c	$1/4$	2	3
d	$1/12$	4	3
		~~Huff~~ Shannon # bits	~~Shannon~~ Huffman # bits

Lower bound:

Thm. The expected length of encoding by a prefix-free code $\{l_i\}$ of a r.v. X is $\geq H(X)$.

Pf.
$$\underbrace{H(X)}_{\sum p_i \log \frac{1}{p_i}} - \sum_{i=1}^{n} p_i l_i$$
$$= \sum p_i \log\left(\frac{1}{p_i 2^{l_i}}\right) \overset{\text{Jensen's}}{\leq} \log\left(\sum p_i \frac{1}{p_i 2^{l_i}}\right)$$
$$\overset{\text{Kraft's}}{\leq} \log 1 = 0. \quad \square$$

Improving +1's

Assume the source outputs a sequence of iid draws from X.

$X_1, \ldots, X_m$

Fact: $H(X_1, \ldots, X_m) = m \cdot H(X)$.

Shannon-Fano Code of $X_1, \ldots, X_m$

$\Downarrow$

$\mathbb{E}$ length $\leq m H(X) + 1$.

$\Rightarrow$ avg. # bits used per symbol $\leq H(X) + \frac{1}{m}$ ✓

$\Rightarrow$ Thm (Fundamental Thm of Source Coding or Noiseless Coding Thm):

$\forall \varepsilon > 0, \exists n$ s.t.
Given iid samples from a r.v. X, it's possible to communicate (on avg.) $\leq H(X) + \varepsilon$ bits/sample to reveal $X_1, \ldots, X_n$ to the other party.

Def. Let X, Y be two (possibly correlated) r.v's, Then

$$H(X,Y) := \sum_{x,y \leftarrow (X,Y)} \underbrace{p(x,y)}_{\Pr(X=x,\, Y=y)} \log \frac{1}{p(x,y)}.$$

If X, Y are independent $\Rightarrow p(x,y) = p(x)p(y)$

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

In general, $p(x,y) = p(x)\, p(y|x)$

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

$$= \underbrace{\sum p(x,y) \log \frac{1}{p(x)}}_{H(X)} + \underbrace{\sum_{x,y} p(x,y) \log \frac{1}{p(y|x)}}_{= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)}}$$

$$= H(X) + \sum_x p(x) \cdot H(Y \mid X = x)$$

$\stackrel{\text{Def}}{=} H(X) + H(Y|X)$.

$\Rightarrow$ <u>Lemma:</u> $H(X,Y) = H(X) + H(Y|X)$.

If X, Y are independent: $H(X,Y) = H(X) + H(Y)$.

& $H(Y|X) = H(Y)$.

<u>Thm (Chain rule):</u>

(for any ordering)

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

$H(X_n | X_1, \ldots, X_{n-1})$.

<u>example:</u> $H(X|X) = 0$

X uniform on $\{0, \ldots, 3\} \Rightarrow H(X) = 2$

$(Y = X \bmod 2)$ " $\{0, 1\} \Rightarrow H(Y) = 1$

$H(X|Y) = 1$. $H(Y|X) = 0$.

$H(X,Y) = 2$.

<u>Lemma</u>. $H(y|x) \leq H(y)$.

<u>Proof</u>. $H(y|x) - H(y) = H(x,y) - H(x) - H(y)$

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

$$- \underbrace{\sum_x p(x)}_{= \sum_{x,y} p(x,y)} \log \frac{1}{p(x)} - \underbrace{\sum_y p(y)}_{= \sum_{x,y} p(x,y)} \log \frac{1}{p(y)}$$

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

$$= \mathop{\mathbb{E}}_{p(x,y)} \log Z \leq \log \underbrace{\mathbb{E}\, Z}_{\sum p(x)p(y) = 1} = 0.$$

Cor. ("Union bound"):

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

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$$

---

* Exercise. (follows from "conditioning cannot increase entropy")

$H(\underline{p})$ is a concave function of the distribution.

i.e., $$\lambda H(\underline{p}) + (1-\lambda) H(\underline{q}) \leq H(\underbrace{\lambda \underline{p} + (1-\lambda) \underline{q}}_{\text{take it as some r.v.}}).$$

Mutual Information

Def. $I(X;Y) := H(X) - H(X|Y)$

(Reduction in uncertainty of X by virtue of knowing Y)

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

$\Rightarrow I(X;Y) \geq 0, \quad I(X;Y) = I(Y;X)$

If X, Y are independent,

$I(X;Y) = 0$.

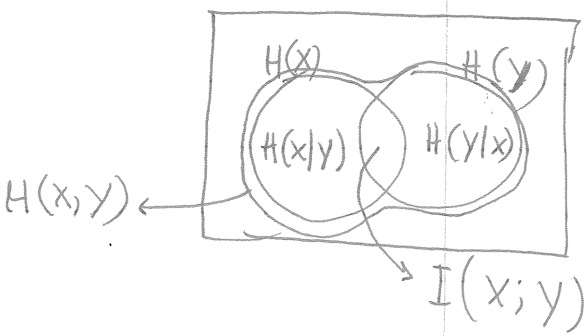
For the $\{0, \ldots, 3\}$ example:

$I(X;Y) = 1$.

Or, Z_i iid $\{0,1\}$.

$X = Z_1, \ldots, Z_5, \quad Y = Z_4 Z_5 Z_6 Z_7$

$\Rightarrow I(X;Y) = 2$.



Conditional Mutual Information.

$$I(X;Y|Z) \overset{def}{:=} H(X|Z) - H(X|Y,Z)$$

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

Exercise: (Chain rule)

$$I(X_1, \dots, X_n ; Y) = \sum_{i=1}^{n} I(X_i ; Y | X_1, \dots X_{i-1})$$

$$= I(X_1 ; Y) + I(X_2 ; Y | X_1) + \dots + I(X_n ; Y | X_1, \dots, X_{n-1})$$

Exercise: Is it always the case that $I(X;Y|Z) \overset{?}{\leq} I(X;Y)$