EE/Ma 126a Information Theory - Homework Set #3

Ling Li, ling@cs.caltech.edu

October 18, 2000

3.1 The AEP and source coding

(a) There are $\binom{100}{w}$ different source sequences that has w 1's. So totally we need

$$\sum_{w=0}^{3} \binom{100}{w} = 166751$$

different codewords. Thus the minimum length is

$$\boxed{\lceil \log_2 166751 \rceil = 18}.$$

(b) No codeword will be assigned iff the source sequence has more than three 1's. Thus, the probability of observing no codeword is

$$\boxed{1 - \sum_{w=0}^{3} \binom{100}{w} p(1)^w p(0)^{100-w} \approx 0.001673}. \tag{1}$$

(c) Let random variable $X_i \in \{0, 1\}$ denotes the ith bit of a source sequence, where $i = 1..100$, and Y be the number of 1's in that sequence. Thus, $X_1, X_2, \ldots, X_{100}$ are i.i.d. $\sim p$, and

$$Y = \sum_{i=1}^{100} X_i.$$

Since $EX_i = p(1)$, $\sigma^2(X_i) = E\left[(X_i - EX_i)^2\right] = p(0)p(1)$, we have $EY = 100p(1) = 0.5$ and $\sigma^2(Y) = 100p(1)p(0) = 0.4975$. With Chebyshev's inequality

$$\Pr\left[(Y - EY)^2 \geq a\right] \leq \frac{\sigma(Y)^2}{a}$$

and $a = 3.5^2$, we have[1]

$$\boxed{\Pr\left[\text{source sequence has 4 or more } 1'\text{s}\right] \leq \frac{0.4975}{3.5^2} \approx 0.04061},$$

which is a much bigger approximation of (1).

[1] Since $(Y - EY)^2 \geq a \Leftrightarrow Y \leq -3$ or $Y \geq 4$, and Y is not negative, we have $(Y - EY)^2 \geq a \Leftrightarrow Y \geq 4$.

3.2 AEP

(a) Since $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$, we have

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n}\log q(X_1, X_2, \ldots, X_n) &= -\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log q(X_i) \\
&= -E_p \log q(X_1) \; \boxed{\text{in probability}} \\
&= \boxed{-\sum_{x=1}^{m} p(x)\log q(x)}.
\end{aligned}$$

The result can also be written as $\boxed{H(p) + D(p||q)}$ which is bigger than $H(p)$ since $p \neq q$.

(b)

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log \frac{q(X_1, X_2, \ldots, X_n)}{p(X_1, X_2, \ldots, X_n)} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log \frac{q(X_i)}{p(X_i)} \\
&= E_p \log \frac{q(X_1)}{p(X_1)} \; \boxed{\text{in probability}} \\
&= \boxed{-D(p||q)},
\end{aligned}$$

which is negative since $p \neq q$.

3.3 Data processing

Summing $p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})$ over all possible x_1, we get

$$\sum_{x_1} p(x_1, x_2, \ldots, x_n) = \sum_{x_1} p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1}),$$

which is

$$p(x_2, x_3, \ldots, x_n) = p(x_2)p(x_3|x_2)\cdots p(x_n|x_{n-1}).$$

Thus

$$\begin{aligned}
p(x_2)p(x_1, x_2, \ldots, x_n) &= p(x_2)\left[p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})\right] \\
&= p(x_1)p(x_2|x_1)\left[p(x_2)p(x_3|x_2)\cdots p(x_n|x_{n-1})\right] \\
&= p(x_1)p(x_2|x_1)p(x_2, x_3, \ldots, x_n),
\end{aligned}$$

i.e.,

$$H(X_2) + H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + H(X_2, X_3, \ldots, X_n).$$

Then we have

$$\begin{aligned}
I(X_1; X_2, \ldots, X_n) &= H(X_1) + H(X_2, X_3, \ldots, X_n) - H(X_1, X_2, \ldots, X_n) \\
&= H(X_2) - H(X_2|X_1) = \boxed{I(X_1; X_2)}.
\end{aligned}$$

3.4 Conditional mutual information

Assume $n \geq 2$, and $k = 1..(n-1)$.

- Since sequence $X_1 X_2 \ldots X_n$ has even number of 1's, we have in $GF(2)$,

$$X_n = X_1 + X_2 + \cdots + X_{n-1}. \tag{2}$$

  That is, with any sequence $X_1 X_2 \ldots X_{n-1}$, we can find exactly one X_n so that $X_1 X_2 \ldots X_n$ has even number of 1's. So

$$p(x_1, x_2, \ldots, x_k) = \sum_{x_{k+1}..x_{n-1}} 2^{-(n-1)} = 2^{n-k-1} \times 2^{-(n-1)} = 2^{-k}.$$

  Thus $H(X_1, X_2, \ldots, X_k) = k$. We also have

$$H(X_k | X_1, X_2, \ldots, X_{k-1}) = H(X_1, X_2, \ldots, X_k) - H(X_1, X_2, \ldots, X_{k-1}) = 1.$$

  When $k = 1$, it says $H(X_1) = 1$.

- Like entropy, the mutual information is a function of the distribution. Since $X_1, X_2, \ldots, X_n$ are symmetric in the problem, we have

$$I(X_k; X_{k+1} | X_1, X_2, \ldots, X_{k-1}) = I(X_k; X_n | X_1, X_2, \ldots, X_{k-1}). \tag{3}$$

  We also have $H(X_n) = H(X_1) = 1$.

- From $I(X;Y|Z) \geq 0$ (corollary 2.93 in [Cover]) and regarding $(X_1, \ldots, X_{k-1})$ as a random vector-valued variable, we have

$$I(X_k; X_n | X_1, X_2, \ldots, X_{k-1}) \geq 0. \tag{4}$$

- (2) says X_n is a function of $X_1, X_2, \ldots, X_{n-1}$. So

$$I(X_1, X_2, \ldots, X_{n-1}; X_n) = H(X_n) = 1. \tag{5}$$

- While given $X_1, \ldots, X_{n-2}$, from (2) we know that X_{n-1} is also a function of X_n, so

$$\boxed{I(X_{n-1}; X_n | X_1, X_2, \ldots, X_{n-2}) = H(X_{n-1} | X_1, X_2, \ldots, X_{n-2}) = 1}. \tag{6}$$

- From the chain rule for information,

$$I(X_1, X_2, \ldots, X_{n-1}; X_n) = \sum_{k=1}^{n-1} I(X_k; X_n | X_1, X_2, \ldots, X_{k-1}).$$

  When considered together with (4), (5), and (6), we have for $k < n-1$,

$$I(X_k; X_n | X_1, X_2, \ldots, X_{k-1}) = 0.$$

  Thus with (3), $I(X_k; X_{k+1} | X_1, X_2, \ldots, X_{k-1}) = 0$ for $k = 1..(n-2)$, i.e.,

$$\boxed{I(X_1; X_2) = I(X_2; X_3 | X_1) = \cdots = I(X_{n-2}; X_{n-1} | X_1, X_2, \ldots, X_{n-3}) = 0}.$$

3.5 Inequalities

(a) $\boxed{H(X,Y|Z) = H(X|Z) + H(Y|Z,X) \geq H(X|Z)}$ with equality iff $\boxed{Y \text{ is a function of } (X,Z)}$;

(b) Use the chain rule for information:

$$\boxed{I(X,Y;Z) = I(X;Z) + I(Y;Z|X) \geq I(X;Z)}$$

with equality iff $\boxed{Y \text{ and } Z \text{ are conditionally independent given } X}$ (corollary 2.93 in [Cover]);

(c) From (b) and

$$\begin{aligned} I(X,Y;Z) &= H(X,Y) + H(Z) - H(X,Y,Z), \\ I(X;Z) &= H(X) + H(Z) - H(X,Z), \end{aligned}$$

we have $\boxed{H(X,Y,Z) - H(X,Y) = H(Z) - I(X,Y;Z) \leq H(Z) - I(X;Z) = H(X,Z) - H(X)}$ with equality iff $\boxed{Y \text{ and } Z \text{ are conditionally independent given } X}$;

(d) From $I(X,Y;Z) = I(X;Z) + I(Y;Z|X) = I(Y;Z) + I(X;Z|Y)$,

$$\boxed{I(X;Z|Y) = I(Z;Y|X) - I(Z;Y) + I(X;Z)}.$$

We can also write

$$I(X;Z|Y) \geq I(Z;Y|X) - I(Z;Y) + I(X;Z)$$

$\boxed{\text{with equality always true}}$.