ECE 776 - Information theory (Spring 2010)
Midterm

Please give well-motivated answers.

Q1 (1 point). Consider random variables X, Y, Z such that $Z = X + Y$. Show that $H(Z) \leq H(X) + H(Y)$. When does equality hold?

Sol.: We have
$$H(Z) = H(X + Y) \leq H(X, Y) \leq H(X) + H(Y).$$

The first inequality is an equality if from $X + Y$ one is always able to recover X and Y (this depends on the alphabets in which X and Y are defined. For instance, if $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{3, 5\}$, this is the case). The second inequality is an equality if X and Y are independent.

Q2 (1 point) Consider the joint distribution $p(x, y)$ given by the following table

$x \backslash y$	0	1
0	1/4	1/4
1	1/2	0

Find the best estimator of X given Y (in the sense of minimizing the error probability) and calculate its probability of error. Then, calculate $H(X|Y)$ and verify the Fano inequality for this estimator.

Sol.: The estimator $\hat{X}(Y)$ that minimizes the probability of error is given by

$$\hat{X}(0) = 1$$
$$\hat{X}(1) = 0.$$

The probability of error is $P_e = p(0,0) + p(1,0) = 1/4$.
We have

$$H(X|Y) = p_Y(1)H(X|Y = 1) + p_Y(0)H(X|Y = 0)$$
$$= \frac{1}{4}0 + \frac{3}{4}H\left(\frac{1}{4} \cdot \frac{4}{3}\right) = \frac{3}{4}H\left(\frac{1}{3}\right) = 0.69.$$

By Fano inequality

$$H(X|Y) \leq P_e \log(2 - 1) + H(P_e)$$
$$\rightarrow 0.69 \leq 0.81.$$

Q3. (1 point) We want to find out if a treasure lies behind door $X = 1$, 2 or 3. Our prior knowledge is summarized in the distribution $p_X(1) = 1/2$, $p_X(2) = 1/4$ and $p_X(3) = 1/4$. We are allowed to ask the question: Is $X \in \mathcal{S}$ (where $\mathcal{S}$ is a subset of $\{1, 2, 3\}$)? The answer to the question is $Y = 1$ if $X \in \mathcal{S}$ and 0 otherwise. Which subset $\mathcal{S}$ would you choose

to maximize the mutual information between X and Y (and thus increasing our chances of finding the treasure)? [Hint: Calculate $I(X;Y)$ for different sets].

Sol.: Let us evaluate

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(\Pr[X \in \mathcal{S}]) - H(Y|X) \\
&= H(\Pr[X \in \mathcal{S}]) \le 1.
\end{aligned}
$$

Thus, the mutual information is maximized if we choose $\mathcal{S} = \{1\}$ or $\mathcal{S} = \{2,3\}$, since in those cases, we obtain $I(X;Y) = 1$. In other words, from the question at hand, we can gather at most 1 bit about the location of the treasure, and these choices of $\mathcal{S}$ allow us to obtain this much information.

Q4 (1 point) We are given the joint pmf $p(x,y)$ defined as below

$x \backslash y$	0	1
0	1/8	1/8
1	1/8	5/8

Are the sequences $x^4 = 1110$ and $y^4 = 0111$ individually ϵ−typical with respect to the marginal distributions $p(x)$ and $p(y)$, respectively, for $\epsilon = 0.1$? Are they jointly ϵ−typical with respect to $p(x,y)$ with $\epsilon = 0.1$?

Sol: We have the marginals $p(x) = p(y) = (1/4, 3/4)$ so that $H(X) = H(Y) = -1/4 \log_2 1/4 - 3/4 \log_2 3/4 = 0.81$ bits. Moreover, the joint entropy is $H(X,Y) = -3/8 \log_2 1/8 - 5/8 \log_2 5/8 = 1.55$ bits.
It is easy to see that x^4, y^4 are individually typical with respect to $p(x)$ and $p(y)$ for any ϵ since

$$
-\frac{1}{4} \log_2 p(x^4) = -\frac{1}{4} \log_2 p(y^4) = H(X) = H(Y).
$$

To check whether they are jointly typical, we must calculate

$$
\begin{aligned}
-\frac{1}{4} \log_2 p(x^4, y^4) &= -\frac{1}{4} \log_2 \left(\frac{5^2}{8} \cdot \frac{1^2}{8} \right) \\
&= 1.84 > H(X,Y) + \epsilon = 1.55 + 0.1.
\end{aligned}
$$

Therefore, the sequences are not jointly typical with respect to the given joint distribution.

P1 (2 points) An i.i.d. source X^n is characterized by the pmf

$$
p(x) = \begin{cases} 0.5 & x = 1 \\ 0.3 & x = 2 \\ 0.2 & x = 3 \end{cases}.
$$

(a) What is the minimal rate required for lossless compression?
(b) Find the codeword lengths for the binary Huffman code and for the Shannon code. Then, calculate the average codeword lengths in the two cases, and compare them with the previous point.

2

(c) Propose a compression scheme that is able to attain the performance limits of point (a).
(d) If X^n is processed according to a function $g(\cdot)$, producing $Y^n = g(X^n)$, does the rate necessary to compress Y^n increase or decrease? Why?

Sol.: (a) It is
$$R = H(X) = 1.485 \text{ bits/ source symbol.}$$

(b) For Shannon, we have

$$
\begin{aligned}
\ell(1) &= \lceil -\log_2(0.5) \rceil = 1 \\
\ell(2) &= \lceil -\log_2(0.3) \rceil = 2 \\
\ell(3) &= \lceil -\log_2(0.2) \rceil = 3,
\end{aligned}
$$

so that the average length is $L = 0.5 + 0.6 + 0.6 = 1.7 > H(X)$.
For Huffman, using the usual procedure, we get

$$
\begin{aligned}
\ell(1) &= 1 \\
\ell(2) &= 2 \\
\ell(3) &= 2,
\end{aligned}
$$

so that the average length is $L = 0.5 + 0.6 + 0.4 = 1.5 > H(X)$ (but quite close to $H(X)$!).
(c) We need to code over blocks of n symbols with n large. As seen in class, we can either use a typicality-based compression scheme, or use Shannon code over blocks of n symbols (see lectures or textbook)
(d) Since $H(Y) \leq H(X)$, the compression rate decreases.

P2 (2 point) The evolution of a random process $X_i \in \{1, 2, 3, 4\}$ depends on another random process $Y_i \in \{0, 1\}$, $i = 1, 2, ..., n$, in the following way. Process $Y_1, Y_2, ..., Y_n$ is a Markov chain with transition probability $p(y_k | y_{k-1})$

$y_{k-1} \backslash y_k$	0	1
0	0.2	0.8
1	0.4	0.6

,

while X^n is such that $p(x^n | y^n) = \prod_{i=1}^{n} p(x_i | y_i)$, where

$$
p(x_i | 0) = \begin{cases} 1/2 & \text{for } x_i = 1 \text{ and } x_i = 2 \\ 0 & \text{otherwise} \end{cases}
$$

and

$$
p(x_i | 1) = \begin{cases} 1/2 & \text{for } x_i = 3 \text{ and } x_i = 4 \\ 0 & \text{otherwise} \end{cases}.
$$

(a) Is process Y^n stationary for any initial distribution $\Pr[Y_1 = 1]$? If not, find an initial distribution that makes Y^n stationary. Assume this distribution for the following questions.
(b) Calculate the entropy rate of Y^n.
(b) Is process X^n stationary? Is it a Markov chain?

(b) Does the entropy rate of X^n exist? Calculate the entropy (not the entropy rate) of process X^n. Do you expect the entropy rate to be larger or smaller?

Sol.: (a) No, from the theory on Markov chains, Y^n is stationary only if the initial distribution is the stationary distribution:

$$\Pr[Y_1 = 1] = \frac{0.8}{0.8 + 0.4} = \frac{2}{3}.$$

(b) We have

$$
\begin{aligned}
H(\mathcal{Y}) &= H(Y_2|Y_1) = \frac{1}{3}H(Y_2|Y_1 = 0) + \frac{2}{3}H(Y_2|Y_1 = 0) \\
&= \frac{1}{3}H(0.2) + \frac{2}{3}H(0.4) \\
&= 0.89.
\end{aligned}
$$

(c) The process X^n is stationary, but is not a Markov chain. To see this, let us calculate

$$
\begin{aligned}
p(x^n) &= \sum_{y^n \in \mathcal{Y}^n} p(y^n)p(x^n|y^n) \\
&= \sum_{y^n \in \mathcal{Y}^n} p(y_1) \prod_{k=2}^{n} p(y_k|y_{k-1}) \prod_{i=1}^{n} p(x_i|y_i),
\end{aligned}
$$

which is the same for all possible time shifts, but does not factorize as for a Markov chain (it is in fact a Hidden Markov Model).

(d) It exists since the process is stationary. To calculate the entropy, we evaluate

$$
\begin{aligned}
p_X(1) &= \sum_{y} p(y)p(x|y) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} \\
p_X(2) &= \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} \\
p_X(3) &= \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3} \\
p_X(4) &= \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3},
\end{aligned}
$$

so that

$$H(X) = 1.91.$$

We expect the entropy rate to be smaller than the entropy since the process is not i.i.d.

P3. Consider the channel described by the following $p(y|x)$ with alphabets $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, e\}$

x/y	0	e	1
0	$1 - \alpha - \epsilon$	α	ϵ
1	ϵ	α	$1 - \alpha - \epsilon$

with $0 \le \alpha, \epsilon \le 1$ and $\alpha + \epsilon \le 1$.

4

(a) Calculate $I(X;Y)$ for a given $p(x)$ (express the mutual information in terms of $p_X(1) = p$) (Hint: To simplify the derivation, define a variable $E = 1$ if $Y = e$ and $E = 0$ otherwise, and use the fact that E is a function of Y...).
(b) Calculate the capacity of the channel at hand.
(c) Argue and verify that if we set $\alpha = 0$ and $\epsilon = 0$ we recover known results.

Sol.:
(a) We have

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(YE) - H(YE|X) \\
&= H(E) + H(Y|E) - H(E|X) - H(Y|EX) \\
&= H(\alpha) + (1-\alpha)H(p) - H(\alpha) - (1-\alpha)H\left(\frac{\epsilon}{1-\alpha}\right)
\end{aligned}
$$

(b) The mutual information is maximized for $p = 1/2$ so that we have

$$
C = (1-\alpha)\left(1 - H\left(\frac{\epsilon}{1-\alpha}\right)\right).
$$

(c) For $\alpha = 0$, we recover the capacity of a binary symmetric channel, whereas for $\epsilon = 0$ we obtain the capacity of a binary erasure channel.