Random codes in communication 11 June 2010
(Dr. Permuter Haim and Mr. Iddo Naiss)

Final Exam - Solution
Total time for the exam: 3 hours!

This is a honor code assignment and each student should do it totaly by himself/herself. Please copy and sign the sentence below on your assignment (an assignment without the following sentence won't be graded.)

"I am respecting the honor code of this assignment: Signature_____"

1) **True or False** (35 points)
Copy each relation to your notebook and write **true** or **false**. Then, if it's true, prove it. If it is false give a counterexample or prove that the opposite is true.
   a) Let X, Y be two independent random variables. Then
   $$H(X-Y) \geq H(X).$$
   True
   $$H(X-Y) \overset{(a)}{\geq} H(X-Y|Y)) \overset{(b)}{\geq} H(X)$$
   (a) Conditioning Reduces Entropy.
   (b) Given Y, $X-Y$ is a Bijective Function.

   b) Let $A, B, C.D$ be random variables with a finite alphabet that form the Markov chain $A-B-C-D$, namely $P(a,b,c,d) = P(a)P(b|a)P(c|b)P(d|c)$. Then,
   $$I(A;C|D) \leq H(B).$$
   True
   c) For any finite alphabet random variables
   $$H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X).$$
   True
   $$H(X,Y,Z) - H(X,Y) =$$
   $$H(X) + H(X|Y) + H(Z|X,Y) - H(X) - H(Y|X) =$$
   $$H(Z|X,Y) \overset{(a)}{\leq}$$
   $$H(Z|X) = H(Z,X) - H(X)$$
   (a) Conditioning Reduces Entropy.

d) Let $\{X_i\}_{i\geq 1}$ be an i.i.d. source distributed according to P_X. In addition, let $\{Y_i\}_{i\geq 1}$ and $\{Z_i\}_{i\geq 1}$ be two i.i.d. side information sequences that may be available at the encoder and decoder of a lossless source coding setting. If $I(X;Y) > I(X;Z)$, then the minimum rate that is needed to compress $\{X_i\}_{i\geq 1}$ losslessly with side information $\{Y_i\}_{i\geq 1}$ is smaller than the minimum rate that is needed to compress $\{X_i\}_{i\geq 1}$ losslessly with with side information $\{Z_i\}_{i\geq 1}$. Assume the side information is known both to the encoder and decoder.

True

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

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

$$I(X;Y) > I(X;Z) \Rightarrow H(X|Z) - H(X|Y) > 0 \Rightarrow H(X|Y) < H(X|Z)$$

e) Suppose that (X, Y, Z) are jointly Gaussian and that $X - Y - Z$ forms a Markov chain. Let X and Y have correlation coefficient ρ_1 and let Y and Z have correlation coefficient ρ_2. Let the variance of each random variable X, Y, and Z be 1. be Let $g(z)$ be any deterministic function. Then

$$h(X|g(Z)) \geq \frac{1}{2}\log 2\pi e(1-\rho_1^2\rho_2^2).$$

True

We've shown that for this Model:

$$I(X;Z) = -\frac{1}{2}\log 2\pi e\left(1-{\rho_1}^2{\rho_2}^2\right)$$

Moreover:

$$H(X) - H(X|Z) = I(X;Z) \overset{(a)}{\geq} I(X;g(Z))$$

Hence we get:

$$-\frac{1}{2}\log 2\pi e\left(1-{\rho_1}^2{\rho_2}^2\right) \geq H(X) - H(X|g(Z)) \Rightarrow$$

$$H(X|G(Z)) \geq H(X) + \frac{1}{2}\log 2\pi e\left(1-{\rho_1}^2{\rho_2}^2\right) \overset{(b)}{\geq} \frac{1}{2}\log 2\pi e\left(1-{\rho_1}^2{\rho_2}^2\right)$$

(a) Data-Processing Inequality.
(b) $H(X) \geq 0$.

f) If $f(x,y)$ is a convex function in the pair (x,y), then for a fixed y, $f(x,y)$ is convex in x, and for a fixed x, $f(x,y)$ is convex in y.

True
If the function is Convex for every combination of (x,y) it is necessarily Convex for Affine Function of the pair.

g) If for a fixed y the function $f(x,y)$ is a convex function in x , and for a fixed x, $f(x,y)$ is convex function in y, then $f(x,y)$ is convex in the pair (x,y). (Examples of such functions are $f(x,y) = f_1(x) + f_2(y)$ or $f(x,y) = f_1(x)f_2(y)$ where $f_1(x)$ and $f_2(y)$ are convex.)

False

2) **Entropy and source coding of a source with infinite alphabet** (15 points)
Let X be an i.i.d. random variable with an infinite alphabet, $\mathcal{X} = \{1, 2, 3, ...\}$. In addition let $P(X = i) = 2^{-i}$.

a) What is the entropy of the random variable?
b) Find an optimal variable length code, and show that it is indeed optimal.

Solution

a)

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$$

$$= -\sum_{i=1}^{\infty} 2^{-i}\log_2(2^{-i})$$

$$= -\sum_{i=1}^{\infty} \frac{-i}{2^i} = 2$$

b) Coding Scheme:

1	0
2	10
3	110
4	1110
5	11110
.	.
.	.
.	.

Average Length:

$$L^* = \sum_{i=1}^{\infty} p(x = i)L(i) = \sum_{i=1}^{\infty} \frac{i}{2^i} = 2 = H(X)$$

Hence it is the Optimal Code.

3) **Empirical distribution of a sequence** (20 points)
A fair dice with 6 faces was thrown n times, where n is a very large number.

a) Using Stirling approximation $n! \approx (\frac{n}{e})^n$, find how many different sequences there exists with an empirical pmf $(p_1, p_2, ..., p_6)$, where p_i is the portion of the sequence that is equal to $i \in \{1, 2, ..., 6\}$.
b) Now, we were told that the portion of odd numbers in the sequence is $2/3$ (i.e., $p_1 + p_3 + p_5 = 2/3$). For n very large, what is the most likely empirical pmf of the sequence.

Solution

a) Number of Combinations is given by:

$$\binom{n}{n_1 n_2 n_3 n_4 n_5 n_6} =$$

$$\frac{n!}{n_1! n_2! n_3! n_4! n_5! n_6!} \overset{(a)}{\approx}$$

$$\frac{\left(\frac{n}{e}\right)^n}{\left(\frac{n_1}{e}\right)^{n_1}\left(\frac{n_2}{e}\right)^{n_2}\left(\frac{n_3}{e}\right)^{n_3}\left(\frac{n_4}{e}\right)^{n_4}\left(\frac{n_5}{e}\right)^{n_5}\left(\frac{n_6}{e}\right)^{n_6}} = N$$

Hence we get:

$$\log N = n\log n - n_1\log n_1 - n_2\log n_2 - \ldots - n_6\log n_6 =$$

$$-\sum_{i=1}^{6} n_i \log\frac{n_i}{n} \Rightarrow N \approx 2^{-\sum_{i=1}^{6} n_i \log\frac{n_i}{n}} =$$

$$2^{-n\sum_{i=1}^{6}\frac{n_i}{n}\log\frac{n_i}{n}} = 2^{nH(\frac{n_1}{n},\frac{n_2}{n},\frac{n_3}{n},\frac{n_4}{n},\frac{n_5}{n},\frac{n_6}{n})} \overset{(b)}{=}$$

$$2^{nH(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})} \overset{(c)}{=} 2^{6nH(\frac{1}{6})}$$

(a) Stirling's Approximation.
(b) $\frac{n_i}{n} = \frac{1}{6}$.
(c) i.i.d.
The result is probable since the Typical Set and the set defined by the Law of Large Numbers converge as the number of samples goes to Infinty.

b)

$$\frac{n_1 + n_3 + n_5}{n} = \frac{2}{3} \Leftrightarrow p_1 + p_2 + p_3 = \frac{2}{3}$$

We would like to find $\{p_1, p_2, p_3, p_4, p_5, p_6\}$ which, under the given constraints Maximizes Entropy. This results in the biggest Typical Set which means the event will be most likely. Equivalently:

$$X = \begin{cases} 1 & p_1 \\ 3 & p_3 \\ 5 & p_5 \end{cases},\ Y = \begin{cases} 2 & p_2 \\ 4 & p_4 \\ 6 & p_6 \end{cases},\ Z = \begin{cases} X & \frac{2}{3} \\ Y & \frac{1}{3} \end{cases}$$

Next, creating an Indicator Auxiliary Variable:

$$I = \begin{cases} 1 & Z = X \\ 0 & Z = Y \end{cases}$$

Now, Maximizing the Entropy of Z.

$$H(Z) \overset{(a)}{=} H(Z, I) = H(I) + H(Z|I) = H(\frac{2}{3}) + \frac{1}{3}H(Y) + \frac{2}{3}H(X)$$

(a) $H(Z, I) = H(Z) + \underbrace{H(I|Z)}_{Deterministic\ Given\ Z} = H(Z)$.

Hence Maximizing $H(Z)$ means Maximizing $H(X),\ H(Y)$.
As We've shown at class Uniform Distribution maximizes Entropy.

$$H(X),\ H(Y) \sim Uniform \Rightarrow p_1,\ p_3,\ p_5 = \frac{2}{9},\ p_2,\ p_4,\ p_6 = \frac{1}{9}$$

4) **For a Gaussian input in an additive channel, the worse noise is Gaussian.** (30 points)

Fig. 1. Fig. 1. An additive channel

Prove, that for an additive $Y = X + Z$ channel with Gaussian input, and power constraint

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P, \text{ with probability 1}$$

the worse noise with varaince $V(Z) \leq N$ is a Gaussian noise. We use the following notation: when a random variable X has a Gaussian distribution we denote it by X_G.
In short, prove that for $Y = X_G + Z$, where $V(Z) \leq N$

$$I(X_G; Y) \geq \frac{1}{2}\log(1 + \frac{P}{N}) = I(X_G; X_G + Z_G),$$

where $Z_G \sim N(0, N)$.
You can prove it on your own, or use the following steps

(a) Prove that for every $\alpha \in \mathbb{R}$,

$$I(X_G; Y) \geq \frac{1}{2}\log(2\pi eP) - \frac{1}{2}\log(2\pi eE((X_G - \alpha Y)^2)),$$

and justify your steps.
(b) Find such α that minimize $E((X_G - \alpha Y)^2)$, i.e., αY is the best linear approximation of X_G (hint: the orthogonality principle, or by differentiating).
(c) Using the parameter α you found, finish the question.

Conclude, what would you have done, as an adversary (someone who oppose the communication, and wants to block it), if you found out the encoder is using a Gaussian codebook?
Solution
The suggested solution is pretty straightforward.
There are 2 more approaches:

a) Using the Characteristic Function of a Normal Random Variable:
Givan $Y = X + Z,\ X \sim Normal,\ Z \perp X$ Then $Y \sim Normal \Leftrightarrow Z \sim Normal$.
The direction in which $Z \sim Normal$ is trivial since $Z \perp X$.
For the other direction we have $Y \sim Normal$. Examing its Characteristic Function:

$$\varphi_Y(t) = E\left[e^{itY}\right] = E\left[e^{it(X+Z)}\right] \overset{(a)}{=} E\left[e^{itX}\right] E\left[e^{itZ}\right] \Rightarrow E\left[e^{itZ}\right] = E\left[e^{itY}\right] / E\left[e^{itX}\right]$$

Now, one should recall that for $X \sim N(\mu,\ \sigma^2)$ The C.F. is $\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$.
Which yields:

$$E\left[e^{itZ}\right] = E\left[e^{itZ}\right] = E\left[e^{itY}\right] / E\left[e^{itX}\right] = e^{it(\mu_Y - \mu_X) - \frac{1}{2}\left({\sigma_Y}^2 - {\sigma_X}^2\right)^2 t^2}$$

Which is the C.F. of a Normal Random Variable, Hence $Z \sim N(\mu_Y - \mu_X,\ {\sigma_Y}^2 - {\sigma_X}^2)$. The resoning for that is simple. One could show that in order to minimize the Mutual Information the constrain on Y would be $Y \sim Normal$ hence the proof holds.

(a) $X,\ Y$ *R.V. s.t.* $X \perp Y \iff \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

b) Using Shannon Inequality (Entropy Power Inequality):

$$I(X_G + Z; X_G) = h(X_G + Z) - h(X_G + Z|X_G) \overset{(a)}{=} h(X_G + Z) - h(Z)$$

By Shannon Inequality (Entropy Power Inequality), for $X \perp Z$:

$$2^{2h(X+Z)} \geq 2^{2h(X)} + 2^{2h(Z)}$$

Defining $f(Z) = \frac{2^{2h(Z)}}{2\pi e}$ yields:

$$h(X_G + Z) - h(Z) \geq \frac{1}{2}\log(2^{2h(X)} + 2^{2h(Z)}) - h(Z) =$$

$$\frac{1}{2}\log((2\pi e)P + 2\pi e f(Z)) - \frac{1}{2}\log(2\pi e f(Z)) = \frac{1}{2}\log(1 + \frac{P}{f(Z)})$$

Now, maximizing $f(Z)$ minimizes $I(X_G; Y)$. By definition fo $f(Z)$, maximizing it means maximizing $h(Z)$ under the Power constrain. As shown at class in order to do so we need to create $Z \sim N(0, N)$.

Good Luck!