ECE 776 - Information theory (Spring 2012)
Final

P1 (1 point). Consider the random process $X_i = UZ_i$, where U equals -1 or 1 with equal probability and Z_i are i.i.d. over index i and distributed as $Z_i \sim f(z) = \lambda e^{-\lambda z}$, with $z \geq 0$ (and $f(z) = 0$ otherwise).
a. What can we say about the limit $\lim_{n\to\infty} 1/n \sum_{i=1}^{n} X_i$?
b. Calculate the differential entropy rate $\lim_{n\to\infty} \frac{1}{n} h(X^n)$.

P2 (1 point). For any random variables $X_1, ..., X_n$, prove the equality

$$H(X_1, ..., X_n) \leq \frac{1}{n-1} \sum_{i=1}^{n} H(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n).$$

(Hint: Note that the equality is equivalent to $nH(X_1, ..., X_n) \leq \sum_{i=1}^{n} H(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) + H(X_1, ..., X_n)$).

P3 (1 point). Derive the equation that describes the typical set for a Gaussian distribution with mean μ and variance σ^2.

P4 (1 point). Suppose that X is a random variable with support $[1, \infty)$ and Y is a random variable with support $[0, 1]$. Moroever, we have $E[X] = \mu_X \geq 0$ and $E[Y] = \mu_Y \geq 0$. Prove the inequality

$$h(X|Y) \leq \log_2 (e(\mu_X - \mu_Y)).$$

P5 (2 points). Generate two random codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ with 2^{nR_1} and 2^{nR_2} codewords of length n symbols, where the symbols are generated i.i.d. with probabilities $p(x_1)$ and $p(x_2)$, respectively. Define the set

$$\mathcal{B} = \{(x_1^n, x_2^n) : \ x_1^n \in \mathcal{C}_1, \ x_2^n \in \mathcal{C}_2, \ (x_1^n, x_2^n) \in A_\epsilon^{(n)}\}$$

where $A_\epsilon^{(n)}$ is the set of jointly typical sequences with respect to a joint pmf $p(x_1, x_2)$ with marginal pmfs $p(x_1)$ and $p(x_2)$. Show that

$$2^{n(R_1+R_1-I(X_1;X_2)-3\epsilon)} \leq E[|\mathcal{B}|] \leq 2^{n(R_1+R_1-I(X_1;X_2)+3\epsilon)}$$

for n large enough. Please provide all the details of the proof (you can use the theorems proved in class).

P6 (1 point). We wish to transmit a $Ber(p)$ process V^n over n channel uses of a binary symmetric channel with crossover probability 0.3. Find necessary and sufficient conditions on p such that V^n can be estimated with vanishing error probability at the receiver.

P7 (1 point). We have the two independent random variables X_1 and X_2 with pmfs $p(x)$ and $q(x)$, respectively, with $x \in \mathcal{X}$. Prove that

$$\Pr[X_1 = X_2] \geq 2^{-(H(p(x))+D(p(x)||q(x)))}.$$

(Hint: $p(x) = 2^{\log_2 p(x)}$).

P8 (1 point) For any source (not necessarily Gaussian) with differential entropy $h(X)$, prove the following bound:

$$R(D) \geq h(X) - \frac{1}{2}\log_2(2\pi e D).$$

When do we have equality? Therefore, is a Gaussian source with the same differential entropy $h(X)$ easier or more difficult to compress?

P9 (2 points) Consider the Gaussian channel with received signal $Y = (Y_1, Y_2)$, where

$$\begin{aligned} Y_1 &= X + Z_1 \\ Y_2 &= X + Z_2, \end{aligned}$$

the power constraint is P and the noises (Z_1, Z_2) are independent and with variances N_1 and N_2.
a. Calculate the capacity.
b. Can we simplify the decoder so that it operates on a single received symbol rather than two for each channel use?

Sol.:
P1. a. The process is stationary but not ergodic. We have that the quantity $1/n\sum_{i=1}^n X_i$ tends to $1/\lambda$ with probability 1/2 and $-1/\lambda$ with probability 1/2 by the (strong) law of large numbers.
b. We have

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}h(X^n) &= \lim_{n\to\infty}\frac{1}{n}h(X^n, U). \\ &= \lim_{n\to\infty}\frac{1}{n}h(X^n|U) + 1 \\ &= \log_2\left(\frac{e}{\lambda}\right) \cancel{+ 1}, \end{aligned}$$

where the first equality holds since U is a function of X^n (except on a set of measure zero).

P2. We have

$$\begin{aligned} H(X_1, ..., X_n) &= H(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) + H(X_i|X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) \\ &\leq H(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) + H(X_i|X_1, ..., X_{i-1}). \end{aligned}$$

Summing over i, we get

$$nH(X_1, ..., X_n) \leq \sum_{i=1}^n H(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) + H(X_1, ..., X_n),$$

which is the desired result.

P3. Please see lecture notes.

P4. We have

$$
\begin{aligned}
h(X|Y) &= \int h(X|Y=y)f(y)dy \\
&= \int h(X-y|Y=y)f(y)dy \\
&= h(X-Y|Y) \\
&\leq h(X-Y) \\
&\leq \log_2\left(e(\mu_X - \mu_Y)\right),
\end{aligned}
$$

where the second inequality follows by the maximum entropy theorem since $X-Y$ has support $[0,\infty)$ and mean $E[X-Y] = \mu_X - \mu_Y$.

P5. Observe that for given codebooks, we have

$$
|\mathcal{B}| = \sum_{w_1=1}^{2^{nR_1}} \sum_{w_1=1}^{2^{nR_2}} 1\{(x_1^n(w_1), x_2^n(w_2)) \in A_\epsilon^{(n)}\},
$$

where $1\{x\} = 0$ if x is false and 1 otherwise. We can then write

$$
\begin{aligned}
E[|\mathcal{B}|] &= \sum_{w_1=1}^{2^{nR_1}} \sum_{w_1=1}^{2^{nR_2}} E[1\{(X_1^n(w_1), X_2^n(w_2)) \in A_\epsilon^{(n)}\}] \\
&= \sum_{w_1=1}^{2^{nR_1}} \sum_{w_1=1}^{2^{nR_2}} \Pr\{(X_1^n(w_1), X_2^n(w_2)) \in A_\epsilon^{(n)}\}.
\end{aligned}
$$

But we know that

$$
(1-\epsilon)2^{-n(I(X_1;X_2)+3\epsilon)} \leq \Pr\{(X_1^n(w_1), X_2^n(w_2)) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X_1;X_2)-3\epsilon)},
$$

where the first inequality holds for n large enough. The desired result follows immediately.

P6. The necessary and sufficient conditions on p is

$$
H(0.3) \leq 1 - H(p),
$$

from which we can derive the condition on p.

P7. We have

$$
\begin{aligned}
\Pr[X_1 &= X_2] = \sum_x p(x)q(x) \\
&= \sum_x p(x) 2^{\log_2 q(x)} \\
&\leq 2^{\sum_x p(x)\log_2 q(x)} \\
&= 2^{\sum_x p(x)\log_2\left(q(x)\frac{p(x)}{p(x)}\right)} \\
&= 2^{\sum_x p(x)\log_2\left(q(x)\frac{p(x)}{p(x)}\right)} \\
&= 2^{-H(p(x)) - D(p(x)||q(x))}.
\end{aligned}
$$

P8. The result follows from the inequality

$$\begin{aligned} I(X;\hat{X}) &= h(X) - h(X|\hat{X}) \\ &\geq h(X) - \frac{1}{2}\log_2(2\pi e D), \end{aligned}$$

where the inequality can be derived as seen in class. We have equality if X is Gaussian. Therefore, a Gaussian source with the same differential entropy $h(X)$ is easier to compress.

P9. Following the same steps seen in class we get

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{N_1} + \frac{P}{N_2}\right).$$

Moreover, it can be checked that a decoder that operates with the received signal $\frac{1}{\sqrt{N_1}}Y_1 + \frac{1}{\sqrt{N_2}}Y_2$ achieves the same capacity.