Random codes in communication 22 July 2012
(Dr. Permuter Haim)

Final Exam, MOED GIMEL (with partial solutions)
Total time for the exam: 3 hours!

1) **Deletion Channel** (15 points)
Consider a binary sequence of length n, denoted by $X^n = (X_1, \ldots, X_n)$. Consider another binary sequence of length n called deletion pattern, denoted by $D^n = (D_1, \ldots, D_n)$, which determines how X_n is to be deleted. Then, the output of the deletion process, denoted by $y(X^n, D^n)$, is derived from X^n by deleting the bits at those locations where the deletion pattern is 1.
Consider the following example for $n = 10$:

$$\begin{aligned} X^n &= (0,1,1,0,0,0,1,1,1,0) \\ D^n &= (0,1,0,1,1,0,0,0,1,0) \\ y(X^n, D^n) &= (0,1,0,1,1,0) \end{aligned}$$

The source sequence $X^n \in \{0,1\}^n$ is i.i.d. Bernoulli$(1/2)$, and the deletion pattern $D^n \in \{0,1\}^n$ is i.i.d. Bernoulli(d), independent of X^n.
We are interested in computing the mutual information between X^n and $y(X^n, D^n)$, $I(X^n; y(X^n, D^n))$. For each relation state if its true or false:

a) $I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) + H(D^n) - H(y(X^n, D^n)|X^n, D^n)$
b) $I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) - H(D^n) + H(D^n|X^n, y(X^n, D^n))$
c) $I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) - H(D^n)$
d) $I(X^n; y(X^n, D^n)) = H(X^n) - H(X^n|y(X^n, D^n)) - H(D^n|X^n, y(X^n, D^n))$

Solution Only (b) is true.

$$\begin{aligned} I(y(X^n, D^n); X^n, D^n) &= H(y(X^n, D^n)) - H(y(X^n, D^n)|X^n, D^n) \\ &= H(y(X^n, D^n)) \\ &= H(X^n, D^n) - H(X^n, D^n|y(X^n, D^n)) \\ &= H(X^n) + H(D^n) - H(X^n|y(X^n, D^n)) - H(D^n|X^n, y(X^n, D^n)) \\ &= I(X^n; y(X^n, D^n)) + H(D^n) - H(D^n|X^n, y(X^n, D^n)) \end{aligned}$$

Rearranging the above terms, we get the desired result

$$I(X^n; y(X^n, D^n)) = H(y(X^n, D^n)) - H(D^n) + H(D^n|X^n, y(X^n, D^n))$$

2) **Bound on each Huffman codeword** (15 points)
The following claim is sometimes found in the literature:

It can be shown that the length of the binary Huffman codeword of a symbol a_i with probability p_i is always less than or equal to $\lceil -\log_2 p_i \rceil$

Is this claim true? Justify your answer
Solution:
This claim is not true in general. To prove it, we shall construct a counterexample. Choose a small number ϵ. Consider the following source X with 4 symbols:

$$X = \begin{cases} x_1, & \text{w.p. } p_1 = \epsilon \\ x_2, & \text{w.p. } p_2 = \frac{1}{3} - \epsilon \\ x_3, & \text{w.p. } p_3 = \frac{1}{3} - \epsilon \\ x_4, & \text{w.p. } p_4 = \frac{1}{3} + \epsilon \end{cases}$$

Now consider the Huffman code corresponding to the above source. Its length function will be given by : $l(x_1) = l(x_2) = 3$, $l(x_3) = 2$ and $l(x_4) = 1$. Now for x_2, the Shannon code length would be $\lceil \log \frac{1}{\frac{1}{3}-\epsilon} \rceil = 2$, for $\epsilon < \frac{1}{12}$. Thus, the binary Huffman codeword length can exceed the Shannon codeword length in general for a symbol.

3) **Condition on length of Huffman codeword to be larger then 1** (20 points)
Consider a source of K symbols, with $p_1 \geq p_2 \geq \ldots \geq p_K$. Find the largest q s.t. $p_1 < q$ implies $l_1 > 1$, where l_i is the length of the binary Huffman codeword associated with symbol i.
Solution:
First we are going to show that in any binary Huffman encoding if $p_1 < \frac{1}{3}$, then the first symbol must be encoded by a codeword length at least 2. Suppose a_1 is encoded by a codeword of length 1. This means that a_1 is only encoded at the last step of the algorithm. Consider the step of the algorithm where exactly 3 symbols are left, say a_1, a_2' and a_3' with probabilities p_1, p_2' and p_3'. Since at this point we merge symbols a_2' and a_3', we must have that p_2',$p_3' \leq p_1$. But then it follows that $1 = p_1 + p_2' + p_3' \leq 3p_1$, so $p_1 \geq \frac{1}{3}$.
Hence, we just proved that if $l_1 = 1$ then $p_1 \geq \frac{1}{3}$, or equivalently, if $p_1 < \frac{1}{3}$, then $l_1 > 1$.
The question that remains is whether $q = \frac{1}{3}$, or there is a $q > \frac{1}{3}$ so that $l_1 > 1$. Yet, we observe that for $(p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ we can construct a Huffman code where $l_1 = 2$. Therefore $q = \frac{1}{3}$.

4) **Power loading with a cost** (30 points)
Consider n parallel additive white Gaussian noise (AWGN) point-to-point channels. For the i-th

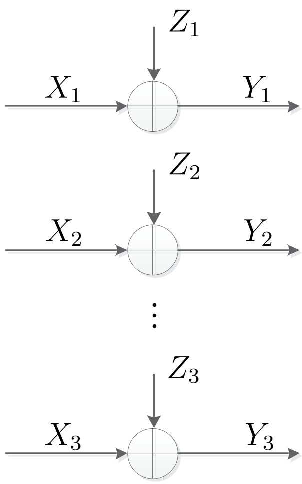

Fig. 1. The n parallel AWGN channels.

channe (where $i \in \{1, 2, \ldots, n\}$) the input and output are denoted by X_i and Y_i respectively and the additive Gaussian noise, which is denoted by Z_i, is distributed according to $Z_i \sim N(0, N_i)$. The channels are additive in the sense that:

$$Y_i = X_i + Z_i, \ \forall \ i \in \{1, 2, \ldots, n\}. \tag{1}$$

The channel is illustrated in Fig. 1 The capacity of this setting is given by:

$$C = \frac{1}{2} \sum_{i=1}^{n} \log \left(1 + \frac{P_i}{N_i}\right), \tag{2}$$

which is achieved by taking the input vector $\mathbf{X} = (X_1, X_2, \ldots, X_n) \sim N(0, \mathbf{\Sigma})$, where the covariance is:

$$\Sigma = \begin{pmatrix} P_1 & 0 & \ldots & 0 \\ 0 & P_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & P_n \end{pmatrix} \tag{3}$$

and is bound to the power constraint:

$$\sum_{i=1}^{n} P_i \leq P. \tag{4}$$

In order to find the optimal distribution of powers $\{P_i\}_{i=1}^n$ between the channels, the following optimization problem is solved:

$$\max_{\{p_i\}_{i=1}^n} \frac{1}{2} \sum_{i=1}^{n} \log\left(1 + \frac{P_i}{N_i}\right), \tag{5}$$

subject to

$$\begin{gathered} \textstyle\sum_{i=1}^{n} P_i \leq P, \\ P_i \geq 0, \ \forall\, i \in \{1, 2, \ldots, n\}. \end{gathered}$$

Solving the optimization in (5) yields the famous ”Water-Filling” solution.
In this question we consider an extension of the classical Water-Filling problem. Here we wish to allocate power to the n parallel channels where the constraint is on the total price (and not power). Lets $\{\beta_i\}_{i=1}^n$ be known, non-negative constants which represent the cost per unit of power in channel i. The transmission is bound to a total cost constraint:

$$\sum_{i=1}^{n} \beta_i P_i \leq B. \tag{6}$$

As in the classical problem, the solution must take into account the non-negativity of the power allocated to each channel, i.e. we must have $P_i \geq 0$ for every $i \in \{1, 2, \ldots, n\}$.

a) State the optimization problem (and the constraints) for the extended setting.
b) Construct the Lagrangian function for this maximization problem.

c) Solve the optimization problem using Lagrange multipliers technique in order to find P_i as a function of (β_i, N_i), for every $i \in \{1, 2, \ldots, n\}$.
You may use the function x^+ which is defined by:

$$x^+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

d) Explain the interpretation of the price allocation for each channel. How does it differs form the classical Water-Filling solution? How should one choose $\{\beta_i\}_{i=1}^n$ and B so that the solution to the extended problem will reduce into the classical one?
e) Now let us consider only two parallel channels, i.e. $n = 2$. Moreover, it is now given that $B = 10$, $N_1 = 3$, $N_2 = 2$, $\beta_1 = 1$, $\beta_2 = 2$. Find (P_1, P_2) and calculate the capacity of channel.

Solutions

$$\begin{aligned} \max_{\{p_i\}_{i=1}^n} \quad & \frac{1}{2}\sum_{i=1}^{n} \log\left(1+\frac{P_i}{N_i}\right), \\ s.t. \quad & \sum_{i=1}^{n} \beta_i P_i \leq B., \\ & P_i \geq 0, \;\forall\; i \in \{1,2,\ldots,n\}. \end{aligned} \tag{7}$$

Notice that this problem is equivalent to

$$\begin{aligned} \max_{\{\beta_i p_i\}_{i=1}^n} \quad & \frac{1}{2}\sum_{i=1}^{n} \log\left(1+\frac{\beta_i P_i}{\beta_i N_i}\right), \\ s.t. \quad & \sum_{i=1}^{n} \beta_i P_i \leq B., \\ & \beta_i P_i \geq 0, \;\forall\; i \in \{1,2,\ldots,n\}. \end{aligned} \tag{8}$$

(Justify why the optimization problem in (7) is equivalent to the optimization problem in (8). Now, we can use a change of variable. Denote

$$P_i' = \beta_i P_i. \tag{9}$$

We get

$$\begin{aligned} \max_{\{p_i'\}_{i=1}^n} \quad & \frac{1}{2}\sum_{i=1}^{n} \log\left(1+\frac{P_i'}{\beta_i N_i}\right), \\ s.t. \quad & \sum_{i=1}^{n} P_i' \leq B., \\ & P_i' \geq 0, \;\forall\; i \in \{1,2,\ldots,n\}. \end{aligned} \tag{10}$$

Now you can solve the problem in (10) as the parallel gaussian channel problem in the lecture notes just using $\beta_i N_i$ rather then N_i and the solution would be for p_i'. Finally, once you have p_i' use (9) to find P_i.

5) **Entropic Sources** (20 points)
Consider a (not necessarily memoryless) source **Z**. Let $H(Z^n)$ denote the entropy of an n-tuple from this source.
Definition 1: A source **Z** is said to be **subentropic** if the total mass of its most likely $\lfloor 2^{(1+\delta)H(Z^n)} \rfloor$ outcomes goes to 1 as $n \to \infty$ for any $\delta > 0$.
Definition 2: A source is **superentropic** if the total mass of its most likely $\lfloor 2^{(1-\delta)H(Z^n)} \rfloor$ outcomes does not go to 1 as $n \to \infty$ for any $\delta > 0$.
Definition 3: A source is **entropic** if it is both subentropic and superentropic.
Answer the following questions: I.

a) Any source is either subentropic or superentropic or both. **True or False ?**
b) For each of the following sources, classify the source as
   a. subentropic but not superentropic
   b. superentropic but not subentropic
   c. entropic
   d. none of the above.

i) $Z^n = (X_1, \cdots, X_n)$ is an i.i.d. source with finite alphabet.

ii) For $0 < q < 1$,

$$P[Z^n = z^n] \;=\; \begin{cases} 1-q & \text{if } z^n = (0, \cdots 0) \\ q/(2^n - 1) & \text{if } z^n \neq (0, \cdots 0). \end{cases} \tag{11}$$

iii) Z^n has four types of outcomes (a)

A) one mass with probability $\frac{1}{2}$;
B) 2^n masses each with probability $(\frac{1}{2} - \frac{1}{n})2^{-n}$;
C) $\lfloor \frac{1}{n} 2^{n^2/2} \rfloor$ masses each with probability $2^{-n^2/2}$;
D) one mass with the remaining probability if $\frac{1}{n}2^{n^2/2}$ is not integer valued.

Solution

a) Note that if a source is not subentropic, then the probability of a set of size $\lfloor 2^{(1+\delta)H(Z^n)} \rfloor$ most likely outcomes is not converging to one. Then a fortiori the mass of the smaller set of size $\lfloor 2^{(1-\delta)H(Z^n)} \rfloor$ most likely outcomes cannot go to one. Thus the source has to be superentropic. Thus no source can be, at the same time, both not subentropic and not superentropic.. Therefore, the correct answer is **True**.

i) Correct answer is (c).
Note that since the Z_i's are i.i.d. on a finite alphabet, the Asymptotic Equipartition Property applies directly. First, we note that $H(Z^n) = nH(Z_1)$. Now we verify that the source is subentropic. Fix a $\delta > 0$. Choose any $\epsilon > 0$ such that $nH(Z_1)(1+\delta) > n(H(Z_1) + \epsilon)$. Consider the typical set $A_\epsilon^{(n)}$. Clearly,

$$P(\text{most likely } \lfloor 2^{(1+\delta)H(Z^n)} \rfloor \text{outcomes}) \geq P(A_\epsilon^{(n)}). \tag{12}$$

Since $P(A_\epsilon^{(n)}) \to 1$, the source is subentropic.
On the other hand, we have shown in class that any set no larger than $2^{n(1-\delta)H(Z_1)}$ has vanishing probability as $n \to \infty$. So in particular,

$$P(\text{most likely } \lfloor 2^{(1-\delta)H(Z^n)} \rfloor \text{outcomes}) \to 0. \tag{13}$$

Therefore, the source is entropic, as it is both subentropic and superentropic.

ii) The correct answer is (b).
Note for the given source that,

$$\begin{aligned} H(Z^n) &= h(q) + qn + q\log(1 - 2^{-n}), & (14) \\ &= qn + o(n). & (15) \end{aligned}$$

So the probability of the $2^{(1+\delta)H(Z^n)}$ most likely outcomes is upper bounded by that of the $2^{(1+2\delta)qn}$ most likely outcomes, which is equal to

$$1 - q + \left(2^{(1+2\delta)qn} - 1\right)\frac{q}{2^n - 1} \to 1 - q, \tag{16}$$

for every $\delta > 0$ sufficiently small. Thus, the source is not subentropic (and hence is superentropic), making (b) the correct choice.

iii) The correct answer is (c).
Recall that the source has four outcomes: (a)

A) one mass with probability $\frac{1}{2}$;
B) 2^n masses each with probability $(\frac{1}{2} - \frac{1}{n})2^{-n}$;
C) $\lfloor \frac{1}{n} 2^{n^2/2} \rfloor$ masses each with probability $2^{-n^2/2}$;
D) one mass with the remaining probability if $\frac{1}{n}2^{n^2/2}$ is not integer valued.

We can compute the entropy $H(Z^n)$, for large n to be

$$\begin{aligned} H(Z^n) &= \frac{1}{2}\log 2 + (\frac{1}{2} - \frac{1}{n})\log(2^n(1/2 - 1/n)^{-1}) + \frac{1}{n}\log\frac{1}{n}2^{n^2/2} \\ &= \frac{1}{2} + (\frac{1}{2} - \frac{1}{n})n + (\frac{1}{2} - \frac{1}{n})\log((1/2 - 1/n)^{-1}) + \frac{n}{2} - \frac{1}{n}\log n \\ &= n + o(n). \end{aligned}$$

Note that the rest of the $o(n)$ terms are sublinear, and do not contribute to the largest exponent in the size of the set, which in this case is n. Thus, a set of size $2^{n(1+\delta)}$, consisting of the most likely outcomes of Z^n, will include the masses in (a) and (b). The total mass of this set will exceed $\frac{1}{2} + (\frac{1}{2} - \frac{1}{n}) = 1 - \frac{1}{n}$. This clearly goes to 1 as $n \to \infty$. Thus, the source is subentropic by Definition 1.
Now consider a set of size $2^{n(1-\delta)}$. It will include the mass in (a) and a fraction of the masses in (b). In particular, the total probability of such a set would not exceed $\frac{1}{2} + (\frac{1}{2} - \frac{1}{n})2^{-n\delta}$ which converges to $\frac{1}{2}$ as $n \to \infty$. Thus the source is also superentropic by Definition 2. Hence the answer is (c) by Definition 3.

Good Luck!