ECE 612: Information Theory
2018 Midterm

Instructor: Patrick Mitran

- This exam consists of 5 problems and XYZ pages.
- The exam duration is 2 hours (120 min).
- You may bring one letter or A4 size sheet of paper with any handwritten notes that you wish (on both sides). There may not be any photocopied content on the sheet.
- You can neither communicate nor collaborate with others during the exam.
- Questions are allowed but will be answered only if you cannot understand the statement of a problem. We will not comment on any other issues.
- Any person that is found cheating on the exam will be immediately reported to the authorities.
- All answers must be written legibly. We reserve our right to reduce your grade if your answer is not written in a legible manner.
- A final correct answer does not mean much to us, if the corresponding approach is not clear and sensible. Please explain your solutions and convince us that your solutions make sense.
- We wish you the best of luck in this exam and also in your academic endeavor.

1
2
3
4
5

Problem I: Let p_X and q_X be two probability mass functions (PMFs) on $\mathcal{X}$ with $|\mathcal{X}|$ finite.

a) Assume $p_X(x) > 0$ and $q_X(x) > 0$ for all $x \in \mathcal{X}$. Let $X_1, X_2, \ldots,$ be drawn independently and identically (iid) according to p_X. Does the following expression converge in probability as $n \to \infty$? If so, to what and why? If not, why not?

$$\frac{1}{n} \log[q_X(X_1) q_X(X_2) \cdots q_X(X_n)].$$

b) With p_X fixed, consider the function of q_X:

$$f(q_X) = \sum_{x \in \mathcal{X}} p_X(x) \log q_X(x).$$

Show that $f(q_X)$ takes its maximum when $q_X = p_X$, i.e., show that

$$\max_{q_X} f(q_X) = f(p_X).$$

Is this maximum unique? Explain.

Solution:

a)

$$\begin{aligned}
\frac{1}{n} \log q_X(X_1, \ldots, X_n) &= \frac{1}{n} \log q_X(X_1) q_X(X_2) \cdots q_X(X_n) \\
&= \frac{1}{n} \sum_{i=1}^{n} \log q_X(X_i) \\
&\to E_{X \sim p_X}[\log q_X(X)] \text{ (in probability by WLLN)} \\
&= \sum_{x \in \mathcal{X}} p_X(x) \log q_X(x).
\end{aligned}$$

b)

$$\begin{aligned}
f(p_X) - f(q_X) &= \sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) - \sum_{x \in \mathcal{X}} p_X(x) \log q_X(x) \\
&= \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)} \\
&= D(p_X || q_X) \\
&\geq 0
\end{aligned}$$

with equality in the last step iff $D(p_X||q_X)$, or equivalenty, iff $p_X = q_X$, so the maximum is unique.

Problem II: Let p_X and q_X be two probability mass functions (PMFs) for the random variable X, and let $p_{Y|X}$ and $q_{Y|X}$ be two conditional PMFs for the random variable Y given X. Assume the following is known:

- $p_X(x) > 0$ **for all** $x \in \mathcal{X}$**,**
- $q_X(x) > 0$ **for all** $x \in \mathcal{X}$**,**
- $I(p_X; p_{Y|X}) = 1$**,**
- $I(q_X; p_{Y|X}) = 2$**,**
- $I(p_X; q_{Y|X}) = 0$**.**

a) Based on this known information, provide a bound on $I(\frac{1}{2}p_X + \frac{1}{2}q_X; p_{Y|X})$ other than the trivial bound that it is non-negative.
b) Based on this known information, what can you say about $I(q_X; q_{Y|X})$?

Solution:

a) Since $I(r_X; r_{Y|X})$ is concave in r_X:

$$I(\frac{1}{2}p_X + \frac{1}{2}q_X; p_{Y|X}) \geq \frac{1}{2}I(p_X; p_{Y|X}) + \frac{1}{2}I(q_X; p_{Y|X}) = 1/2 + 2/2 = 3/2.$$

b) Since $I(p_X; q_{Y|X}) = 0$, then this implies that X and Y are independent under the distribution $p_X(x)q_{Y|X}(y|x)$, i.e., $p_X(x)q_{Y|X}(y|x) = p_X(x)q_Y(y)$ or, since $p_X(x) > 0$, $q_{Y|X}(y|x) = q_Y(y)$. Therefore, $q_X(x)q_{Y|X}(y|x) = q_X(x)q_Y(y)$, i.e., they are independent again. So $I(q_X; q_{Y|X}) = 0$.

Problem III: Let $X_0, X_1, X_2, \ldots$ be a stationary random process. For each of the three statements below, either prove the result, or give an example that shows the statement is not always true.

a) $H(X_m|X_\ell) = H(X_\ell|X_m)$ **for all** $m \geq 0, k \geq 0$**.**

b) $H(X_iX_j|X_k) = H(X_k|X_iX_j)$ **for all** $i \geq 0, j \geq 0, k \geq 0$**.**

c) $H(X_iX_j|X_kX_\ell) = H(X_kX_\ell|X_iX_j)$ **for all** $i \geq 0, j \geq 0, k \geq 0, \ell \geq 0$**.**

Solution:

a)

$$
\begin{aligned}
I(X_m; X_\ell) &= H(X_m) - H(X_m|X_\ell) \\
&= H(X_\ell) - H(X_\ell|X_m).
\end{aligned}
$$

Thus, $H(X_m) - H(X_m|X_\ell) = H(X_\ell) - H(X_\ell|X_m)$. Now, since the process is stationary, $H(X_m) = H(X_\ell)$. Thus, we must have $H(X_m|X_\ell) = H(X_\ell|X_m)$.

b) Let X_ℓ be a sequence of iid random variables. Then $H(X_iX_j|X_k) = H(X_i)+H(X_j) = 2H(X_1)$ and $H(X_k|X_iX_j) = H(X_k) = H(X_1)$. These are not equal unless $H(X_1) = 0$. Hence, if X_1 is a Bernoulli random variable with parameter 1/2, then $H(X_1) = 1$ provides a counterexample.

c) Let X_0, X_1, X_2 be iid Bernoulli random variables with parameter 1/2. For $n \geq 0$ let

$$
X_n = \begin{cases} X_0 & \text{when } (n \mod 3) = 0 \\ X_1 & \text{when } (n \mod 3) = 1 \\ X_2 & \text{when } (n \mod 3) = 2 \end{cases}
$$

Then $H(X_0X_1|X_2X_5) = H(X_0X_1) = 2$, but $H(X_2X_5|X_0X_1) = H(X_2X_5) = H(X_2) = 1$.

Problem IV: In class we saw that a prefix (or instantaneous) code over a D-ary alphabet must satisfy the Kraft inequality

$$\sum_i D^{-\ell_i} \leq 1$$

where $\ell_1, \ldots, \ell_m$ are the lengths of the codewords.

Now consider the case that we wish to "reserve space for future expansion of the code", i.e., suppose we wish to optimize

$$\min_{\{\ell_i\}} \sum_{i=1}^{m} p_i \ell_i$$

$$\textbf{subject to:} \qquad \sum_i D^{-\ell_i} \leq f$$

where f is a known constant $0 < f < 1$ that accounts for the amount of space we want to reserve for future use and $\ell_1, \ldots, \ell_m$ are non-negative integers.

a) Solve the optimization problem for the case when the lengths $\ell_1, \ldots, \ell_m$ are relaxed to real values. For the relaxed problem, what are the optimal lengths $\ell_1^*, \ldots, \ell_m^*$ and what is $L^* = \sum_{i=1}^m p_i \ell_i^*$?

b) Under what conditions on f and $p_1, \ldots, p_m$ is the solution in a) optimal for the original problem where $\ell_1, \ldots, \ell_m$ must be integers?

c) Design a good prefix code for the case that $D = 2$, $p_1 = 4/7$, $p_2 = p_3 = p_4 = 1/7$ and $f = 7/8$. Specifically, provide the 4 codewords of the good prefix code.

Solution:

a) The optimal solution will have equality in the constraint since otherwise, we could do better. The Lagrangian is then:

$$J = \sum_i \ell_i p_i + \lambda(\sum_i D^{-\ell_i} - f).$$

Taking derivatives with respect to ℓ_j results in:

$$p_j - \lambda D^{-\ell_j} \ln D = 0,$$

from which we get

$$D^{-\ell_j} = \frac{p_j}{\lambda \ln D},$$

and therefore

$$\sum_i D^{-\ell_j} = \sum_i \frac{p_i}{\lambda \ln D} = \frac{1}{\lambda \ln D} = f$$

and hence $\lambda = \frac{1}{f \ln D}$. Substituting this back in, we get

$$D^{-\ell_j} = f p_j,$$

or $\ell_j^* = -\log_D p_j - \log_D f$, and

$$L^* = \sum_{i=1}^{m} p_i \ell_i^* = -\sum_i p_i [\log_D p_j + \log_D f] = H_D(X) + \log_D(1/f),$$

so we pay a penalty of $\log_D(1/f)$ in rate.

b) We have equality provided $\ell_j = -\log_D(f p_j)$ are all integers. (Note: this is always non-negative since both $f \leq 1$ and $p_j \leq 1$). This is equivalent to $f p_j = D^{-\ell_j}$ for integers ℓ_j.

c) Let's pick the lengths to be $\ell_j = \lceil -\log_2(f p_j) \rceil$. Then

$$\ell_1 = \lceil -\log_2(7/8 \times 4/7) \rceil = 1$$

$$\ell_2 = \lceil -\log_2(7/8 \times 1/7) \rceil = 3$$

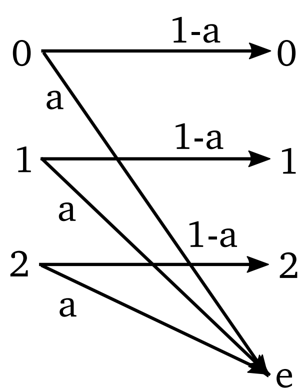
$$\ell_3 = \lceil -\log_2(7/8 \times 1/7) \rceil = 3$$

$$\ell_4 = \lceil -\log_2(7/8 \times 1/7) \rceil = 3$$

So, without loss of generality, we pick $c_1 = 0$. The next codewords must start with a 1. So we pick $c_2 = 100$, $c_3 = 101$ and $c_4 = 110$. This leaves codewords that start with 111 for future expansion, and these "occupy 1/8th of the codespace".

Problem V: Consider the ternary erasure channel with $\mathcal{X} = \{0, 1, 2\}$ and $\mathcal{Y} = \{0, 1, 2, e\}$ where the output $Y = e$ denotes an erasure. The channel has $p_{Y|X}$ as below where a is a known constant.

$$p_{Y|X}(y|x) = \begin{cases} 1-a & y = x \\ a & y = e \end{cases}$$



What is the channel capacity in terms of the constant a? Prove your answer.

Solution 1: One could try to prove this by setting $p_X = (b_0, b_1, b_2)$ where the latter is a probability vector, and then optimizing $I(X;Y)$ over b_0, b_1 and b_3 subject to the constraint $1 = b_0 + b_1 + b_2$, but this is rather complicated.

Instead, consider

$$p_X^0 = (b_0, b_1, b_2)$$
$$p_X^1 = (b_1, b_2, b_0)$$
$$p_X^2 = (b_2, b_0, b_1).$$

Then by symmetry

$$I(p_X^0; p_{Y|X}) = I(p_X^1; p_{Y|X}) = I(p_X^2; p_{Y|X}).$$

Hence, by concavity of $I(p_X; p_{Y|X})$ in p_X:

$$I(\frac{1}{3}p_X^0 + \frac{1}{3}p_X^1 + \frac{1}{3}p_X^2; p_{Y|X}) \geq \frac{1}{3}I(p_X^0; p_{Y|X}) + \frac{1}{3}I(p_X^1; p_{Y|X}) + \frac{1}{3}I(p_X^2; p_{Y|X})$$
$$= I(p_X^0; p_{Y|X}).$$

But, $\frac{1}{3}p_X^0 + \frac{1}{3}p_X^1 + \frac{1}{3}p_X^2 = (1/3, 1/3, 1/3)$, i.e., the uniform distribution on X. Hence,

$$I((\frac{1}{3}, \frac{1}{3}, \frac{1}{3}); p_{Y|X}) \geq I(p_X^0; p_{Y|X}).$$

Since this is true for any p_X^0, the uniform distribution maximizes $I(X;Y)$. Since $p_X = (1/3, 1/3, 1/3)$ maximizes capacity, we compute $I(X;Y) = H(Y) - H(Y|X)$ under this input distribution. Then,

$$p_Y = \left(\frac{1-a}{3}, \frac{1-a}{3}, \frac{1-a}{3}, a\right),$$

hence

$$\begin{aligned} H(Y) &= 3 \times -\frac{1-a}{3} \log \frac{1-a}{3} - a \log a \\ &= -(1-a) \log \frac{1-a}{3} - a \log a \\ &= -(1-a) \log(1-a) - a \log a + (1-a) \log 3 \\ &= h(a) + (1-a) \log 3. \end{aligned}$$

Also,

$$H(Y|X) = \sum_x H(Y|X=x) p_X(x) = \sum_x h(a) \frac{1}{3} = h(a).$$

Therefore

$$C = h(a) + (1-a) \log 3 - h(a) = (1-a) \log 3.$$

Solution 2: This approach is similar to the BEC in class.

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

and

$$H(Y|X) = \sum_x H(Y|X=x) p_X(x) = \sum_x h(a) p_X(x) = h(a),$$

then

$$C = \max_{p_X} [H(Y) - h(a)].$$

Now, define the randome variable $E = 1(Y = e)$, i.e., $E = 1$ if $Y = e$ and $Y = 0$ otherwise. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E). \end{aligned}$$

Now $H(E) = h(a)$ since $P[Y = e] = \sum_x P[Y = e|X = x]p_X(x) = \sum_x ap_X(x) = a$. Also, if $Y \neq e$, then $Y = X$. Hence

$$\begin{aligned} H(Y|E) &= H(Y|E = 0)P[E = 0] + H(Y|E = 1)P[E = 1] \\ &= H(X)P[E = 0] + 0P[E = 1] \\ &= H(X)P[E = 0] \\ &= H(X)(1 - a) \end{aligned}$$

Hence,

$$\begin{aligned} C &= \max_{p_X} [H(Y) - h(a)] \\ &= \max_{p_X} [h(a) + (1 - a)H(X) - h(a)] \\ &= \max_{p_X} (1 - a)H(X) \\ &= (1 - a) \log 3. \end{aligned}$$