EE5139 Information Theory for Communication Systems
Solutions for Final Exam 2020

1 Question 1: Warmup (Total: 12 points)

The following are TRUE/FALSE questions. Explanations are not needed. Please just write TRUE or FALSE. Please indicate the part of the question $(1, 2$, etc.) in your answer script clearly.

1. (1 point) We always have $I(X;Y \mid Z) \geq I(X;Y)$.

   FALSE.

2. (1 point) We have $D(P\|Q) \neq 0$ unless $P = Q$ for any two pmfs P and Q.

   TRUE.

3. (1 point) There is a binary prefix-free variable-length source code with lengths (1,3,3,3,4,4) .

   TRUE.

4. (1 point) It is possible to reliably communicate a source X with entropy $H(X) = 1.2$ bits per source symbol over an additive white Gaussian noise channel with SNR 3 .

   FALSE.

5. (1 point) The capacity of a DMC with input alphabet $\mathcal{X} = \{a, b, c\}$ and output alphabet $\mathcal{Y} = \{d, e, f, g\}$ can be 2 bits per channel use.

   FALSE.

6. (1 point) In channel coding, if we transmit at a rate R strictly below capacity C, then there is a sequence of codes such that the error probability vanishes.

   TRUE.

7. (2 points) There exists a binary error correcting code with $n = 5, d = 3$ and $|C| = 8$.

   FALSE.

8. (1 point) For a discrete real-valued random variable X, we have $H(X) = H(cX)$ for all $c \in \mathbb{R}$ with $c \neq 0$

   TRUE.

9. (1 point) For a continuous real-valued random variable X, we have $h(X) = h(cX)$ for all $c \in \mathbb{R}$ with $c \neq 0$

   FALSE.

   Please answer the following question in at most two sentences.

10. (2 points) Explain the difference between a strong and a weak converse for the example of source coding for a DMC.

    The strong converse prohibits any compression at a rate below H(X) with nonzero success probability. The weak converse only prohibits such compression with vanishing error.

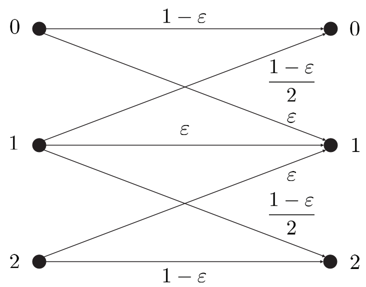
2 Question2: Channel capacity (Total: 11 points)

Let W be a ternary channel with input and output symbols $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$ given by

$$W(y \mid x) = \begin{cases} 1-\varepsilon & \text{if } x = y \neq 1 \\ \frac{1-\varepsilon}{2} & \text{if } x = 1 \text{ and } y \neq 1 \\ \varepsilon & \text{if } y = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

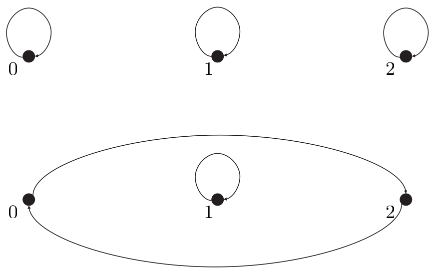
1. (1 point) Draw a diagram with the transition probabilities for this channel.

   Solution:



2. (2 points) Observe any symmetries of the channel. List the permutations π on $\mathcal{X}$ for which there exists a permutation π' on $\mathcal{Y}$ such that $W(y \mid x) = W(\pi'(y) \mid \pi(x))$ for all x, y. Permutations on $\mathcal{X}$ are conveniently represented as directed graphs with the elements of $\mathcal{X}$ as nodes and the edges showing the transitions.

   Solution:



3. (3 points) We are interested in the channel mutual information, $I(W) = \max_{P \in \mathcal{P}(X)} I(P, W)$ where $I(P, W) = I(X : Y)$ is evaluated for $P_{XY}(x, y) = P(x) W(y \mid x)$. Using the symmetry of the channel, prove that the input distribution achieving $I(W)$ can be chosen of the following form:

$$P_X(x) = \begin{cases} \frac{1-p}{2} & \text{if } x \neq 1 \\ p & \text{if } x = 1 \end{cases} \tag{2}$$

   for some $p \in [0, 1]$ still to be determined.

   Hint: You may use (without proof) the concavity of $I(P, W)$ in P and the fact that $I(P, W)$ does not change when we apply a permutation from Part 2 to the input pmf P.

   Solution: Take any $P_x = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$, then

$$I(P_x, W) = I(\tilde{P}, W) \tag{3}$$

where $\tilde{P} = \begin{bmatrix} p_3 \\ p_2 \\ p_1 \end{bmatrix}$ by symmetry. Thus,

$$\begin{aligned} I(P_x, W) &= \frac{1}{2}(I(P_x, W) + I(\tilde{P}, W)) \\ &\leq I(\frac{1}{2}P + \frac{1}{2}\tilde{P}, W) = I(\bar{P}, W) \end{aligned} \tag{4}$$

where

$$\bar{P} = \begin{bmatrix} \frac{1}{2}(p_1 + p_3) \\ p_2 \\ \frac{1}{2}(p_1 + p_3) \end{bmatrix}. \tag{5}$$

Thus $\bar{p}$ is always at least as good as P_X and has the desired structure.

4. (4 points) Using the simplification established in Part 3, compute $I(P, W)$ as a function of p. Determine $I(W)$ and the value p^* that achieves the maximum.

   Solution:

$$P_Y = \begin{bmatrix} \frac{1-p}{2}(1-\varepsilon) + p\frac{1-\varepsilon}{2} \\ \varepsilon \\ \frac{1-p}{2}(1-\epsilon) + p\frac{1-\varepsilon}{2} \end{bmatrix} = \begin{bmatrix} \frac{1-\varepsilon}{2} \\ \varepsilon \\ \frac{1-\varepsilon}{2} \end{bmatrix} \tag{6}$$

$$\begin{aligned} I(P, W) &= H(Y) - H(Y|X) && (7) \\ &= H(Y) - 2\frac{1-p}{2}h(\varepsilon) - pH(Y) && (8) \\ &= (1-p)(H(Y) - h(\varepsilon)) && (9) \\ &= (1-p)(2\frac{1-\varepsilon}{2}\log\frac{2}{1-\varepsilon} - (1-\varepsilon)\log\frac{1}{1-\varepsilon}) && (10) \\ &= (1-p)(1-\varepsilon) && (11) \end{aligned}$$

   This is maximized for $p^* = 0$, when $I(W) = 1 - \varepsilon$.

5. (1 point) If we fix $p = p^*$ the channel effectively reduces to a well-known channel—which one?

   Solution: BEC (Binary Erasure Channel)

3 Question 3: Entropy inequalities (Total: 6 points)

Consider three discrete random variables X_1, X_2 and X_3

1. (2 points) Show that

$$H(X_1, X_2) + H(X_2, X_3) \geq H(X_2) + H(X_1, X_2, X_3) \tag{12}$$

   Solution:

$$H(X_1|X_2) \geq H(X_1|X_2, X_3) \tag{13}$$

   Then

$$H(X_1, X_2) + H(X_2, X_3) \geq H(X_2) + H(X_1, X_2, X_3) \tag{14}$$

2. (4 points) Show that

$$H(X_1, X_2, X_3) \leq \frac{1}{2}(H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3)) \tag{15}$$

Solution:

$$H(X_1, X_2, X_3) \leq H(X_1, X_3) + H(X_2|X_3) \tag{16}$$

Using (1), we have

$$\begin{aligned} H(X_1, X_2, X_3) &\leq \frac{1}{2}(H(X_1, X_2) + H(X_1, X_3) + H(X_2|X_3) + H(X_3|X_2)) \\ &\leq \frac{1}{2}(H(X_1, X_2) + H(X_1, X_3) + H(X_2) + H(X_3|X_2)) \\ &= \frac{1}{2}(H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3)) \end{aligned} \tag{17}$$

4 Question 4: Group testing (Total: 11 points)

You are presented with nine people and a guarantee that exactly one of them has a rare disease, and you even know the probability with which each person has the disease, given by the vector

$$(0.2, 0.1, 0.05, 0.15, 0.25, 0.1, 0.05, 0.04, 0.06) \tag{18}$$

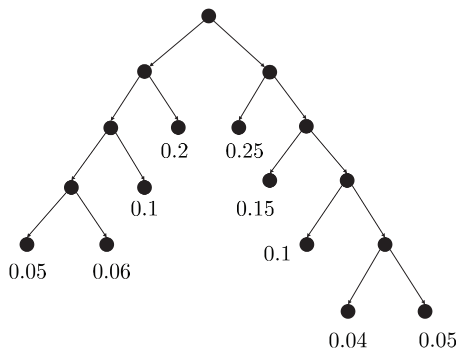
You are tasked to find the person carrying the disease. An expensive test is available that allows you test if any group of people carries the disease or not with certainty.

1. (2 point) Devise a protocol (a sequence of tests, where the following test is always chosen as a function of the previous test results) to find the infected person that requires at most 4 tests.

   Solutions: Use binary search (see below).

2. (4 points) Devise a protocol that minimizes the expected number of tests.

   Solutions:



   Starting from the root, test at each node test the group on the left branch vs. the group on the right branch, until a leaf is reached.

3. (1 point) What is the maximal number of tests required (in the worst case) for the latter protocol?

   Solutions: 5

   Now consider a more realistic scenario where an imperfect test yields false positives or false negatives with finite probability.

4. (2 points) Devise a scheme that requires 7 tests and and allows you to identify the infected person with certainty even if at most one of the seven tests yields the wrong outcome.

   Solutions: Associate a 4-bits string $X_1X_2X_3X_4$ to each patient. The 4 tests in part (1) test for each of these bits separately, e.g. we test patients with $X_1 = 0$ against those with $X_1 = 1$ to find the value of X_1.

   Now we add 3 tests according to the Hamming code:

$$X_2 \oplus X_3 \oplus X_4 \oplus = 0 \text{ vs. } X_2 \oplus X_3 \oplus X_4 \oplus = 1$$
$$X_1 \oplus X_3 \oplus X_4 \oplus = 0 \text{ vs. } X_1 \oplus X_3 \oplus X_4 \oplus = 1$$
$$X_1 \oplus X_2 \oplus X_4 \oplus = 0 \text{ vs. } X_1 \oplus X_2 \oplus X_4 \oplus = 1$$

Information comes to light that puts into question the model that allowed you to arrive at the probabilities in Eq. (5) in the first hand. An alternative model would yield the probabilities

$$(0.2, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1) \tag{19}$$

5. (2 points) Assuming both models are equally likely, what is the probability of correctly distinguishing between them with a single test? You can assume a perfect (noiseless) test here.

   Solutions: Compute the trace distance, $\Delta = 0.05 + 0.15 = 0.2$. Then,

$$P_{\text{guess}} = \frac{1}{2}(1 + \delta) = 0.6\,. \tag{20}$$

5 Question 5: Alternative channel coding converse (Total: 10 points)

Show the weak converse to the channel coding theorem for a DMC with state. The channel state $s \in \mathcal{S}$ is i.i.d. according to some pmf P_S and known to both the encoder and decoder. The DMC is then determined by a conditional pmf $W_{Y|XS}$. For a fixed blocklength $n \in \mathbb{N}$, the setup is depicted here:

Consider an arbitrary code for sending $M_n = \lceil 2^{nR} \rceil$ messages over W^n. Such a code is given by an encoder $e_n : [M_n] \times \mathcal{S}^n \to \mathcal{X}^n$ and decoder $d_n : \mathcal{Y}^n \times \mathcal{S}^n \to [M_n]$. As usual we consider the case where $M \in [M_n]$ follows a uniform distribution. Let $\varepsilon_n = \Pr[M \neq \hat{M}]$. We want to show that for any rate $R > I(W)$, where

$$I(W) := \max_{P_{X|S}} I(X : Y \mid S) \tag{21}$$

we must have $\lim_{n\to\infty} \varepsilon_n > 0$. Construct a proof of the converse by following the steps outlined below:

1. (1 point) Show that $H(M \mid Y^n S^n) \leq 1 + \varepsilon_n \cdot nR$ for any code with blocklength n, rate R, and probability of error ε_n.

   Solution:

   $H(M|\hat{M}) \leq 1 + \varepsilon_n nR$ by Fano's inequality.

   $H(M|Y^n S^n) \leq H(M|\hat{M})$ by DPI for the decoder.

2. (2 points) Show that $nR \leq H(M \mid Y^n S^n) + I(M : Y^n \mid S^n)$.

   Solution:

$$\begin{aligned} I(M : Y^n|S^n) &= H(M|S^n) - H(M|Y^n S^n) \\ &= H(M) - H(M|Y^n S^n) \\ &\geq nR - H(M|Y^n S^n). \end{aligned} \tag{22}$$

   Thus $nR \leq I(M : Y^n|S^n) - H(M|Y^n S^n)$.

3. (3 points) Show that $I(M : Y^n \mid S^n) \leq I(W^n)$.

   Solution: Since e_n is injective, we can define its inverse $e_n^{-1} : \mathcal{X} \times \mathcal{S}^n \to [M_n]$ for each $s^n \in \mathcal{S}^n$. By the DPI, we have

$$\begin{aligned} I(X^n : Y^n|S^n) &= H(Y^n|S^n) - H(Y^n|X^n S^n) \\ &\geq H(Y^n|S^n) - H(Y^n|M S^n) \\ &= I(M : Y^n|S^n). \end{aligned} \tag{23}$$

   We further relax the bound by optimizing over $P_{X^n|S^n}$.

4. (3 points) Show that $I(\cdot)$ is additive, i.e., $I\left(W^n\right) = nI(W)$. You may use, without proof, that for any random variables X_1, X_2, Y_1, Y_2, S_1 and S_2 the following relation holds:

$$I(X_1, X_2 : Y_1, Y_2 \mid S_1, S_2) = I(X_1 : Y_1, Y_2 \mid S_1, S_2) + I(X_2 : Y_1, Y_2 \mid X_1, S_1, S_2). \tag{24}$$

Simplify this further for the situation at hand and show the equality for $n = 2$ first.

Solution:

$$\begin{aligned} I(W^2) &= I(X_1 : Y_1, Y_2 \mid S_1, S_2) + I(X_2 : Y_1, Y_2 \mid X_1, S_1, S_2) \qquad (25) \\ &= I(X_1 : Y_1 \mid S_1) + I(X_2 : Y_2 \mid X_1, S_2) \\ &\leq \max_{P_{X1}} I(X_1 : Y_1 \mid S_1) + \max_{P_{X2}} I(X_2 : Y_2 \mid S_2) \\ &= 2I(W), \qquad (26) \end{aligned}$$

where in Eq. (25) we chose the maximizing input distribution $P_{X_1 X_2 | S_1 S_2}$. We can chain this argument to show $I(W^n) \leq nI(W)$.

5. (1 points) Use the above facts to complete the proof of the weak converse.

Solution: Summarizing the above inequalities, we have

$$nR \leq 1 + \varepsilon_n nR + nI(W)\,. \tag{27}$$

Then, $\varepsilon_n \geq \frac{R - I(W)}{R} - \frac{1}{nR}$, and thus

$$\lim_{n \to \infty} \varepsilon_n \geq \frac{R - I(W)}{R} > 0 \quad \text{by assumption on } R\,. \tag{28}$$