Choose any three (EE5139) or two (EE6139) of the below exercises. The remainder of the exercises will serve as a good preparation for the final exam.

Exercise 10.1 Binary Huffman Code

Suppose there is a source X with alphabet $\{1, 2, \ldots, n\}$ and each symbol has an equal probability. We use the Huffman algorithm to generate a binary code. Calculate the code length of each symbol and the average code length when $n = 2^L$ and $n = 2^L + 1$, where L is a positive integer.

Exercise 10.2 A Testing Problem

Let there be 9 visually identical balls. One of the balls is heavier than the rest, and the remaining 8 balls are of equal weight. The task is to find out the heavier ball with a balance. We let X be a random variable denoting the number of the heaviest ball, *i.e.*, $X = i$ if and only if the i-th ball is heaviest, Assume equal probability for each ball being the heavier one, *i.e.*, $p_X(i) = 1/9$ or each $i = 1, \ldots, 9$.

a.) Suppose, for the first weighing, you decide to weigh 3 balls against another 3 balls. Denote the outcome by random variable Y. What is the resultant conditional entropy $H(X|Y)$ (*i.e.*, the average uncertainty of X after observing Y)?

b.) Calculate the mutual information $I(X : Y) = H(X) - H(X|Y)$ and what is its interpretation in this context?

c.) Calculate $H(Y)$ and $H(Y|X)$ directly from their distributions. Using these term, recalculate the mutual information $I(X : Y) = H(Y) - H(Y|X)$. (Of course, you should end up with same result. Otherwise, check your answer.)

d.) Suppose, instead, you decide to weigh 2 balls against another 2 balls for the first weighing, and denote the outcome by random variable Y. What is the resultant conditional entropy $H(X|Y)$ and hence, calculate $I(X : Y)$.

e.) Suppose, instead, you decide to weigh only one ball against another ball for the first weighing, and denote the outcome by random variable Y. What is the resultant conditional entropy $H(X|Y)$ and hence, calculate $I(X : Y)$.

f.) Based on above steps, what is the optimal weighing strategy for the first weighing in order to maximize mutual information $I(X : Y)$?

g.) Suppose now you have narrowed down the candidate for the heavier ball to just 3 balls, following the argument above. What should be the optimal weighing pattern be for the second weighing?

Exercise 10.3 Another Testing Problem

Given 8 bottles of water, among which one of them is sugared. The likelihood of the x-th bottle to be sugared (prior to any testing) is described by the probability p_X. Suppose, for each $i \in \{1, \ldots, 8\}$, $p_X(i) = p_i$ where $(p_1, p_2, \ldots, p_8) = (\frac{7}{32}, \frac{6}{32}, \frac{6}{32}, \frac{5}{32}, \frac{3}{32}, \frac{3}{32}, \frac{1}{32}, \frac{1}{32})$. The task is to taste the water to figure out which bottle is sugared. (Obviously, you need to taste at most 7 times.)

a.) Without mixing the water, what is the minimum expected number of tastes? In what order for tasting of the bottles, is this minimum expectation attained?

b.) Allowing mixing, what is the strategy that minimizes the expected number of drinks that need to be tasted? What is the expected and maximal number of taste for this strategy? Which mixture should be tasted first? **Hint:** You may think about this problem as a coding problem.

Exercise 10.4 Joint Typicality

Let (X, Y) be a pair of random variables with X and Y distributed on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let (X^n, Y^n) be n i.i.d. copies of (X, Y). We define the set of joint typical sequences in $(\mathcal{X} \times \mathcal{Y})^n$ as (note that this is slightly different from Exercise 9.3)

$$\mathcal{A}_\epsilon^{(n)}(XY) := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| \frac{1}{n} \log \frac{1}{p_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y})} - H(XY) \right| \leq \epsilon \right\}.$$

A sequence $(\boldsymbol{x}, \boldsymbol{y})$ is said to be jointly typical if and only if it is in $\mathcal{A}_\epsilon^{(n)}(XY)$.

a.) Show that, for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}(XY)$, it holds that

$$2^{-n(H(X)+H(Y|X)+\epsilon)} \leq p_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y}) \leq 2^{-n(H(X)+H(Y|X)-\epsilon)}.$$

b.) For a typical $\boldsymbol{x} \in \mathcal{A}_\epsilon^{(n)}(X)$, a sequence $\boldsymbol{y} \in \mathcal{Y}^n$ is said to be *relatively typical* to $\boldsymbol{x}$ if and only if $(\boldsymbol{x}, \boldsymbol{y})$ is jointly typical. Given $\boldsymbol{y}$ to be relatively typical to $\boldsymbol{x}$, show that

$$2^{-n(H(Y|X)+2\epsilon)} \leq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x}) \leq 2^{-n(H(Y|X)-2\epsilon)}.$$

c.) Prove that the probability that Y^n is relatively typical to X^n tends to 1 as $n \to \infty$, namely,

$$p_{X^n,Y^n}(\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \boldsymbol{x} \in \mathcal{A}_\epsilon^{(n)}(X), (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}(XY)\}) \to 1$$

as $n \to \infty$.

Exercise 10.5 Information Spectrum Analysis

Let $p_{Y|X}$ be a channel from input set $\mathcal{X}$ to output set $\mathcal{Y}$. $\mathcal{X}$ and $\mathcal{Y}$ need not be discrete.

a.) Suppose we use the channel *once*. Show that there exist a code with M codewords with average probability of error ϵ satisfying

$$\epsilon \leq \Pr\left[\log \frac{p_{Y|X}(Y|X)}{p_Y(Y)} \leq \log M + \gamma\right] + 2^{-\gamma}.$$

for any choice of $\gamma > 0$ and any input distribution P_X where $P_Y(y) = \sum_x P_{Y|X}(y|x)P_X(x)$. **Hint:** Generate codewords independently according to P_X. Instead of using typical set for decoding, use $\hat{m} \in \{1, \ldots, M\}$ as the transmitted message if it is the unique one satisfying

$$\log \frac{P_{Y|X}(y|x(\hat{m}))}{P_Y(y)} \geq \log M + \gamma.$$

If there is no unique $\hat{m}$ satisfying the above condition, declare an error. The analysis to arrive at the one-shot (finite blocklength) bound above is very similar to typical set decoding. A stronger version of this bound (for maximum error) was shown by Feinstein.

b.) Based on part (a), prove the channel coding theorem for finite $\mathcal{X}, \mathcal{Y}$ and memoryless channels. **Hint:** Set P_{X^n} above to be the n-fold product distribution corresponding to a capacity-achieving input distribution $P_X \in \arg\max_{P_X} I(X;Y)$. Set γ above to be $n\gamma'$ for some $\gamma' > 0$. Set $\log M = n(C - 2\gamma')$. Apply the law of large numbers to the first term to see that there exists a sequence of $(n, 2^{n(C-2\gamma')})$-codes with vanishing average error probabilities.

c.) Again consider the setup in (b). Let

$$V := \mathrm{Var}\left(\log \frac{P^\star_{Y|X}(Y|X)}{P^\star_Y(Y)}\right)$$

be evaluated at a capacity-achieving input distribution. Based on part (a), show that, by the central limit theorem, there exists a sequence of codes indexed by blocklength n, with sizes M_n satisfying

$$\log M_n = nC + \sqrt{nV}\Phi^{-1}(\epsilon) + o(\sqrt{n})$$

such that the average error probability is no larger than $\epsilon + o(1)$.

Exercise 10.6 Channel Coding and Capacity

Consider that there is a binary symmetric channel (BSC) with crossover probability ϵ. We use a coding scheme on this channel that encodes messages a_1 and a_2 as 000 and 111, respectively. On the other hand, we use a decoding scheme that relies on the majority principle, *i.e.*, $001 \mapsto 0$ and $011 \mapsto 1$.

a.) Merge the above encoder, the BSC, and the above decoder into a new channel. Describe the input/output alphabets and the channel rule (*i.e.*, the conditional probability of outputs given inputs) of this new channel.

b.) What is the capacity of the new channel if the crossover probability of the original BSC $\epsilon = 0.1$?

c.) What is the capacity of the original BSC channel with $\epsilon = 0.1$?

d.) Suppose that we have 4 input messages (each to be sent with equal probability) instead. Find the best code of length 3 that minimizes the average decoding error.

e.) Consider a binary erasure channel (BEC) with erasure probability $\epsilon = 0.1$. Suppose we encode the messages a_1 and a_2 as 000 and 111, respectively. On the decoder side, if we receive $\perp\perp\perp$, we will randomly assign it to 000 or 111 with equal probability; otherwise, we decode according to the survived symbol, *i.e.*, $0 \perp\perp \mapsto a_1$ and $\perp\perp 1 \mapsto a_2$. What is the average decoding error in this case?

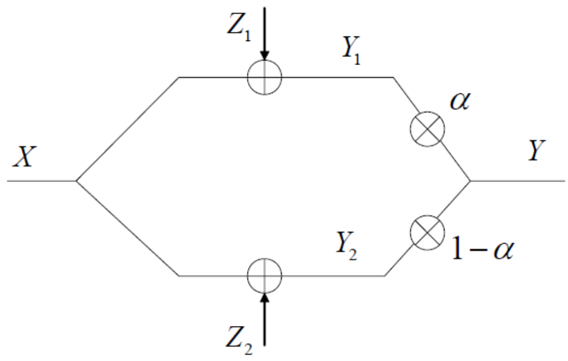
f.) Merge the encoder (from step e.)), the BEC, and the decoder (from step e.)) into another new channel. Show, without computation, the capacity of the new channel is no larger than three times the capacity of the BEC.

Exercise 10.7 A Gaussian channel

Consider a Gaussian channel shown below, in which the transmitted signal X with $\mathbb{E}[X^2] = P$ is received by two antennas with $Y_1 = X + Z_1$ and $Y_2 = X + Z_2$ where Z_1 and Z_2 are independent with $\mathbb{E}[Z_i^2] = \sigma_i^2$ ($\sigma_1^2 < \sigma_2^2$). The signals at the two antennas are combined as

$$Y = \alpha Y_1 + (1-\alpha) Y_2$$

before decoding, where $0 \le \alpha \le 1$.



a.) Find the capacity of this channel for a given α. Please provide the units of capacity.

b.) Using your result at the previous point, find the optimal α that maximizes the capacity and write down the corresponding maximum capacity in terms of P, σ_1^2 and σ_2^2.