EECS229: Information Theory
In Class Midterm
Write name and ID number on each page
Turn in your notes along with exam
100 points is a good score.

Problem 1.1 *Warm Up*

X_i are i.i.d. Bernoulli random variables with $P(X_i = 1) = \frac{3}{4}$.
The distortion measure is:

$$d(x,\hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ +\infty & \text{if } x \neq \hat{x} \end{cases}$$

For block-length N, a fixed rate R source code is an encoder/decoder pair whose encoder maps $\{0,1\}^N$ into $\{0,1\}^{RN}$ and the decoder maps the result back into $\{0,1\}^N$.

For block-length N, a variable rate source code has the encoder map $\{0,1\}^N$ into a prefix-free set of binary codewords of length l_i each and the decoder maps the binary codewords back into $\{0,1\}^N$. The rate R of the code is considered to be the ensemble average of the codeword length per input letter:

$$R = \frac{E[L]}{N}$$

a. *10pts Draw the subset of the R, D plane that can be reached in the limit of large N using fixed rate codes. Argue why this is the optimal region and how you would achieve points within it.*

b. 10pts Draw the subset of the R, D plane that can be reached in the limit of large N using variable rate codes. Argue why this is the optimal region and how you would achieve points within it.

c. *10pts Is it possible to achieve small expected distortion by appropriate coding over a binary symmetric channel with crossover probability* $\epsilon = \frac{1}{100}$*? How or why not?*

Problem 1.2 *Constrained Capacity*

You are given a discrete memoryless channel with input alphabet $\mathcal{X} = \{0, 1, 2, \ldots, J\}$, *finite output alphabet* $\mathcal{Y} = \{0, 1, 2, \ldots, K\}$, *and transition probabilities* $p(Y = k|X = j) = p_{j,k}$.

You are also given a cost function $+\infty > f(j) > 0$ *which represents the cost of using input* j *on the channel and an average cost constraint* W *on your use of the channel. By this, we mean that for a block code containing* 2^{NR} *codewords* $\vec{x}(m)$, *each of length* N *that:*

$$\frac{1}{2^{NR}} \sum_{m=1}^{2^{NR}} \frac{1}{N} \sum_{i=1}^{N} f(x_i(m)) \leq W$$

This problem asks you to show that the capacity of the channel is

$$C = \sup_{P(X)|E[f(X)] \leq W} I(X;Y)$$

a. *35pts Show that for any desired probability of error* ϵ, *and rate* $R < C$, *that there exists a rate* R *block-code that satisfies the cost constraint and has probability of error less than* ϵ.

(HINT: Setup the appropriate random codebook and use jointly typical decoding where the notion of joint typicality includes the cost constraint...)

b. 35pts Show the converse — that if we have a sequence of block-codes at rate R that have block error probability $P_e \to 0$ while each also satisfies the average cost constraint, then the rate $R \leq C$.

(HINT: Use Fano's inequality, the data processing inequality, and then look at the mutual information on each channel use. Use the convexity of mutual information to complete the story.)

Problem 1.3 *Conditional Compression*

The source process is i.i.d. and generates a pair of discrete random variables (X_i, Y_i) *at every given time.* X_i *and* Y_i *are* **not** *independent of each other. The twist in this problem is that* Y_i *is known perfectly at the decoder as well as the encoder.*

So the encoder has access to both (X_i, Y_i) *while the decoder only has access to* Y_i.

In this context, a rate R *block code with blocklength* N *consists of an encoder that maps* (X_1^N, Y_1^N) *into* RN *bits* B_1^{RN}, *together with a decoder that maps* (B_1^{RN}, Y_1^N) *into* $\hat{X}_1^N$.

a. *35pts Suppose that we want to achieve* ϵ*-lossless compression using a block code. Give the tightest possible bound you can on the required rate* R *and give a scheme that gets arbitrarily close to it in the limit of large* N.

   (HINT: Think about jointly typical sequences involving both Y_i and X_i.)

b. *35pts* *Continue with ϵ-lossless compression as in part (a). Give the best bound on rate that you can and show that it is not possible for any block code to do substantially better than your bound.*

(HINT: This is the converse to part a. You can use the jointly typical set again.)

c. 40pts (harder) Suppose you wanted to do lossy compression in this setup and you have an additive distortion measure d. What is the analogue to $R(D)$ and how would you approach it? Show that you can do no better.