1. For each of the following, explain the difference between the two quantities listed. Do not just give the definitions, but explain, in words, **why** they represent different concepts.

   (a) (**5 pts.**) $H(X, Y)$ and $I(X; Y)$

   (b) (**5 pts.**) $H(X)$ and $h(X)$

   (c) (**6 pts.**) C and $C^{(I)}$

2. Let $q_0(x)$ and $q_1(x)$ be two different distributions on alphabet $\mathcal{X}$, and let $\lambda \in [0, 1]$. Consider random variables X, Z where $Z \sim \text{Bern}(\lambda)$ and

$$p(x|z) = q_z(x) \text{ for } z = 0, 1.$$

(a) (**8 pts.**) Write $H(X|Z)$ and $H(X)$ in terms of q_0, q_1, and λ.

(b) (**8 pts.**) Use the fact that conditioning reduces entropy to prove that the entropy function is concave.

3. (**14 pts.**) Consider the channel with discrete input $X \in \{1, 2\}$, and real-valued output Y, where Y is distributed uniformly on the interval $[0, X]$. That is, if $X = 1$, then Y is uniform on $[0, 1]$, and if $X = 2$, then Y is uniform on $[0, 2]$.

   (a) (**7 pts.**) Let $q = \Pr(X = 2)$. Find the mutual information $I(X; Y)$ in terms of q.

   (b) (**7 pts.**) Maximize the mutual information over q to find the capacity of this channel.

4. Each of the following gives a rate-distortion problem. For each one, find the extreme points of the rate-distortion function. That is, find (i) the smallest achievable distortion with zero rate, and (ii) the smallest achievable rate for zero distortion. You do **not** need to compute the entire rate-distortion function.

   (a) (**6 pts.**) Let X be a discrete random variable with probabilities $(p_1, p_2, \ldots, p_k)$ where $p_1 \geq p_2 \geq \cdots \geq p_k$. The distortion function is the Hamming distortion.

   (b) (**5 pts.**) Let X be uniform on the set $\mathcal{X} = \{1, 2, 3, 4\}$. The reconstruction alphabet is also $\hat{\mathcal{X}} = \{1, 2, 3, 4\}$, and the distortion function is

   $$d(x, \hat{x}) = \begin{cases} 0, & x - \hat{x} \text{ is even} \\ 1, & x - \hat{x} \text{ is odd.} \end{cases}$$

   (c) (**5 pts.**) Let X be a continuous random variable uniformly distributed on the interval $[-1, 1]$. The distortion function is the squared error distortion $d(x, \hat{x}) = (x - \hat{x})^2$.

5. Consider 3 parallel Gaussian channels with noise variances $N_1 = 5$, $N_2 = 8$, $N_3 = 10$, and total power P. Calculate the capacity when:

   (a) (**5 pts.**) $P = 2$

   (b) (**5 pts.**) $P = 5$

   (c) (**6 pts.**) $P = 13$

6. (**14 pts.**)

(a) (**7 pts.**) Let P be a distribution on the ternary alphabet $\mathcal{X} = \{1, 2, 3\}$ with $P(1) \geq \frac{2}{3}$. Show that $H(P) \leq H(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$.

(b) (**7 pts.**) Let A be the set of sequences in $\mathcal{X}^n$ where at least 2/3 of the elements are 1s. Use part (a) and the method of types to show that

$$|A| \leq (n+1)^3 \, 2^{nH(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})}.$$

7. Consider the hypothesis test between two i.i.d. distributions

$$H_0 : X^n \sim P_0^n$$
$$H_1 : X^n \sim P_1^n$$

where P_0 and P_1 are distributions on the alphabet $\{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$ given by

$$P_0(x) = \begin{cases} 1/3, & x = \mathsf{a} \\ 1/3, & x = \mathsf{b} \\ 1/3, & x = \mathsf{c} \end{cases} \qquad P_1(x) = \begin{cases} 2/3, & x = \mathsf{a} \\ 1/3, & x = \mathsf{b} \\ 0, & x = \mathsf{c} \end{cases}$$

The following test based on the type of x^n has been suggested to decide between H_0 and H_1:

$$g(x^n) = \begin{cases} 1, & P_{x^n}(\mathsf{a}) > 1/2 \\ 0, & P_{x^n}(\mathsf{a}) \leq 1/2. \end{cases}$$

Let $\alpha_n = P_1^n(g(X^n) = 0)$ and $\beta_n = P_0^n(g(X^n) = 1)$ be the two error probabilities for this test.

(a) (**6 pts.**) The Neyman-Pearson lemma states that all optimal tests are likelihood ratio tests. Is g a likelihood ratio test?

(b) (**5 pts.**) Find $\lim_{n\to\infty} -\frac{1}{n} \log \alpha_n$.

(c) (**5 pts.**) Find $\lim_{n\to\infty} -\frac{1}{n} \log \beta_n$.

8. Consider the discrete memoryless channel with input and output alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, and the following channel transition matrix (rows correspond to X, columns to Y):

$$\begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

(a) (**5 pts.**) Find the capacity of this channel.

(b) (**5 pts.**) Now we impose the following channel cost:

$$b(x) = \begin{cases} 0, & x = 0 \text{ or } x = 1 \\ 1, & x = 2. \end{cases}$$

Recall that the capacity-cost function is given by

$$C(B) = \max_{p(x):\mathbb{E}b(X) \leq B} I(X;Y).$$

Find $C(0)$ for this channel.

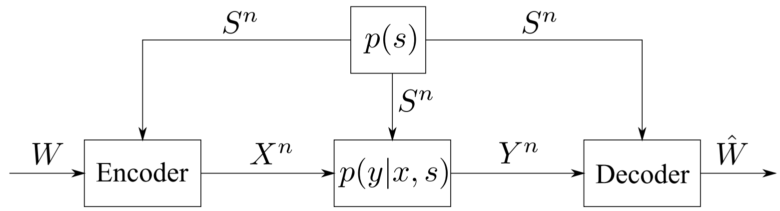
(c) (**6 pts.**) Find $C(B)$ for all $B \geq 0$.

9. Consider a channel with input X and output Y. Assume that the channel inputs are binary taking on the values 0 or 1. Let Z be a continuous random variable that is uniform on the interval $(-1, 1)$. Assuming that Z is independent of X.

   (a) What is the capacity of this channel if the input-output relationship is given by $Y = X + Z$.

   (b) What is the capacity of this channel if the input-output relationship is given by $Y = XZ$.

10. (**24 pts.**) Consider the variation on channel coding diagrammed below:



The variable $S \in \mathcal{S}$ is called the *channel state*: it determines which of several different channel transition probabilities occur (i.e. the channel given by $p(y|x, S=1)$ is different from $p(y|x, S=2)$). The state sequence S^n is drawn i.i.d. from $p(s)$ (independently from the choice of message), and then it is revealed to both the encoder and the decoder. For the ith channel use, the encoder chooses X_i, and then Y_i is drawn from $p(y|x,s)$ conditioned on X_i and S_i. In this problem, you will show that the capacity of this channel is

$$C = \max_{p(x|s)} I(X;Y|S).$$

In above expression, the conditional mutual information is with respect to the distribution $p(s)p(x|s)p(y|x,s)$, where $p(s)$ and $p(y|x,s)$ are given in the problem description, and $p(x|s)$ is chosen by the maximization.
This problem has 4 parts. Be sure to attempt them all.

(a) (**5 pts.**) For each $s \in \mathcal{S}$, let C_s be the usual channel capacity of the channel from X to Y with fixed $S = s$. That is

$$C_s = \max_{p(x)} I(X;Y|S=s).$$

Show that

$$\sum_{s \in \mathcal{S}} p(s) C_s = C$$

where C is defined above.

(b) (**5 pts.**) Prove that $\lim_{n\to\infty} \Pr\Big(N(s|S^n) \geq n(1-\epsilon)p(s) \text{ for all } s \in \mathcal{S}\Big) = 1.$

(c) (**9 pts.**) For each $s \in \mathcal{S}$, assume the existence of a standard channel code for channel $p(y|x, S=s)$ with blocklength $n(1-\epsilon)p(s)$, rate C_s, and probability of error ϵ. Using parts (a) and (b), show that for sufficiently large n, there exists a code with rate $(1-\epsilon)C$ and probability of error ϵ' where $\epsilon' \to 0$ as $\epsilon \to 0$.

(d) (**5 pts.**) Given a sequence of codes with rate R and probability of error $P_e^{(n)} \to 0$, prove the converse by justifying each of the following steps ($\mathcal{E}_n$ is a sequence decreasing to 0 as $n \to \infty$):

$$\begin{aligned}
nR &\overset{\text{(i)}}{=} I(W;Y^n|S^n) + H(W|Y^n,S^n) \\
&\overset{\text{(ii)}}{\leq} I(W;Y^n|S^n) + n\mathcal{E}_n \\
&\overset{\text{(iii)}}{=} \sum_{i=1}^{n} \Big[H(Y_i|S^n, Y_1, \ldots, Y_{i-1}) - H(Y_i|S^n, W, Y_1, \ldots, Y_{i-1})\Big] + n\mathcal{E}_n \\
&\overset{\text{(iv)}}{\leq} \sum_{i=1}^{n} \Big[H(Y_i|S_i) - H(Y_i|X_i, S_i)\Big] + n\mathcal{E}_n \\
&\overset{\text{(v)}}{\leq} nC + n\mathcal{E}_n.
\end{aligned}$$