1. Each of the following lists two quantities. Determine (i) whether $\leq$ or $\geq$ holds between them, and (ii) under what conditions they are equal. Don't forget to justify your answers. Assume X, Y are random variables, $f(X)$ is some function of X, and a_i for $i = 1, \ldots, n$ are real numbers.

   (a) $H(X,Y)$ and $H(X) + H(Y)$

   (b) $H(X,Y)$ and $H(X)$

   (c) $\max_{i \in \{1,\ldots,n\}} a_i$ and $\frac{1}{n}\sum_{i=1}^{n} a_i$

   (d) $I(X;Y|f(X))$ and $I(X;Y)$

2. This problem illustrates that any entropy can be written in terms of the binary entropy function. Recall that the binary entropy function is given by

$$H(p) = -p \log p - (1-p) \log(1-p).$$

Let $X \in \{1, 2, 3\}$ be a random variable with probabilities (p_1, p_2, p_3). Let

$$Y = \begin{cases} 1 & \text{if } X = 1 \\ 0 & \text{if } X = 2 \text{ or } X = 3. \end{cases}$$

(a) Prove that $H(X) = H(Y) + (p_2 + p_3)H(X|Y = 0)$. *Hint:* First show that $H(X) = H(X, Y)$.

(b) Given part (a), write $H(X)$ in terms of p_1, p_2, p_3 using only the binary entropy function.

(c) Now consider the random variable $Z \in \{1, 2, 3, 4\}$ with probabilities (p_1, p_2, p_3, p_4). Using a similar method as above, write $H(Z)$ using only the binary entropy function.

3. Consider a random variable X with alphabet $\mathcal{X}$ and distribution $p(x)$. Given real numbers a, b, define the set

$$\mathcal{S} = \{x \in \mathcal{X} : a \leq p(x) \leq b\}.$$

Prove the following:

$$\frac{\Pr\{X \in \mathcal{S}\}}{b} \leq |\mathcal{S}| \leq \frac{\Pr\{X \in \mathcal{S}\}}{a}.$$

4. Consider a distribution on four letters $\{A, B, C, D\}$ with probabilities 3/8, 5/16, 1/4, 1/16.

   (a) Find a Huffman code for this distribution. (Not just the code-lengths. Give a specific code.)

   (b) Find a Shannon code for this distribution. Recall that a Shannon code is a *prefix* code with code lengths given by $\lceil \log \frac{1}{p(x)} \rceil$.

5. The m-ary symmetric channel is a generalization of the binary symmetric channel. The input and output alphabets each consist of the integers from 1 to m; i.e.

$$\mathcal{X} = \mathcal{Y} = \{1, 2, \ldots, m\}.$$

The channel transition matrix is given by

$$p(y|x) = \begin{cases} 1-p & \text{if } y = x \\ \frac{p}{m-1} & \text{if } y \neq x. \end{cases}$$

(a) For any distribution on X, calculate $H(Y|X)$ for the m-ary symmetric channel.

(b) Find the capacity of the m-ary symmetric channel.

(c) Use Fano's inequality to find an upper bound on $H(Y|X)$ using the fact that $p = \Pr\{Y \neq X\}$. Compare this bound to the answer to part (a).

6. Consider the following source code on alphabet $\{a, b, c, d, e\}$ which is *not* a prefix code

$$
\begin{aligned}
a &\to 0\,0 \\
b &\to 0\,1 \\
c &\to 1\,1 \\
d &\to 0\,0\,1 \\
e &\to 0\,1\,1
\end{aligned}
$$

(a) Even though this is not a prefix code, it is still decodable. Decode the following bit string, which represents the concatenation of several codewords:

$$0\,0\,1\,1\,1\,0\,1\,0\,1\,1\,0\,0\,1\,1$$

(b) Find a prefix code that is equivalent to this code in that it has the same codeword lengths.

(c) Describe the advantage of a prefix code.

7. Consider the steps of the converse proof to the channel coding theorem shown below. Justify each of the steps (a)–(e).

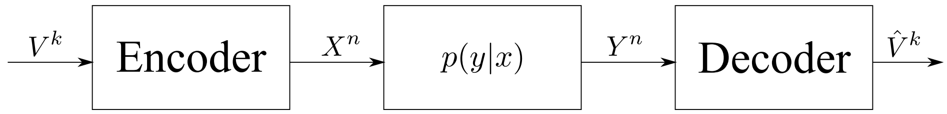
$$
\begin{aligned}
nR &\overset{(a)}{\leq} I(W;Y^n) + n\epsilon_n \\
&\overset{(b)}{\leq} I(X^n;Y^n) + n\epsilon_n \\
&\overset{(c)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) + n\epsilon_n \\
&\overset{(d)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) + n\epsilon_n \\
&\overset{(e)}{\leq} nC^{(I)} + n\epsilon_n
\end{aligned}
$$

where $\epsilon_n = \frac{1}{n} + P_e^{(n)} R$.

8. Consider the following channel. The input $X \in \{0, 1\}$ is sent through two independent binary erasure channels, each with erasure probability α. Let Y_a and Y_b be the two outputs. The output of the channel is the pair (Y_a, Y_b). Find the capacity of this channel.

9. Let $p(x)$ be the uniform distribution on the finite alphabet $\mathcal{X}$, and $A_\epsilon^{(n)}$ the typical set with respect to $p(x)$. Prove that every sequence $x^n \in \mathcal{X}^n$ is typical.

10. Consider the following joint source-channel coding problem in which the sequence V^k is to be transmitted across a channel. Note that the length of the source sequence is k and the number of channel uses is n.



Assume V^k is drawn from an i.i.d. distribution Bern(q) (i.e., 0 with probability $1-q$ and 1 with probability q), and assume the channel is a binary symmetric channel with crossover probability p.

(a) Under what circumstances can $\Pr\{\hat{V}^k \neq V^k\}$ be made arbitrarily small?

(b) Suppose $n/k = m$ is an integer, and the encoder uses a repetition strategy. That is, it chooses

$$X^n = (\underbrace{V_1, V_1, \ldots, V_1}_{m \text{ times}}, \underbrace{V_2, V_2, \ldots, V_2}_{m \text{ times}}, \ldots, \underbrace{V_k, V_k, \ldots, V_k}_{m \text{ times}}).$$

The decoder chooses $\hat{V}_i$ as whichever of 0 or 1 appeared more often among the m bits associated with V_i. For example, if $m = 3$ and $k = 2$, then the channel output string $Y^n = (0, 0, 1, 1, 0, 1)$ would be decoded to $\hat{V}^k = (0, 1)$. In the case that $m = 3$, find $\Pr\{\hat{V}^k \neq V^k\}$.

(c) Now consider an arbitrary but fixed m. Does the repetition strategy result in arbitrarily small probability of error by taking $k \to \infty$? What about for fixed k and $m \to \infty$?