ASSIGNMENT - 1

INFORMATION

& COMMUNICATION

SOURCE: Probability and Computing by Michael and Eli

1. **Exercise 1.19:** Give examples of events where $\Pr(A \mid B) < \Pr(A)$, $\Pr(A \mid B) = \Pr(A)$, and $\Pr(A \mid B) > \Pr(A)$.

2. **Exercise 1.21:** Give an example of three random events X, Y, Z for which any pair are independent but all three are not mutually independent.

3, **Exercise 1.22:** **(a)** Consider the set $\{1, \ldots, n\}$. We generate a subset X of this set as follows: a fair coin is flipped independently for each element of the set; if the coin lands heads then the element is added to X, and otherwise it is not. Argue that the resulting set X is equally likely to be any one of the 2^n possible subsets.

(b) Suppose that two sets X and Y are chosen independently and uniformly at random from all the 2^n subsets of $\{1, \ldots, n\}$. Determine $\Pr(X \subseteq Y)$ and $\Pr(X \cup Y = \{1, \ldots, n\})$. (*Hint:* Use the part (a) of this problem.)

Optional

Exercise 1.26: Tic-tac-toe always ends up in a tie if players play optimally. Instead, we may consider random variations of tic-tac-toe.

(a) First variation: Each of the nine squares is labeled either X or O according to an independent and uniform coin flip. If only one of the players has one (or more) winning tic-tac-toe combinations, that player wins. Otherwise, the game is a tie. Determine the probability that X wins. (You may want to use a computer program to help run through the configurations.)

(b) Second variation: X and O take turns, with the X player going first. On the X player's turn, an X is placed on a square chosen independently and uniformly at random from the squares that are still vacant; O plays similarly. The first player to have a winning tic-tac-toe combination wins the game, and a tie occurs if neither player achieves a winning combination. Find the probability that each player wins. (Again, you may want to write a program to help you.)

SOURCE: Elements of Information Theory by Thomas and Joy

4.

2.2 *Entropy of functions*. Let X be a random variable taking on a finite number of values. What is the (general) inequality relationship of $H(X)$ and $H(Y)$ if

(a) $Y = 2^X$?

(b) $Y = \cos X$?

5.

2.8 *Drawing with and without replacement*. An urn contains r red, w white, and b black balls. Which has higher entropy, drawing $k \geq 2$ balls from the urn with replacement or without replacement? Set it up and show why. (There is both a difficult way and a relatively simple way to do this.)

6.

2.9 *Metric*. A function $\rho(x, y)$ is a metric if for all x, y,

- $\rho(x, y) \geq 0$.
- $\rho(x, y) = \rho(y, x)$.
- $\rho(x, y) = 0$ if and only if $x = y$.
- $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

(a) Show that $\rho(X, Y) = H(X|Y) + H(Y|X)$ satisfies the first, second, and fourth properties above. If we say that $X = Y$ if there is a one-to-one function mapping from X to Y, the third property is also satisfied, and $\rho(X, Y)$ is a metric.

7.

2.14 *Entropy of a sum*. Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$.

(a) Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus, the addition of *independent* random variables adds uncertainty.

(b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) Under what conditions does $H(Z) = H(X) + H(Y)$?

8.

5.28 *Shannon code.* Consider the following method for generating a code for a random variable X that takes on m values $\{1, 2, \ldots, m\}$ with probabilities $p_1, p_2, \ldots, p_m$. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_m$. Define

$$F_i = \sum_{k=1}^{i-1} p_k, \tag{5.148}$$

the sum of the probabilities of all symbols less than i. Then the codeword for i is the number $F_i \in [0, 1]$ rounded off to l_i bits, where $l_i = \lceil \log \frac{1}{p_i} \rceil$.

(a) Show that the code constructed by this process is prefix-free and that the average length satisfies

$$H(X) \leq L < H(X) + 1. \tag{5.149}$$

(b) Construct the code for the probability distribution (0.5, 0.25, 0.125, 0.125).

Optional

5.32 *Bad wine.* One is given six bottles of wine. It is known that precisely one bottle has gone bad (tastes terrible). From inspection of the bottles it is determined that the probability p_i that the ith bottle is bad is given by $(p_1, p_2, \ldots, p_6) = (\frac{8}{23}, \frac{6}{23}, \frac{4}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23})$. Tasting will determine the bad wine. Suppose that you taste the wines one at a time. Choose the order of tasting to minimize the expected number of tastings required to determine the bad bottle. Remember, if the first five wines pass the test, you don't have to taste the last.

(a) What is the expected number of tastings required?

(b) Which bottle should be tasted first?

Now you get smart. For the first sample, you mix some of the wines in a fresh glass and sample the mixture. You proceed, mixing and tasting, stopping when the bad bottle has been determined.

(a) What is the minimum expected number of tastings required to determine the bad wine?

(b) What mixture should be tasted first?

9.

5.40 *Code rate.* Let X be a random variable with alphabet $\{1, 2, 3\}$ and distribution

$$X = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ 2 & \text{with probability } \frac{1}{4} \\ 3 & \text{with probability } \frac{1}{4}. \end{cases}$$

The data compression code for X assigns codewords

$$C(x) = \begin{cases} 0 & \text{if } x = 1 \\ 10 & \text{if } x = 2 \\ 11 & \text{if } x = 3. \end{cases}$$

Let $X_1, X_2, \ldots$ be independent, identically distributed according to this distribution and let $Z_1 Z_2 Z_3 \cdots = C(X_1)C(X_2)\cdots$ be the string of binary symbols resulting from concatenating the corresponding codewords. For example, 122 becomes 01010.

(a) Find the entropy rate $H(\mathcal{X})$ and the entropy rate $H(\mathcal{Z})$ in bits per symbol. Note that Z is not compressible further.

(b) Now let the code be

$$C(x) = \begin{cases} 00 & \text{if } x = 1 \\ 10 & \text{if } x = 2 \\ 01 & \text{if } x = 3 \end{cases}$$

and find the entropy rate $H(\mathcal{Z})$.

(c) Finally, let the code be

$$C(x) = \begin{cases} 00 & \text{if } x = 1 \\ 1 & \text{if } x = 2 \\ 01 & \text{if } x = 3 \end{cases}$$

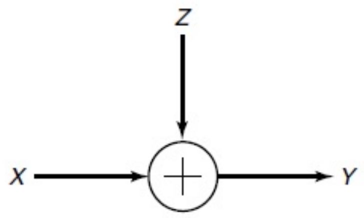
and find the entropy rate $H(\mathcal{Z})$.

10. **7.1** *Preprocessing the output.* One is given a communication channel with transition probabilities $p(y|x)$ and channel capacity $C = \max_{p(x)} I(X; Y)$. A helpful statistician preprocesses the output by forming $\tilde{Y} = g(Y)$. He claims that this will strictly improve the capacity.

(a) Show that he is wrong.

(b) Under what conditions does he not strictly decrease the capacity?

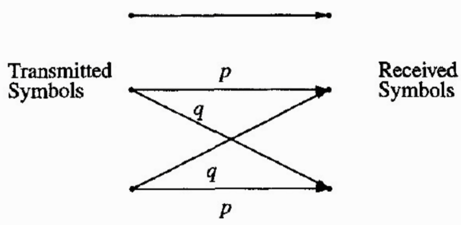
11. **7.2** *Additive noise channel.* Find the channel capacity of the following discrete memoryless channel:



where $\Pr\{Z = 0\} = \Pr\{Z = a\} = \frac{1}{2}$. The alphabet for x is $\mathbf{X} = \{0, 1\}$. Assume that Z is independent of X. Observe that the channel capacity depends on the value of a.

SOURCE: Error Correction Coding by Todd K. Moon

12. Consider the channel model shown here, which accepts three different symbols.



The first symbol is not affected by noise, while the second and third symbols have a probability p of not being corrupted, and a probability q of being changed into the other of the pair. Let $\alpha = -p \log p - q \log q$, and let P be the probability that the first symbol is chosen and let Q be the probability that either of the other two is chosen, so that $P + 2Q = 1$.

(a) Show that $H(X) = -P \log P - 2Q \log Q$.

(b) Show that $H(X|Y) = 2Q\alpha$.

(c) Choose the input distribution (i.e., choose P and Q) in such a way to maximize $I(X; Y) = H(X) - H(X|Y))$ subject to $P + 2Q = 1$. What is the capacity for this channel?

13. Consider a series of M BSCs, each with transition probability p, where the outputs of each BSC is connected to the inputs of the next in the series. Show that the resulting overall channel is a BSC and determine the crossover probability as a function of M. What happens as $M \to \infty$? *Hint*: To simplify, consider the difference of $(x + y)^n$ and $(x - y)^n$.