Homework1

I. Must done

1. (Prob. 2.1 of [1])

Coin flips. A fair coin is flipped until the first head occurs. Let X denote the number of flips required.

(a) Find the entropy $H(X)$ in bits. The following expressions may be useful:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \qquad \sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}.$$

(b) A random variable X is drawn according to this distribution. Find an "efficient" sequence of yes–no questions of the form,

"Is X contained in the set S?" Compare $H(X)$ to the expected number of questions required to determine X.

2. (Prob. 2.3 of [1])

Minimum entropy. What is the minimum value of $H(p_1, \ldots, p_n) = H(\mathbf{p})$ as $\mathbf{p}$ ranges over the set of n-dimensional probability vectors? Find all $\mathbf{p}$'s that achieve this minimum.

3. (Prob. 2.4 of [1])

Entropy of functions of a random variable. Let X be a discrete random variable. Show that the entropy of a function of X is less than or equal to the entropy of X by justifying the following steps:

$$H(X, g(X)) \overset{\text{(a)}}{=} H(X) + H(g(X) \mid X) \tag{2.168}$$

$$\overset{\text{(b)}}{=} H(X), \tag{2.169}$$

$$H(X, g(X)) \overset{\text{(c)}}{=} H(g(X)) + H(X \mid g(X)) \tag{2.170}$$

$$\overset{\text{(d)}}{\geq} H(g(X)). \tag{2.171}$$

Thus, $H(g(X)) \leq H(X)$.

4. (Prob. 2.5 of [1])

Zero conditional entropy. Show that if $H(Y|X) = 0$, then Y is a function of X [i.e., for all x with $p(x) > 0$, there is only one possible value of y with $p(x, y) > 0$].

5. (Prob. 2.10 of [1])

Entropy of a disjoint mixture. Let X_1 and X_2 be discrete random variables drawn according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over the respective alphabets $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m+1, \ldots, n\}$. Let

$$X = \begin{cases} X_1 & \text{with probability } \alpha, \\ X_2 & \text{with probability } 1-\alpha. \end{cases}$$

(a) Find $H(X)$ in terms of $H(X_1)$, $H(X_2)$, and α.

(b) Maximize over α to show that $2^{H(X)} \le 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

6. (Prob. 2.12 of [1])

Example of joint entropy. Let $p(x, y)$ be given by

$X \backslash Y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

Find:

(a) $H(X)$, $H(Y)$.

(b) $H(X \mid Y)$, $H(Y \mid X)$.

(c) $H(X, Y)$.

(d) $H(Y) - H(Y \mid X)$.

(e) $I(X; Y)$.

(f) Draw a Venn diagram for the quantities in parts (a) through (e).

7. (Prob. 2.16 of [1])

Bottleneck. Suppose that a (nonstationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus, $X_1 \to X_2 \to X_3$, that is,

$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$, for all $x_1 \in \{1, 2, \ldots, n\}$, $x_2 \in \{1, 2, \ldots, k\}$, $x_3 \in \{1, 2, \ldots, m\}$.

(a) Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \le \log k$.

(b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.

8. (Prob. 2.18 of [1])

World Series. The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, and BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, and $H(X|Y)$.

9. (Prob. 2.25 of [1])

Venn diagrams. There isn't really a notion of mutual information common to three random variables. Here is one attempt at a definition: Using Venn diagrams, we can see that the mutual information common to three random variables X, Y, and Z can be defined by

$$I(X; Y; Z) = I(X; Y) - I(X; Y|Z).$$

This quantity is symmetric in X, Y, and Z, despite the preceding asymmetric definition. Unfortunately, $I(X; Y; Z)$ is not necessarily nonnegative. Find X, Y, and Z such that $I(X; Y; Z) < 0$, and prove the following two identities:

(a) $I(X; Y; Z) = H(X, Y, Z) - H(X) - H(Y) - H(Z) + I(X; Y) + I(Y; Z) + I(Z; X)$.

(b) $I(X; Y; Z) = H(X, Y, Z) - H(X, Y) - H(Y, Z) - H(Z, X) + H(X) + H(Y) + H(Z)$.

The first identity can be understood using the Venn diagram analogy for entropy and mutual information. The second identity follows easily from the first.

10. (Prob. 2.29 of [1])

Inequalities. Let X, Y, and Z be joint random variables. Prove the following inequalities and find conditions for equality.

(a) $H(X, Y|Z) \geq H(X|Z)$.

(b) $I(X, Y; Z) \geq I(X; Z)$.

(c) $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$.

(d) $I(X; Z|Y) \geq I(Z; Y|X) - I(Z; Y) + I(X; Z)$.

11. (Prob. 2.32 of [1])

Fano. We are given the following joint distribution on (X, Y):

$X \backslash Y$	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $P_e = \Pr\{\hat{X}(Y) \neq X\}$.

(a) Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated P_e.

(b) Evaluate Fano's inequality for this problem and compare.

12. (Prob. 2.35 of [1])

Relative entropy is not symmetric.
Let the random variable X have three possible outcomes $\{a, b, c\}$. Consider two distributions on this random variable:

Symbol	$p(x)$	$q(x)$
a	$\frac{1}{2}$	$\frac{1}{3}$
b	$\frac{1}{4}$	$\frac{1}{3}$
c	$\frac{1}{4}$	$\frac{1}{3}$

Calculate $H(p)$, $H(q)$, $D(p||q)$, and $D(q||p)$. Verify that in this case, $D(p||q) \neq D(q||p)$.

II. Recommended

1. (Prob. 2.20 of [1])

Run-length coding. Let $X_1, X_2, \ldots, X_n$ be (possibly dependent) binary random variables. Suppose that one calculates the run lengths $\mathbf{R} = (R_1, R_2, \ldots)$ of this sequence (in order as they occur). For example, the sequence $\mathbf{X} = 0001100100$ yields run lengths $\mathbf{R} = (3, 2, 2, 1, 2)$. Compare $H(X_1, X_2, \ldots, X_n)$, $H(\mathbf{R})$, and $H(X_n, \mathbf{R})$. Show all equalities and inequalities, and bound all the differences.

2. (Prob. 2.21 of [1])

Markov's inequality for probabilities. Let $p(x)$ be a probability mass function. Prove, for all $d \geq 0$, that

$$\Pr\{p(X) \leq d\} \log \frac{1}{d} \leq H(X). \qquad (2.175)$$

3. (Prob. 2.24 of [1])

Average entropy. Let $H(p) = -p \log_2 p - (1-p)\log_2(1-p)$ be the binary entropy function.

(a) Evaluate $H(\frac{1}{4})$ using the fact that $\log_2 3 \approx 1.584$. (*Hint:* You may wish to consider an experiment with four equally likely outcomes, one of which is more interesting than the others.)

(b) Calculate the average entropy $H(p)$ when the probability p is chosen uniformly in the range $0 \le p \le 1$.

(c) (*Optional*) Calculate the average entropy $H(p_1, p_2, p_3)$, where (p_1, p_2, p_3) is a uniformly distributed probability vector. Generalize to dimension n.

4. (Prob. 2.27 of [1])

Grouping rule for entropy. Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ be a probability distribution on m elements (i.e., $p_i \ge 0$ and $\sum_{i=1}^m p_i = 1$).

Define a new distribution $\mathbf{q}$ on $m-1$ elements as $q_1 = p_1, q_2 = p_2, \ldots, q_{m-2} = p_{m-2}$, and $q_{m-1} = p_{m-1} + p_m$ [i.e., the distribution $\mathbf{q}$ is the same as $\mathbf{p}$ on $\{1, 2, \ldots, m-2\}$, and the probability of the last element in $\mathbf{q}$ is the sum of the last two probabilities of $\mathbf{p}$]. Show that

$$H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1}+p_m}, \frac{p_m}{p_{m-1}+p_m}\right). \tag{2.179}$$

5. (Prob. 2.30 of [1])

Maximum entropy. Find the probability mass function $p(x)$ that maximizes the entropy $H(X)$ of a nonnegative integer-valued random variable X subject to the constraint

$$EX = \sum_{n=0}^{\infty} np(n) = A$$

for a fixed value $A > 0$. Evaluate this maximum $H(X)$.

6. (Prob. 2.37 of [1])

Relative entropy. Let X, Y, Z be three random variables with a joint probability mass function $p(x, y, z)$. The relative entropy between the joint distribution and the product of the marginals is

$$D(p(x, y, z) || p(x)p(y)p(z)) = E\left[\log \frac{p(x, y, z)}{p(x)p(y)p(z)}\right]. \tag{2.180}$$

Expand this in terms of entropies. When is this quantity zero?

7. (Prob. 2.43 of [1])

Mutual information of heads and tails

(a) Consider a fair coin flip. What is the mutual information between the top and bottom sides of the coin?

(b) A six-sided fair die is rolled. What is the mutual information between the top side and the front face (the side most facing you)?

8. (Prob. 2.46 of [1])

Axiomatic definition of entropy (Difficult). If we assume certain axioms for our measure of information, we will be forced to use a logarithmic measure such as entropy. Shannon used this to justify his initial definition of entropy. In this book we rely more on the other properties of entropy rather than its axiomatic derivation to justify its use. The following problem is considerably more difficult than the other problems in this section.

If a sequence of symmetric functions $H_m(p_1, p_2, \ldots, p_m)$ satisfies the following properties:

- Normalization: $H_2\left(\frac{1}{2}, \frac{1}{2}\right) = 1$,
- Continuity: $H_2(p, 1-p)$ is a continuous function of p,
- Grouping: $H_m(p_1, p_2, \ldots, p_m) = H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$,

prove that H_m must be of the form

$$H_m(p_1, p_2, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i, \qquad m = 2, 3, \ldots. \tag{2.181}$$

There are various other axiomatic formulations which result in the same definition of entropy. See, for example, the book by Csiszár and Körner [149].

[1] Thomas M. Cover, “*Elements of Information Theory*”, Second Edition, 2006

[149] Csiszár and J. Körner, “*Information Theory: Coding Theorems for Discrete Memoryless Systems*”, Academic Press, New York, 1981.