Cover and Thomas, EIT, Chapter Homework Collection

HW#1

Problem 2.1 Coin flips

2.1 *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.

2.2 *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) $Y = \cos X$?

2.5 *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 $p(x, y) > 0$].

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.)

2.12 *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).

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 $H(Z) = H(X) + H(Y)$?

2.18 *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 $H(X|Y)$.

- Showing the convexity of $f(x) = e^x$ is easy. Use the Calculus: Take the derivatives twice and show that it's positive everywhere. Now, prove the convexity of $f(x)$ using the general convexity proving technique learned in this lecture.

- (Challenge; Optional) Consider arbitrary random variables X_1, X_2, and

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$$

  where the matrix elements $[a_{ij}]$ are arbitrary non zero constants and N_1 and N_2 are independent random variables. Let's denote $\mathbf{X} := \binom{X_1}{X_2}$.
  Prove or disprove $I(\mathbf{X}; Y_1, Y_2) \leq I(\mathbf{X}; Y_1) + I(\mathbf{X}; Y_2)$.

- $I(X_1, X_2; Y)$ and $I(\mathbf{X}; Y)$. Are they different?
- Recall the HW#0 problem on the joint distribution of U and V.
  (a) For the first case where $p_1 = 0.1$ and $p_2 = 0.2$, find the following measures: H(U), H(V), $H(U|\boldsymbol{\theta}_1)$, $H(V|\boldsymbol{\theta}_2)$, H(U|V), H(V|U), H(U, V), I(U; V), $I(U; \boldsymbol{\theta})$, $I(V; \boldsymbol{\theta})$.
  (b) Repeat for p_1=0.01 and $p_2 = 0.02$.
  (c) Note there is a notable change in I(U; V) between (a) and (b). Describe this change and make qualitative statements explaining the change. What would happen to I(U; V) when p_1 and p_2 approach zero? What would happen if they both approach 1/2.

Note Theta_1 and theta_2 here are supposed to be equal to e_1 and e_2 respectively.

HW Set#2

2.6 *Conditional mutual information vs. unconditional mutual information.* Give examples of joint random variables X, Y, and Z such that

(a) $I(X; Y \mid Z) < I(X; Y)$.

(b) $I(X; Y \mid Z) > I(X; Y)$.

2.23 *Conditional mutual information.* Consider a sequence of n binary random variables $X_1, X_2, \ldots, X_n$. Each sequence with an even number of 1's has probability $2^{-(n-1)}$, and each sequence with an odd number of 1's has probability 0. Find the mutual informations

$$I(X_1; X_2), \quad I(X_2; X_3|X_1), \ldots, \quad I(X_{n-1}; X_n|X_1, \ldots, X_{n-2}).$$

2.26 *Another proof of nonnegativity of relative entropy.* In view of the fundamental nature of the result $D(p||q) \geq 0$, we will give another proof.

(a) Show that $\ln x \leq x - 1$ for $0 < x < \infty$.

(b) Justify the following steps:

$$-D(p||q) = \sum_x p(x) \ln \frac{q(x)}{p(x)} \tag{2.176}$$

$$\leq \sum_x p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \tag{2.177}$$

$$\leq 0. \tag{2.178}$$

(c) What are the conditions for equality?

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

2.34 *Entropy of initial conditions*. Prove that $H(X_0|X_n)$ is nondecreasing with n for any Markov chain.

2.40 *Discrete entropies*. Let X and Y be two independent integer-valued random variables. Let X be uniformly distributed over $\{1, 2, \ldots, 8\}$, and let $\Pr\{Y = k\} = 2^{-k}$, $k = 1, 2, 3, \ldots$.

(a) Find $H(X)$.

(b) Find $H(Y)$.

(c) Find $H(X + Y, X - Y)$.

2.43 *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)?

2.48 *Sequence length*. How much information does the length of a sequence give about the content of a sequence? Suppose that we consider a Bernoulli $(\frac{1}{2})$ process $\{X_i\}$. Stop the process when the first 1 appears. Let N designate this stopping time. Thus, X^N is an element of the set of all finite-length binary sequences $\{0, 1\}^* = \{0, 1, 00, 01, 10, 11, 000, \ldots\}$.

(a) Find $I(N; X^N)$.

(b) Find $H(X^N|N)$.

(c) Find $H(X^N)$.

Let's now consider a different stopping time. For this part, again assume that $X_i \sim \text{Bernoulli}(\frac{1}{2})$ but stop at time $N = 6$, with probability $\frac{1}{3}$ and stop at time $N = 12$ with probability $\frac{2}{3}$. Let this stopping time be independent of the sequence $X_1 X_2 \cdots X_{12}$.

(d) Find $I(N; X^N)$.

(e) Find $H(X^N|N)$.

(f) Find $H(X^N)$.

HW#3

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

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

2.30 *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 $H(X)$.

2.32 *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.

3.1 *Markov's inequality and Chebyshev's inequality*

(a) (*Markov's inequality*) For any nonnegative random variable X and any $t > 0$, show that

$$\Pr\{X \geq t\} \leq \frac{EX}{t}. \tag{3.31}$$

Exhibit a random variable that achieves this inequality with equality.

(b) (*Chebyshev's inequality*) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y - \mu)^2$, show that

for any $\epsilon > 0$,

$$\Pr\{|Y - \mu| > \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}. \tag{3.32}$$

(c) (*Weak law of large numbers*) Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Let $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ be the sample mean. Show that

$$\Pr\left\{\left|\overline{Z}_n - \mu\right| > \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}. \tag{3.33}$$

Thus, $\Pr\left\{\left|\overline{Z}_n - \mu\right| > \epsilon\right\} \to 0$ as $n \to \infty$. This is known as the *weak law of large numbers*.

3.2 *AEP and mutual information*. Let (X_i, Y_i) be i.i.d. $\sim p(x, y)$. We form the log likelihood ratio of the hypothesis that X and Y are independent vs. the hypothesis that X and Y are dependent. What is the limit of

$$\frac{1}{n} \log \frac{p(X^n)p(Y^n)}{p(X^n, Y^n)}?$$

3.4 *AEP*. Let X_i be iid $\sim p(x)$, $x \in \{1, 2, \ldots, m\}$. Let $\mu = EX$ and $H = -\sum p(x) \log p(x)$. Let $A^n = \{x^n \in \mathcal{X}^n : |-\frac{1}{n} \log p(x^n) - H| \le \epsilon\}$. Let $B^n = \{x^n \in \mathcal{X}^n : |\frac{1}{n} \sum_{i=1}^n X_i - \mu| \le \epsilon\}$.

(a) Does $\Pr\{X^n \in A^n\} \longrightarrow 1$?

(b) Does $\Pr\{X^n \in A^n \cap B^n\} \longrightarrow 1$?

3.10 *Random box size*.
An n-dimensional rectangular box with sides $X_1, X_2, X_3, \ldots, X_n$ is to be constructed. The volume is $V_n = \prod_{i=1}^n X_i$. The edge length l of a n-cube with the same volume as the random box is $l = V_n^{1/n}$. Let $X_1, X_2, \ldots$ be i.i.d. uniform random variables over the unit interval [0, 1]. Find $\lim_{n\to\infty} V_n^{1/n}$ and compare to $(EV_n)^{\frac{1}{n}}$. Clearly, the expected edge length does not capture the idea of the volume of the box. The geometric mean, rather than the arithmetic mean, characterizes the behavior of products.

3.13 *Calculation of typical set*. To clarify the notion of a typical set $A_\epsilon^{(n)}$ and the smallest set of high probability $B_\delta^{(n)}$, we will calculate the set for a simple example. Consider a sequence of i.i.d. binary random variables, $X_1, X_2, \ldots, X_n$, where the probability that $X_i = 1$ is 0.6 (and therefore the probability that $X_i = 0$ is 0.4).

(a) Calculate $H(X)$.

(b) With $n = 25$ and $\epsilon = 0.1$, which sequences fall in the typical set $A_\epsilon^{(n)}$? What is the probability of the typical set? How many elements are there in the typical set? (This involves computation of a table of probabilities for sequences with k 1's, $0 \le k \le 25$, and finding those sequences that are in the typical set.)

(c) How many elements are there in the smallest set that has probability 0.9?

(d) How many elements are there in the intersection of the sets in parts (b) and (c)? What is the probability of this intersection?

Note the following table has incorrect entries. You should produce this table first on your own.

k	$\binom{n}{k}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$-\frac{1}{n} \log p(x^n)$
0	1	0.000000	1.321928
1	25	0.000000	1.298530
2	300	0.000000	1.275131
3	2300	0.000001	1.251733
4	12650	0.000007	1.228334
5	53130	0.000054	1.204936
6	177100	0.000227	1.181537
7	480700	0.001205	1.158139
8	1081575	0.003121	1.134740
9	2042975	0.013169	1.111342
10	3268760	0.021222	1.087943
11	4457400	0.077801	1.064545
12	5200300	0.075967	1.041146
13	5200300	0.267718	1.017748
14	4457400	0.146507	0.994349
15	3268760	0.575383	0.970951
16	2042975	0.151086	0.947552
17	1081575	0.846448	0.924154
18	480700	0.079986	0.900755
19	177100	0.970638	0.877357
20	53130	0.019891	0.853958
21	12650	0.997633	0.830560
22	2300	0.001937	0.807161
23	300	0.999950	0.783763
24	25	0.000047	0.760364
25	1	0.000003	0.736966

HW#4

4.4 *Second law of thermodynamics.* Let $X_1, X_2, X_3, \ldots$ be a stationary first-order Markov chain. In Section 4.4 it was shown that $H(X_n \mid X_1) \geq H(X_{n-1} \mid X_1)$ for $n = 2, 3, \ldots$. Thus, conditional uncertainty about the future grows with time. This is true although the unconditional uncertainty $H(X_n)$ remains constant. However, show by example that $H(X_n|X_1 = x_1)$ does not necessarily grow with n for every x_1.

4.7 *Entropy rates of Markov chains*

(a) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix}.$$

(b) What values of p_{01}, p_{10} maximize the entropy rate?

(c) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ 1 & 0 \end{bmatrix}.$$

(d) Find the maximum value of the entropy rate of the Markov chain of part (c). We expect that the maximizing value of p should be less than $\frac{1}{2}$, since the 0 state permits more information to be generated than the 1 state.

(e) Let $N(t)$ be the number of allowable state sequences of length t for the Markov chain of part (c). Find $N(t)$ and calculate

$$H_0 = \lim_{t \to \infty} \frac{1}{t} \log N(t).$$

[*Hint*: Find a linear recurrence that expresses $N(t)$ in terms of $N(t-1)$ and $N(t-2)$. Why is H_0 an upper bound on the entropy rate of the Markov chain? Compare H_0 with the maximum entropy found in part (d).]

4.11 *Stationary processes*. Let $\ldots, X_{-1}, X_0, X_1, \ldots$ be a stationary (not necessarily Markov) stochastic process. Which of the following statements are true? Prove or provide a counterexample.

(a) $H(X_n|X_0) = H(X_{-n}|X_0)$.

(b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

(c) $H(X_n|X_1, X_2, \ldots, X_{n-1}, X_{n+1})$ is nonincreasing in n.

(d) $H(X_n|X_1, \ldots, X_{n-1}, X_{n+1}, \ldots, X_{2n})$ is nonincreasing in n.

4.24 *Entropy rates*. Let $\{X_i\}$ be a stationary process. Let $Y_i = (X_i, X_{i+1})$. Let $Z_i = (X_{2i}, X_{2i+1})$. Let $V_i = X_{2i}$. Consider the entropy rates $H(\mathcal{X})$, $H(\mathcal{Y})$, $H(\mathcal{Z})$, and $H(\mathcal{V})$ of the processes $\{X_i\}$,$\{Y_i\}$, $\{Z_i\}$, and $\{V_i\}$. What is the inequality relationship $\leq$, $=$, or $\geq$ between each of the pairs listed below?

(a) $H(\mathcal{X}) \gtreqless H(\mathcal{Y})$.

(b) $H(\mathcal{X}) \gtreqless H(\mathcal{Z})$.

(c) $H(\mathcal{X}) \gtreqless H(\mathcal{V})$.

(d) $H(\mathcal{Z}) \gtreqless H(\mathcal{X})$.

5.3 *Slackness in the Kraft inequality*. An instantaneous code has word lengths $l_1, l_2, \ldots, l_m$, which satisfy the strict inequality

$$\sum_{i=1}^{m} D^{-l_i} < 1.$$

The code alphabet is $\mathcal{D} = \{0, 1, 2, \ldots, D-1\}$. Show that there exist arbitrarily long sequences of code symbols in $\mathcal{D}^*$ which cannot be decoded into sequences of codewords.

5.6 *Bad codes*. Which of these codes cannot be Huffman codes for any probability assignment?

(a) {0, 10, 11}

(b) {00, 01, 10, 110}

(c) {01, 10}

5.7 *Huffman 20 questions*. Consider a set of n objects. Let $X_i = 1$ or 0 accordingly as the ith object is good or defective. Let $X_1, X_2, \ldots, X_n$ be independent with $\Pr\{X_i = 1\} = p_i$; and $p_1 > p_2 > \cdots > p_n > \frac{1}{2}$. We are asked to determine the set of all defective objects. Any yes–no question you can think of is admissible.

(a) Give a good lower bound on the minimum average number of questions required.

(b) If the longest sequence of questions is required by nature's answers to our questions, what (in words) is the last question we should ask? What two sets are we distinguishing with this question? Assume a compact (minimum average length) sequence of questions.

(c) Give an upper bound (within one question) on the minimum average number of questions required.

5.16 *Huffman codes*. Consider a random variable X that takes six values $\{A, B, C, D, E, F\}$ with probabilities 0.5, 0.25, 0.1, 0.05, 0.05, and 0.05, respectively.

(a) Construct a binary Huffman code for this random variable. What is its average length?

(b) Construct a quaternary Huffman code for this random variable [i.e., a code over an alphabet of four symbols (call them a, b, c and d)]. What is the average length of this code?

(c) One way to construct a binary code for the random variable is to start with a quaternary code and convert the symbols into binary using the mapping $a \to 00$, $b \to 01$, $c \to 10$, and $d \to 11$. What is the average length of the binary code for the random variable above constructed by this process?

(d) For any random variable X, let L_H be the average length of the binary Huffman code for the random variable, and let L_{QB} be the average length code constructed by first building a quaternary Huffman code and converting it to binary. Show that

$$L_H \le L_{QB} < L_H + 2. \tag{5.146}$$

(e) The lower bound in the example is tight. Give an example where the code constructed by converting an optimal quaternary code is also the optimal binary code.

(f) The upper bound (i.e., $L_{QB} < L_H + 2$) is not tight. In fact, a better bound is $L_{QB} \le L_H + 1$. Prove this bound, and provide an example where this bound is tight.

5.18 *Classes of codes*. Consider the code $\{0, 01\}$.

(a) Is it instantaneous?

(b) Is it uniquely decodable?

(c) Is it nonsingular?

HW#5

5.12 *Shannon codes and Huffman codes.* Consider a random variable X that takes on four values with probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$.

(a) Construct a Huffman code for this random variable.

(b) Show that there exist two different sets of optimal lengths for the codewords; namely, show that codeword length assignments $(1, 2, 3, 3)$ and $(2, 2, 2, 2)$ are both optimal.

(c) Conclude that there are optimal codes with codeword lengths for some symbols that exceed the Shannon code length $\left\lceil \log \frac{1}{p(x)} \right\rceil$.

5.16 *Huffman codes.* Consider a random variable X that takes six values $\{A, B, C, D, E, F\}$ with probabilities 0.5, 0.25, 0.1, 0.05, 0.05, and 0.05, respectively.

(a) Construct a binary Huffman code for this random variable. What is its average length?

(b) Construct a quaternary Huffman code for this random variable [i.e., a code over an alphabet of four symbols (call them a, b, c and d)]. What is the average length of this code?

(c) One way to construct a binary code for the random variable is to start with a quaternary code and convert the symbols into binary using the mapping $a \to 00$, $b \to 01$, $c \to 10$, and $d \to 11$. What is the average length of the binary code for the random variable above constructed by this process?

(d) For any random variable X, let L_H be the average length of the binary Huffman code for the random variable, and let L_{QB} be the average length code constructed by first building a quaternary Huffman code and converting it to binary. Show that

$$L_H \le L_{QB} < L_H + 2. \qquad (5.146)$$

(e) The lower bound in the example is tight. Give an example where the code constructed by converting an optimal quaternary code is also the optimal binary code.

(f) The upper bound (i.e., $L_{QB} < L_H + 2$) is not tight. In fact, a better bound is $L_{QB} \le L_H + 1$. Prove this bound, and provide an example where this bound is tight.

5.18 *Classes of codes*. Consider the code $\{0, 01\}$.

(a) Is it instantaneous?

(b) Is it uniquely decodable?

(c) Is it nonsingular?

5.20 *Huffman codes with costs*. Words such as "Run!", "Help!", and "Fire!" are short, not because they are used frequently, but perhaps because time is precious in the situations in which these words are required. Suppose that $X = i$ with probability p_i, $i = 1, 2, \ldots, m$. Let l_i be the number of binary symbols in the codeword associated with $X = i$, and let c_i denote the cost per letter of the codeword when $X = i$. Thus, the average cost C of the description of X is $C = \sum_{i=1}^{m} p_i c_i l_i$.

(a) Minimize C over all $l_1, l_2, \ldots, l_m$ such that $\sum 2^{-l_i} \leq 1$. Ignore any implied integer constraints on l_i. Exhibit the minimizing $l_1^*, l_2^*, \ldots, l_m^*$ and the associated minimum value C^*.

(b) How would you use the Huffman code procedure to minimize C over all uniquely decodable codes? Let C_{Huffman} denote this minimum.

(c) Can you show that

$$C^* \leq C_{\text{Huffman}} \leq C^* + \sum_{i=1}^{m} p_i c_i?$$

5.19 *The game of Hi-Lo*

(a) A computer generates a number X according to a known probability mass function $p(x), x \in \{1, 2, \ldots, 100\}$. The player asks a question, "Is $X = i$?" and is told "Yes," "You're too high," or "You're too low." He continues for a total of six questions. If he is right (i.e., he receives the answer "Yes") during this sequence, he receives a prize of value $v(X)$. How should the player proceed to maximize his expected winnings?

(b) Part (a) doesn't have much to do with information theory. Consider the following variation: $X \sim p(x)$, prize $= v(x)$, $p(x)$ known, as before. But *arbitrary* yes–no questions are asked sequentially until X is determined. ("Determined" doesn't mean that a "Yes" answer is received.) Questions cost 1 unit each. How should the player proceed? What is the expected payoff?

(c) Continuing part (b), what if $v(x)$ is fixed but $p(x)$ can be chosen by the computer (and then announced to the player)? The computer wishes to minimize the player's expected return. What should $p(x)$ be? What is the expected return to the player?

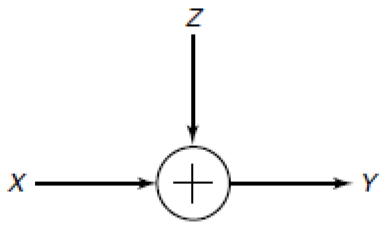
5.21 *Conditions for unique decodability*. Prove that a code C is uniquely decodable if (and only if) the extension

$$C^k(x_1, x_2, \ldots, x_k) = C(x_1)C(x_2)\cdots C(x_k)$$

is a one-to-one mapping from $\mathcal{X}^k$ to D^* for every $k \geq 1$. (The "only if" part is obvious.)

HW#6

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.

7.3 *Channels with memory have higher capacity.* Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $\Pr\{Z_i = 1\} = p = 1 - \Pr\{Z_i = 0\}$, but that $Z_1, Z_2, \ldots, Z_n$ are not necessarily independent. Assume that Z^n is independent of the input X^n. Let $C = 1 - H(p, 1 - p)$. Show that $\max_{p(x_1, x_2, \ldots, x_n)} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \geq nC$.

7.4 *Channel capacity.* Consider the discrete memoryless channel $Y = X + Z \pmod{11}$, where

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3} \end{pmatrix}$$

and $X \in \{0, 1, \ldots, 10\}$. Assume that Z is independent of X.

(a) Find the capacity.

(b) What is the maximizing $p^*(x)$?

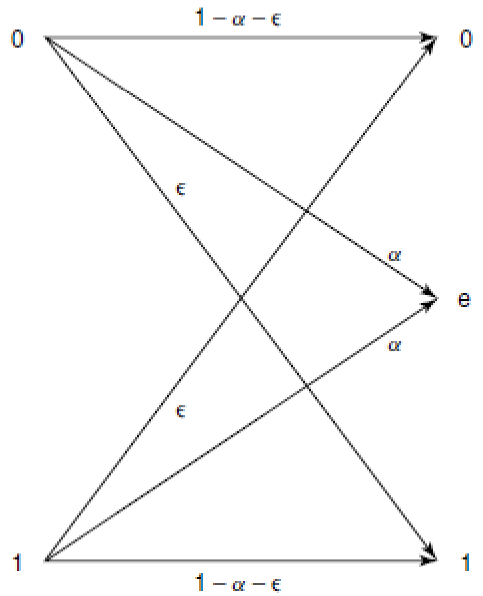
7.8 *Z-channel.* The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad x, y \in \{0, 1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

7.9 *Suboptimal codes.* For the Z-channel of Problem 7.8, assume that we choose a $(2^{nR}, n)$ code at random, where each codeword is a sequence of *fair* coin tosses. This will not achieve capacity. Find the maximum rate R such that the probability of error $P_e^{(n)}$, averaged over the randomly generated codes, tends to zero as the block length n tends to infinity.

7.13 *Erasures and errors in a binary channel.* Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be ϵ and the probability of erasure be α, so the channel is follows:



(a) Find the capacity of this channel.

(b) Specialize to the case of the binary symmetric channel ($\alpha = 0$).

(c) Specialize to the case of the binary erasure channel ($\epsilon = 0$).

HW#7

7.16 *Encoder and decoder as part of the channel.* Consider a binary symmetric channel with crossover probability 0.1. A possible coding scheme for this channel with two codewords of length 3 is to encode message a_1 as 000 and a_2 as 111. With this coding scheme, we can consider the combination of encoder, channel, and decoder as forming a new BSC, with two inputs a_1 and a_2 and two outputs a_1 and a_2.

(a) Calculate the crossover probability of this channel.

(b) What is the capacity of this channel in bits per transmission of the original channel?

(c) What is the capacity of the original BSC with crossover probability 0.1?

(d) Prove a general result that for any channel, considering the encoder, channel, and decoder together as a new channel from messages to estimated messages will not increase the capacity in bits per transmission of the original channel.

7.26 *Noisy typewriter.* Consider the channel with $x, y \in \{0, 1, 2, 3\}$ and transition probabilities $p(y|x)$ given by the following matrix:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

(a) Find the capacity of this channel.

(b) Define the random variable $z = g(y)$, where

$$g(y) = \begin{cases} A & if\, y \in \{0, 1\} \\ B & if\, y \in \{2, 3\}. \end{cases}$$

For the following two PMFs for x, compute $I(X; Z)$:

(i)

$$p(x) = \begin{cases} \frac{1}{2} & if\, x \in \{1, 3\} \\ 0 & if\, x \in \{0, 2\}. \end{cases}$$

(ii)

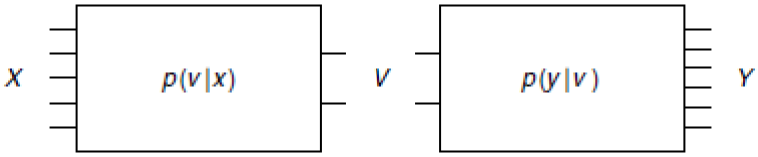
$$p(x) = \begin{cases} 0 & if\, x \in \{1, 3\} \\ \frac{1}{2} & if\, x \in \{0, 2\}. \end{cases}$$

(c) Find the capacity of the channel between x and z, specifically where $x \in \{0, 1, 2, 3\}$, $z \in \{A, B\}$, and the transition probabilities $P(z|x)$ are given by

$$p(Z = z|X = x) = \sum_{g(y_0)=z} P(Y = y_0|X = x).$$

(d) For the X distribution of part (i) of (b), does $X \to Z \to Y$ form a Markov chain?

7.25 *Bottleneck channel.* Suppose that a signal $X \in \mathcal{X} = \{1, 2, \ldots, m\}$ goes through an intervening transition $X \longrightarrow V \longrightarrow Y$:



where $x = \{1, 2, \ldots, m\}$, $y = \{1, 2, \ldots, m\}$, and $v = \{1, 2, \ldots, k\}$. Here $p(v|x)$ and $p(y|v)$ are arbitrary and the channel has transition probability $p(y|x) = \sum_v p(v|x)p(y|v)$. Show that $C \leq \log k$.

7.26 *Noisy typewriter.* Consider the channel with $x, y \in \{0, 1, 2, 3\}$ and transition probabilities $p(y|x)$ given by the following matrix:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

(a) Find the capacity of this channel.

(b) Define the random variable $z = g(y)$, where

$$g(y) = \begin{cases} A & if\, y \in \{0, 1\} \\ B & if\, y \in \{2, 3\}. \end{cases}$$

For the following two PMFs for x, compute $I(X; Z)$:

(i)

$$p(x) = \begin{cases} \frac{1}{2} & if\, x \in \{1, 3\} \\ 0 & if\, x \in \{0, 2\}. \end{cases}$$

(ii)

$$p(x) = \begin{cases} 0 & if\, x \in \{1, 3\} \\ \frac{1}{2} & if\, x \in \{0, 2\}. \end{cases}$$

(c) Find the capacity of the channel between x and z, specifically where $x \in \{0, 1, 2, 3\}$, $z \in \{A, B\}$, and the transition probabilities $P(z|x)$ are given by

$$p(Z = z|X = x) = \sum_{g(y_0)=z} P(Y = y_0|X = x).$$

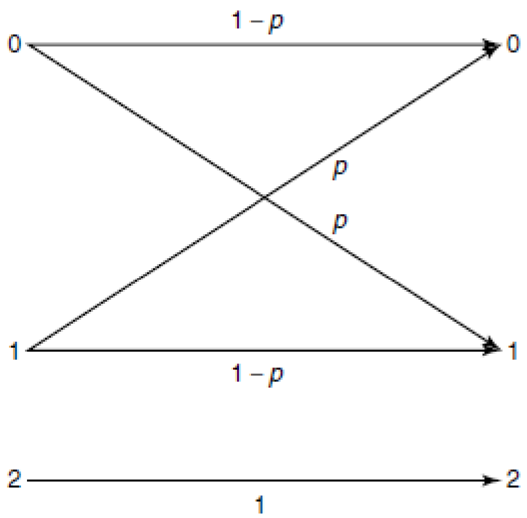
(d) For the X distribution of part (i) of (b), does $X \to Z \to Y$ form a Markov chain?

7.28 *Choice of channels.* Find the capacity C of the union of two channels $(\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$, where at each time, one can send a symbol over channel 1 or channel 2 but not both. Assume that the output alphabets are distinct and do not intersect.

(a) Show that $2^C = 2^{C_1} + 2^{C_2}$. Thus, 2^C is the effective alphabet size of a channel with capacity C.

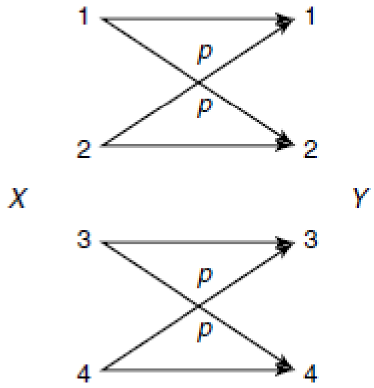
(b) Compare with Problem 2.10 where $2^H = 2^{H_1} + 2^{H_2}$, and interpret part (a) in terms of the effective number of noise-free symbols.

(c) Use the above result to calculate the capacity of the following channel.

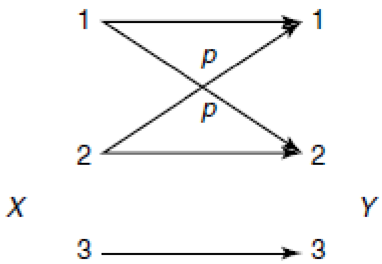


7.34 *Capacity*. Find the capacity of

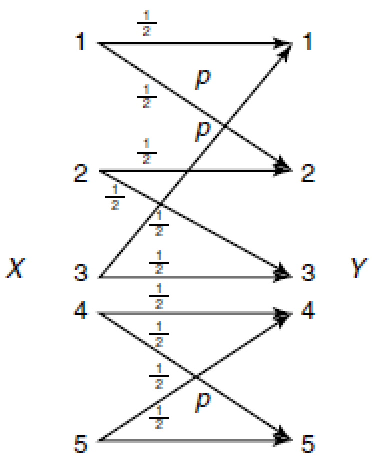
(a) Two parallel BSCs:



(b) BSC and a single symbol:



(c) BSC and a ternary channel:



(d) Ternary channel:

$$p(y|x) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}. \tag{7.167}$$

7.36 *Channel with memory*. Consider the discrete memoryless channel $Y_i = Z_i X_i$ with input alphabet $X_i \in \{-1, 1\}$.

(a) What is the capacity of this channel when $\{Z_i\}$ is i.i.d. with

$$Z_i = \begin{cases} 1, & p = 0.5 \\ -1, & p = 0.5? \end{cases} \tag{7.168}$$

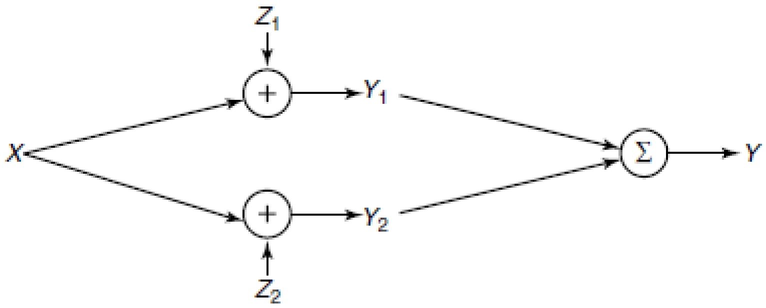
Now consider the channel with memory. Before transmission begins, Z is randomly chosen and fixed for all time. Thus, $Y_i = Z X_i$.

(b) What is the capacity if

$$Z = \begin{cases} 1, & p = 0.5 \\ -1, & p = 0.5? \end{cases} \tag{7.169}$$

HW#8

9.7 *Multipath Gaussian channel*. Consider a Gaussian noise channel with power constraint P, where the signal takes two different paths and the received noisy signals are added together at the antenna.

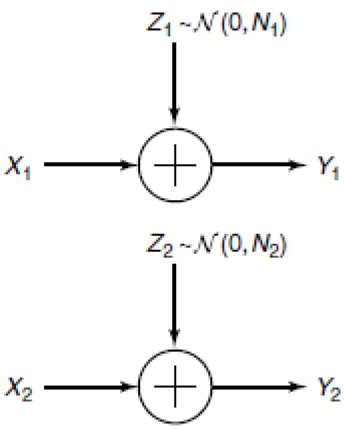


(a) Find the capacity of this channel if Z_1 and Z_2 are jointly normal with covariance matrix

$$K_Z = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}.$$

(b) What is the capacity for $\rho = 0$, $\rho = 1$, $\rho = -1$?

9.8 *Parallel Gaussian channels*. Consider the following parallel Gaussian channel:



where $Z_1 \sim \mathcal{N}(0,N_1)$ and $Z_2 \sim \mathcal{N}(0,N_2)$ are independent Gaussian random variables and $Y_i = X_i + Z_i$. We wish to allocate power to the two parallel channels. Let β_1 and β_2 be fixed. Consider a total cost constraint $\beta_1 P_1 + \beta_2 P_2 \leq \beta$, where P_i is the power allocated to the ith channel and β_i is the cost per unit power in that channel. Thus, $P_1 \geq 0$ and $P_2 \geq 0$ can be chosen subject to the cost constraint β.

(a) For what value of β does the channel stop acting like a single channel and start acting like a pair of channels?

(b) Evaluate the capacity and find P_1 and P_2 that achieve capacity for $\beta_1 = 1$, $\beta_2 = 2$, $N_1 = 3$, $N_2 = 2$, and $\beta = 10$.

9.9 *Vector Gaussian channel.* Consider the vector Gaussian noise channel

$$Y = X + Z,$$

where $X = (X_1, X_2, X_3)$, $Z = (Z_1, Z_2, Z_3)$, $Y = (Y_1, Y_2, Y_3)$, $E\|X\|^2 \leq P$, and

$$Z \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}\right).$$

Find the capacity. The answer may be surprising.

9.10 *Capacity of photographic film.* Here is a problem with a nice answer that takes a little time. We're interested in the capacity of photographic film. The film consists of silver iodide crystals, Poisson distributed, with a density of λ particles per square inch. The film is illuminated without knowledge of the position of the silver iodide particles. It is then developed and the receiver sees only the silver iodide particles that have been illuminated. It is assumed that light incident on a cell exposes the grain if it is there and otherwise results in a blank response. Silver iodide particles that are not illuminated and vacant portions of the film remain blank. The question is: What is the capacity of this film?

We make the following assumptions. We grid the film very finely into cells of area dA. It is assumed that there is at most one silver iodide particle per cell and that no silver iodide particle is intersected by the cell boundaries. Thus, the film can be considered to be a large number of parallel binary asymmetric channels with crossover probability $1 - \lambda dA$. By calculating the capacity of this binary asymmetric channel to first order in dA (making the

necessary approximations), one can calculate the capacity of the film in bits per square inch. It is, of course, proportional to λ. The question is: What is the multiplicative constant?
The answer would be λ bits per unit area if both illuminator and receiver knew the positions of the crystals.

9.11 *Gaussian mutual information.* Suppose that (X, Y, Z) are jointly Gaussian and that $X \to Y \to Z$ forms a Markov chain. Let X and Y have correlation coefficient ρ_1 and let Y and Z have correlation coefficient ρ_2. Find $I(X; Z)$.

9.14 *Additive noise channel.* Consider the channel $Y = X + Z$, where X is the transmitted signal with power constraint P, Z is independent additive noise, and Y is the received signal. Let

$$Z = \begin{cases} 0 & \text{with probability } \frac{1}{10} \\ Z^* & \text{with probability } \frac{9}{10}, \end{cases}$$

where $Z^* \sim N(0, N)$. Thus, Z has a mixture distribution that is the mixture of a Gaussian distribution and a degenerate distribution with mass 1 at 0.

(a) What is the capacity of this channel? This should be a pleasant surprise.

(b) How would you signal to achieve capacity?

9.15 *Discrete input, continuous output channel.* Let $\Pr\{X = 1\} = p$, $\Pr\{X = 0\} = 1 - p$, and let $Y = X + Z$, where Z is uniform over the interval $[0, a]$, $a > 1$, and Z is independent of X.

(a) Calculate

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

(b) Now calculate $I(X; Y)$ the other way by

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

(c) Calculate the capacity of this channel by maximizing over p.