Telecom Paris ACCQ202

ASSIGNMENT 2

Exercise 1 (Block coding)**.** Suppose a source generates $X_1, X_2, \ldots, X_n$ in an i.i.d. fashion and suppose we encode these symbols all at once, instead of symbol-by-symbol. Exhibit a coding scheme whose per-symbol expected length lies between $H(X)$ and $H(X) + 1/n$.

Solution. Use a Shannon code over a super-symbol $(X_1, X_2, \ldots, X_n)$. □

Exercise 2 (Bad codes)**.** Which of the following binary codes cannot be a Huffman code for any distribution? Why?

a. 0, 10, 111, 101

b. 00, 010, 011, 10, 110

c. 1, 000, 001, 010, 011

Solution. a. A Huffman code is a prefix free code but here we have 10 which is a prefix of 101.

b. This is not a Huffman code since codeword 110 does not have any sibling. Hence, the code could be improved by replacing this codeword with 11.

c. This is a Huffman code for distribution $(0.4, 0.15, 0.15, 0.15, 0.15)$ for instance.

□

Exercise 3 (Huffman codes)**.** For the distribution $(p_1, ..., p_n)$, where

$$p_1 > p_2 > \cdots > p_n > 0,$$

we have an optimal binary prefix code. Show that

a. If $p_1 > 2/5$ then the corresponding codeword has length 1.

b. If $p_1 < 1/3$ then the corresponding codeword has length at least 2.

Solution. Consider the algorithm for constructing Huffman codes. Let $(q_1, ..., q_k)$, $k \geq 1$ be the distribution at the $(n-k)$-th iteration of the algorithm, sorted in the decreasing order. Note that for $k = n$, $(q_1, ..., q_k) = (p_1, ..., p_n)$. In the next iteration, the two smallest probabilities, q_{k-1} and q_k are replaced by their sum $q_{k-1} + q_k$, then a Huffman code for set of probabilities $(q_1, ..., q_{k-2}, q_{k-1} + q_k)$ is constructed. Suppose the corresponding codes are $(C_1, \cdots, C_{k-1})$, then the Huffman code for distribution $(q_1, ..., q_k)$ will be $(C_1, \cdots, C_{k-2}, C_{k-1} * 0, C_{k-1} * 1)$ where $*$ denotes concatenation.

a. Suppose, by contradiction, that the codeword for p_1 is greater or equal than 2, and consider the first place where p_1 becomes the second largest probability. More precisely, let

$$q_1 \geq q_2 \geq \cdots \geq q_{k+1},$$

$k \geq 3$, $q_1 = p_1$ and $q_k + q_{k+1} \geq q_1$. Now, notice that $q_2 \geq q_k \geq \frac{q_k + q_{k+1}}{2} \geq \frac{q_1}{2}$. So, we have

$$\begin{aligned} 1 = \sum_{i=1}^{k+1} q_i &\geq q_1 + q_2 + q_k + q_{k+1} \\ &\geq q_1 + \frac{q_1}{2} + q_1 = \frac{5}{2} p_1 \\ &> \frac{5}{2} \cdot \frac{2}{5} = 1, \end{aligned}$$

a contradiction.

b. Similarly as above consider (q_1, q_2, q_3) with

$$q_1 \geq q_2 \geq q_3,$$

and $q_1 = p_1$. Then

$$1 = \sum_{i=1}^{3} q_i \leq 3q_1 = 3p_1 < 3 \cdot \frac{1}{3} = 1,$$

a contradiction.

□

Exercise 4 (Huffman code for a wrong source)**.** The purpose of this problem is to see what happens when you design a code for the wrong set of probabilities. Consider a Huffman code that is designed for a symbol source whose probability is given by P. Suppose that we use this code for the source with distribution Q. Find the average number of binary code symbols per source symbol and compare it with the entropy of the source for the following.

1. $P = (0.5, 0.3, 0.2), \;\; Q = (0.65, 0.2, 0.15)$
2. $P = (0.5, 0.3, 0.2), \;\; Q = (0.15, 0.2, 0.65)$
3. $P = (0.5, 0.3, 0.1, 0.1), \;\; Q = (0.3, 0.2, 0.3, 0.2)$

Can the optimal codes for P and Q be the same?

Solution. Let $L(X)$ denote the length of the codeword for symbol X. Let $\mathbb{E}_Q[L]$ denote the expected value of $L(X)$ and $H_Q(X)$ denote the entropy when X has distribution Q.

1. A code for P is $(0, 10, 11)$ and $\mathbb{E}_Q[L] = 0.65 \times 1 + 0.2 \times 2 + 0.15 \times 2 = 1.35$. We calculate the entropy to be $H_Q(X) \approx 1.28$. The optimal code for P and Q could be the same.

2. A code for P is $(0, 10, 11)$ and $\mathbb{E}_Q[L] = 0.15 \times 1 + 0.2 \times 2 + 0.65 \times 2 = 1.65$. The entropy is the same as in the case above. The optimal code for P and Q are different but the set of codewords could be the same.

3. A code for P is $(0, 10, 110, 111)$ and $\mathbb{E}_Q[L] = 0.3 \times 1 + 0.2 \times 2 + 0.3 \times 3 + 0.2 \times 3 = 2.2$. We calculate the entropy to be $H_Q(X) \approx 1.97$. The code and the set of codewords are different for P and Q.

□

Exercise 5 (Shannon code, divergence)**.** Suppose we wrongly estimate the probability of a source of information, and that we use a Shannon code for a distribution Q whereas the true distribution is P. Show that

$$H(P) + D(P||Q) \leq L(C) \leq H(P) + D(P||Q) + 1\,.$$

So $D(P||Q)$ can be interpreted as the increase in descriptive complexity due to incorrect information.

Solution. For a Shannon code for distribution Q, the length of the codeword of a symbol X is $\left\lceil \log \frac{1}{Q(X)} \right\rceil$. Let $\mathbb{E}_P[\cdot]$ denote the expectation under the distribution P. Observe that

$$\log \frac{1}{Q(X)} \leq \left\lceil \log \frac{1}{Q(X)} \right\rceil \leq \log \frac{1}{Q(X)} + 1.$$

Then, the result follows from the following.

$$\begin{aligned}
\mathbb{E}_P \left[\log \frac{1}{Q(X)} \right] &= \sum_x P(x) \log \frac{1}{Q(x)} \\
&= \sum_x P(x) \log \left(\frac{P(x)}{Q(x)} \frac{1}{P(x)} \right) \\
&= \sum_x P(x) \log \frac{P(x)}{Q(x)} + \sum_x P(x) \log \frac{1}{P(x)} \\
&= D(P||Q) + H(P)
\end{aligned}$$

□

Exercise 6 (Huffman Codes)**.** The sequence of six independent realizations of source X is encoded symbol-by-symbol using a binary Huffman code. The resulted string is 10110000101. We know that the alphabet of X has five elements and that its distribution is either $(0.4, 0.3, 0.2, 0.05, 0.05)$ or $(0.3, 0.25, 0.2, 0.2, 0.05)$. Which of them is the distribution of X?

Solution. By the result in Exer.$3b$., every codeword in a Huffman code for the second distribution should be of length at least 2. We know that there are 6 realizations of X and hence the string 10110000101 (of length 11) could not have been produced by a Huffman code for the second distribution. A possible Huffman code for the first distribution, namely $(0.4, 0.3, 0.2, 0.05, 0.05)$ is $(1, 01, 000, 0010, 0011)$ (Note that Huffman codes are not unique!). Using this code, one can decode the string 10110000101 as $1, 01, 1, 000, 01, 01$. Hence, the probability distribution of X is $(0.4, 0.3, 0.2, 0.05, 0.05)$. □

Exercise 7 (Guessing, Huffman)**.** There are 6 bottles of wine, one of which you know has gone bad. You do not know which bottle contains the bad wine, but you know that the probability of each bottle being bad is $(8/23, 6/23, 4/23, 2/23, 2/23, 1/23)$. The bad wine has a distinctive taste. To find the bad wine your friend suggests you to taste a little bit of each wine until you find the bad wine.

a. To have the least number of tastings on average, what should your strategy be? Which bottle should be tasted first?

b. What is the average number of tastings to find the bad wine?

c. Calculate the minimum average number of tastings if you are allowed to taste a mixture of different wines and detect a bad wine's taste inside (the distinctive taste is retained even when mixed with other good wines).

d. Is the strategy studied in (a) optimal if you are allowed to mix wines?

Solution. a. A guessing strategy for a random variable X can be written as a vector $G = (g_1, g_2, \ldots)$ with $g_i \in \mathcal{X}$ being the i-th guess of X. With this notation, the expected number of guesses is given by $\mathbb{E}(G) = \sum_i i\mathbb{P}(X = g_i)$. Now assume that for some $i < j$ we have $\mathbb{P}(X = g_j) > \mathbb{P}(X = g_i)$, and consider the new strategy G' where g_i and g_j are swapped. It then follows that $E(G) - E(G') = (j - i)(\mathbb{P}(X = g_j) - \mathbb{P}(X = g_i)) > 0$. It follows that the strategy that guesses the values of X in decreasing order of proabilities minimizes the expected number of guesses.

b. $56/23$

c. A sequence of questions is equivalent to a code. Indeed, any question depends on the sequence of answers to the questions before it. Since the sequence of answers uniquely determines a particular sample of X, if we represent the sequence of yes-no answers by 0 and 1, each sample of X is associated to a codeword. Conversely, from a binary code for each possible sample of X, we can find a sequence of questions that corresponds to the code. The i-th question is "Is the i-th bit bit equal to 1?" or, more specifically, "Does the X belongs to the set of samples whose codewords have the i-th bit equal to 1?"

Therefore, from the equivalence between quessing strategy and code, finding a quessing strategy that minimizes the number of questions is equivalent to finding a code whose average length is minimal. An optimal strategy to identify the bad bottle is thus obtained via the construction of the Huffman code of the bad bottle probability distribution. Note that we use here the fact that we are allowed to mix wines, hence we can ask, at each step, whether the bad wine belongs to some particular subset of bottles or not.

□

Exercise 8 (Entropy and Yes/No questions)**.** We are asked to determine an object by asking yes-no questions. The object is drawn randomly from a finite set according to a certain distribution. Playing optimally, we need 38.5 questions on the average to find the object. At least how many elements does the finite set have?

Solution. An optimal yes/no scheme corresponds to an optimal source code whose expected length is at most $H(X) + 1$, where X is the hidden object. Hence $H(X) + 1 \geq 38.5$. On the other hand we have $\log |\mathcal{X}| \geq H(X)$. These two yield that $\log |\mathcal{X}| \geq 37.5$, and so $n \geq \lceil 2^{37.5} \rceil$. □

Exercise 9 (Pure randomness from biased distributions)**.** Let $X_1, X_2, \ldots, X_n$ denote the outcomes of independent flips of a biased coin. Thus, for $i = 1, \ldots, n$ we have $\Pr(X_i = 1) = p, \Pr(X_i = 0) = 1 - p$, where p is unknown. We wish to obtain a sequence $Z_1, Z_2, \ldots, Z_K$ of fair coin flips from $X_1, X_2, \ldots, X_n$. To this end let $f : \mathcal{X}^n \to \{0,1\}^\star$ (where $\{0,1\}^\star = \{\Lambda, 0, 1, 00, 01, \ldots\}$ is the set of all finite length binary sequences including the null string Λ) be a mapping $f(X_1, X_2, \ldots, X_n) = (Z_1, Z_2, ..., Z_K)$, such that $Z_i \sim$ Bernoulli$(1/2)$ and where K possibly depends on $(X_1, ..., X_n)$. For the sequence $Z_1, Z_2, ..., Z_K$ to correspond to fair coin flips, the map f from biased coin flips to fair flips must have the property that all 2^k sequences $(z_1, z_2, \ldots, z_k)$ of a given length k have equal probability (possibly 0). For example, for $n = 2$, the map $f(01) = 0, f(10) = 1, f(00) = f(11) = \Lambda$ has the property that $\Pr(Z_1 = 1|K = 1) = \Pr(Z_1 = 0|K = 1) = 1/2$.

a. Justify the following (in)equalities

$$\begin{aligned} nH_b(p) &\overset{(a)}{=} H(X_1, ..., X_n) \\ &\overset{(b)}{\geq} H(Z_1, Z_2, \ldots, Z_K, K) \\ &\overset{(c)}{=} H(K) + H(Z_1, Z_2, ..., Z_K|K) \\ &\overset{(d)}{=} H(K) + E(K) \\ &\overset{(e)}{\geq} E(K) \end{aligned}$$

where $E(K)$ denotes the expectation of K. Thus, on average, no more than $nH_b(p)$ fair coin tosses can be derived from $(X_1, ..., X_n)$.

b. Exhibit a good map f on sequences of length $n = 4$.

Solution. a. (a.) the X_i's are i.i.d. Bernoulli(p) distributed; (b) (Z^K, K) is a function of X^n; (c) chain rule; (d) given $K = k$, $(Z_1, Z_2, \ldots, Z_k)$ is an i.i.d. Bernoulli$(1/2)$ sequence, hence $H(Z_1, Z_2, ..., Z_K|K = k) = k$, from which the result follows; (d) non-negativity of the entropy.

b. One possibility is as follows. Let T_k be the set of binary sequences of length 4 with exactly k ones ($k \in \{0, 1, 2, \ldots, 4\}$). Observe that T_1 and T_3 each have four elements, and each contains equiprobable elements (obviously, the elements in T_1 have a different probability than those in T_3). We map the 4 elements in T_1 in 00, 01, 10, and 11, and similarly for T_3. I follows that, given $K = 2$, (Z_1, Z_2) are purely random. To see this note that for any pair of bit (i, j)

$$\begin{aligned} \text{Pr}((Z_1, Z_2) = (i, j)|K = 2) &= \text{Pr}((Z_1, Z_2) = (i, j)|X^4 \in T_1 \cup T_3) \\ &= \text{Pr}((Z_1, Z_2) = (i, j)|X^4 \in T_1)\text{Pr}(X^4 \in T_1|X^4 \in T_1 \cup T_3) \\ &+ \text{Pr}((Z_1, Z_2) = (i, j)|X^4 \in T_3)\text{Pr}(X^4 \in T_3|X^4 \in T_1 \cup T_3) \\ &= \frac{1}{4}\text{Pr}(X^4 \in T_1|X^4 \in T_1 \cup T_3) + \frac{1}{4}\text{Pr}(X^4 \in T_3|X^4 \in T_1 \cup T_3) \\ &= \frac{1}{4}\,. \end{aligned}$$

All the elements in T_0, T_2, and T_4 are mapped into Λ.

□

Exercise 10 (Entropy bound)**.** Let $p(x)$ be a probability mass function of random variable X. Prove that

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

for any $d \geq 0$. *Hint* – Use Markov's inequality.

Solution.

$$\begin{aligned} Pr\{p(X) \leq d\} &= Pr\{-\log p(X) \geq -\log d\} \\ &\leq \frac{\mathbb{E}\left[-\log p(X)\right]}{-\log d} \end{aligned}$$

by Markov's inequality. The result follows by observing that $\mathbb{E}\left[-\log p(X)\right] = H(X)$. □