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 $H(X) + 1/n$.

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

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.

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?

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.

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?

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?

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?

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 $n = 4$.

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.