INFORMATION AND CODING THEORY

Exercise sheet n° 4 :

4-1. *Lempel-Ziv code.*
(a) Consider a source with alphabet $\{\texttt{A}, \texttt{B}, \texttt{C}, \texttt{\_}\}$. Encode the sequence $\texttt{AA\_ABABBABC\_ABABC}$ with the Lempel-Ziv code. Establish a method to encode the position, and calculate the number of bits that are necessary to transmit this sequence compressed by Lempel-Ziv. What do you encounter at the last ABC? What would you encounter if the sequence was $\texttt{AA\_ABABBABC\_ACABC}$?
(b) Consider a source with alphabet $\{\texttt{A}, \texttt{B}\}$. Encode the sequence $\texttt{ABAAAAAAAAAAAAABB}$ with the Lempel-Ziv code. Calculate the number of bits necessary to transmit the sequence compressed by Lempel-Ziv and compare it with a naive encoding.
(c) Consider a source with alphabet $\{\texttt{A}, \texttt{B}\}$. Encode the sequence $\texttt{ABAAAAAAAAAAAAAAAAAAA}$ with the Lempel-Ziv code. Calculate the number of bits necessary to transmit the sequence compressed by Lempel Ziv and compare it with a naive encoding.

(d) The following is the encoded sequence that is obtained by using the Lempel-Ziv code. Reconstruct the original sequence.

$$
\begin{aligned}
&0, \texttt{A} \\
&1, \texttt{B} \\
&2, \texttt{C} \\
&0, \texttt{\_} \\
&2, \texttt{B} \\
&0, \texttt{B} \\
&6, \texttt{B} \\
&7, \texttt{B} \\
&0, \texttt{.}
\end{aligned}
$$

4-2. Each object in a set of n objects can either be faulty or intact. The random variable X_i takes the value 1 if the i-th object is faulty and 0 otherwise. We assume that the variables X_1, $X_2, \cdots, X_n$ are independent, with $\text{Prob}\{X_i = 1\} = p_i$ and $p_1 > p_2 > \cdots > p_n > 1/2$. The problem is to determine the set of all faulty objects with an optimal method.

(a) How can one find the optimal sequence of yes/no-questions that determines all objects that are faulty?

(b) What is the last question that has to be asked? Which two ensembles can one distinguish with this question?

4-3. A player A choses an object from a set and a player B tries to identify the chosen object with a series of yes/no-questions. We assume that player B uses an optimal code (which gives the minimal expected length) with respect to the distribution of objects of player A. One observes that, on average, player B needs to ask 35 questions to identify the object chosen by A. Find a lower bound on the number of objects in the set.

4-4. One is given n coins, where one of them may be counterfeit. A counterfeit coin has a different weight from an authentic one. In order to find the counterfeit coin, all coins are weighed.

(a) What is the probability distribution of the counterfeit coin that maximizes the Shannon entropy? Treat the existence or non-existence of a counterfeit coin as an extra element in the probability distribution.

(b) Establish the equation that bounds from below and above the expected number of weighs, if one uses an optimal method.

(c) Deduce the maximal number of coins n, given an average number of weighs k.

(d) Show that this maximum can be attained.

4-5. A random variable X can take m possible values and has an entropy $H(X)$. Suppose one is given an instantaneous code C for the source X, which has an expected length $L(C) = H(X)/\log_2 5 = H_5(X)$.

(a) Show that each symbol of X has a probability of the form 5^{-n}, where n is a positive integer.

(b) Show that m has to satisfy $m = 4k + 1$ where k is a positive integer.

Web site :

http ://quic.ulb.ac.be/teaching/