ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 13 Information Theory and Coding
Homework 6 Oct. 24, 2017

PROBLEM 1. Consider an information source that generates a sequence of independent identically distributed letters form a finite alphabet $\mathcal{X}$, with $p(x)$ denoting the probability that the letter x is generated. However we believe that the probability that the source generates x is $q(x)$. (Unless q is the same as p, we are wrong in our belief.)

Suppose that we have designed a prefix-free code C to minimize the expected codeword length under our belief. That is, C is chosen to minimize $\sum_x q(x) \operatorname{length}[C(x)]$.

(a) Let X be a source letter. Show that $E\big[-\log_2 q(X)\big] = H(p) + D(p\|q)$.

(b) Assuming that $q(x)$ is an integer power of $1/2$ for every x, express $\operatorname{length}[C(x)]$ in terms of $q(x)$.

(c) Still assuming that $q(x)$ is an integer power of $1/2$, express the difference between the true average codeword length and the true entropy of the source.

PROBLEM 2. From the notes on the Lempel–Ziv algorithm, we know that the maximum number of distinct words c a string of length n can be parsed into satisfies

$$n > c \log_K (c/K^3)$$

where K is the size of the alphabet the letters of the string belong to. This inequality lower bounds n in terms of c. We will now show that n can also be upper bounded in terms of c.

(a) Show that, if $n \geq \frac{1}{2}m(m-1)$, then $c \geq m$.

(b) Find a sequence for which the bound in (a) is met with equality.

(c) Show now that $n < \frac{1}{2}c(c+1)$.

PROBLEM 3. Let X be the channel input. Assume that the channel output Y is passed through a data processor in such a way that no information is lost. That is,

$$I(X;Y) = I(X;Z)$$

where Z is the processor output. Find an example where $H(Y) > H(Z)$ and find an example where $H(Y) < H(Z)$.
Hint: The data processor does not have to be deterministic

PROBLEM 4. Consider the discrete memoryless channel $Y = X + Z \pmod{11}$, where

$$\Pr(Z=1) = \Pr(Z=2) = \Pr(Z=3) = 1/3$$

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

(a) Find the capacity.

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

Problem 5. We are given a memoryless stationary binary symmetric channel BSC(ϵ). Namely, if $X_1, \dots, X_n \in \{0,1\}$ are the input of this channel and $Y_1, \dots, Y_n \in \{0,1\}$ are the output, we have:

$$P(Y_i|X_i, X^{i-1}, Y^{i-1}) = P(Y_i|X_i) = \begin{cases} 1-\epsilon & \text{if } Y_i = X_i, \\ \epsilon & \text{otherwise.} \end{cases}$$

Let W be a random variable that is uniform in $\{0,1\}$ and consider a communication system with feedback which transmits the value of W to the receiver as follows:

- At time $t = 1$, the transmitter sends $X_1 = W$ through the channel.
- At time $t = i + 1 \le n$, the transmitter gets the value of Y_i from the feedback and sends $X_{i+1} = Y_i$ through the channel.

(a) Give the capacity C of the channel in terms of ϵ, and show that $C = 0$ when $\epsilon = \frac{1}{2}$.

(b) Show that if $\epsilon = \frac{1}{2}$, $I(X^n; Y^n) = n - 1$. This means that $I(X^n; Y^n) \le nC$ does not hold for this system.

(c) Show that although $I(X^n; Y^n) > nC$ when $\epsilon = \frac{1}{2}$, we still have $I(W; Y^n) \le nC$.

Note that since W is the useful information that is being transmitted, it is the value of $I(W; Y^n)$ that we are interested in when we want to compute the amount of information that is shared with the receiver.