Telecom ParisTech SOCOM205, Information Theory
Prof. Aslan Tchamkerten February 2017

ASSIGNMENT 1: SOLUTIONS

Exercise 1 (Alternative definition of unique decodability)**.** An $f : \mathcal{X} \to \mathcal{Y}$ code is called uniquely decodable if for any messages $u = u_1 \cdots u_k$ and $v = v_1 \cdots v_k$ (where $u_1, v_1, \cdots, u_k, v_k \in \mathcal{X}$) with

$$f(u_1)f(u_2)\cdots f(u_k) = f(v_1)f(v_2)\cdots f(v_k),$$

we have $u_i = v_i$ for all i. That is, as opposed to the definition given in class, we require that the codes of any pair of messages with the same length are equal. Prove that the two definitions are equivalent.

Solution. The definition given class is as follows. An $f : \mathcal{X} \to \mathcal{Y}^*$ code is called uniquely decodable if for any messages $u = u_1 \cdots u_k$ and $v = v_1 \cdots v_m$ (where $u_1, \cdots, u_k, v_1, \cdots, v_m \in \mathcal{X}$) with

$$f(u_1)f(u_2)\cdots f(u_k) = f(v_1)f(v_2)\cdots f(v_m),$$

we have $k = m$ and $u_i = v_i$ for all i.

Trivially, by letting $m = k$ this definition implies the definition stated in the exercise. Now, we show the converse.

Suppose there exist $a = a_1 \cdots a_k$ and $b = b_1 \cdots b_m$ (where $a_1, \cdots, a_k, b_1, \cdots, b_m \in \mathcal{X}$) with

$$f(a_1)f(a_2)\cdots f(a_k) = f(b_1)f(b_2)\cdots f(b_m). \tag{1}$$

We should prove that $k = m$ and $a_i = b_i$ for all i. Take

$$u = ab = a_1 \cdots a_k b_1 \cdot b_m$$

$$v = ba = b_1 \cdots b_m a_1 \cdots a_k.$$

Now since $|u| = |v| = k + m$ and

$$f(u) = f(a_1)\cdots f(a_k)f(b_1)\cdots f(b_m) = f(b_1)\cdots f(b_m)f(a_1)\cdots f(a_k) = f(v)$$

due to the fact that $f(a_1)\cdots f(a_k) = f(b_1)\cdots f(b_m)$, the definition of the exercise implies that $u = v$, *i.e.*

$$a_1 \cdots a_k b_1 \cdots b_m = b_1 \cdots b_m a_1 \cdots a_k. \tag{2}$$

Now, suppose $k < m$, then the above equation implies that $a_i = b_i$ for $1 \leq i \leq k$ and so

$$f(a_1)f(a_2)\cdots f(a_k) = f(b_1)f(b_2)\cdots f(b_k).$$

This together with (1) gives

$$f(b_{k+1})\cdots f(b_m) = \phi,$$

a contradiction. With a similar argument one shows that $k > m$ cannot happen, and therefore $k = m$. So from (2), we have $a_i = b_i$ for all i. □

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

a. 0, 10, 111, 101

b. 00, 010, 011, 10, 110

c. 1, 000, 001, 010, 011

Solution. For a., a Huffman code is a prefix free code but here we have 10 which is a prefix of 101. For b., again this is not a Huffman code since codeword 110 doesn't have any sibling hence the code could be improved by replacing this codeword with 11. For c., this is a Huffman code for distribution $\{0.4, 0.3, 0.15, 0, 1, 0, 1\}$ for instance. □

Exercise 3 (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. Given in the course. □

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 three symbol source whose probability is given by $(0.5, 0.3, 0.2)$. Suppose that we use this code for the source $(0.15, 0.2, 0.65)$. Find the average number of binary code symbols per source symbol and compare it with the entropy of the source.

Solution. The required Huffman code for probability distribution $(0.5, 0.3, 0.2)$ is

$$X = \begin{pmatrix} A & B & C \\ 0.5 & 0.3 & 0.2 \\ 0 & 10 & 11 \end{pmatrix}$$

which by using probability distribution $(0.15, 0.2, 0.65)$ has the average length per source symbol is 1.85 and $H(X) \simeq 1.28$. □

Exercise 5 (Entropy)**.** Let X and Y be the outcomes of a pair of dice thrown independently (hence each independently takes on values in $\{1, 2, 3, 4, 5, 6\}$ with equal probabilities). Let $Z = X + Y$ and let $Q = Z \mod 2$. Compute the following entropies: $H(X)$, $H(Y)$, $H(Z)$, $H(Q)$.

Solution. X and Y are uniform random variables over $\{1, 2, 3, 4, 5, 6\}$, so

$$H(X) = H(Y) = \log_2(6).$$

The probability distribution of Z is

$$Z = \left(\begin{array}{ccccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{array} \right)$$

So, $H(Z) = 3.27$. Finally since Q takes values 0 and 1 equiprobably, we get $H(Q) = 1$.

□

Exercise 6 (Entropy)**.** Let X be a random variable taking values in M points $a_1, \ldots, a_M$ and let $p_X(a_M) = \alpha$. Show that

$$H(X) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) + (1-\alpha)H(Y)$$

where Y is a random variable taking values in $M-1$ points $a_1, \ldots, a_{M-1}$ with probabilities $P_Y(a_j) = P_X(a_j)/(1-\alpha)$ for $1 \leq j \leq M-1$. Show that

$$H(X) \leq -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) + (1-\alpha)\log(M-1)$$

and determine the condition for equality.

Solution.

$$\begin{aligned} H(X) &= -\sum_{i=1}^{M} p(a_i)\log p(a_i) \\ &= -\alpha\log\alpha - \sum_{i=1}^{M-1} p(a_i)\log p(a_i) \\ &= -\alpha\log\alpha - (1-\alpha)\sum_{i=1}^{M-1}\frac{p(a_i)}{1-\alpha}\log\left(\frac{p(a_i)}{(1-\alpha)}(1-\alpha)\right) \\ &= (1-\alpha)\log(1-\alpha) - (1-\alpha)H(Y) \end{aligned}$$

where for the fourth equality we used that $\sum_{i=1}^{M-1} p(a_i) = 1-\alpha$.

To prove that

$$H(X) \leq -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) + (1-\alpha)\log(M-1)$$

it suffices to observe that Y takes at most $M-1$ values, hence its entropy is at most $\log(M-1)$. □

Exercise 7 (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. Corresponding Huffman codes for distributions $\{0.4, 0.3, 0.2, 0.05, 0.05\}$ and $\{0.3, 0.25, 0.2, 0.2, 0.05\}$ are $\{1, 01, 000, 0010, 0011\}$ and $\{00, 01, 11, 100, 101\}$, respectively (Note that Huffman codes are not unique!). The first code can construct in this way $1, 01, 1, 000, 01, 01$, hence the probability distribution is $\{0.4, 0.3, 0.2, 0.05, 0.05\}$. Note that the second code cannot give the string since we know that there are 6 realizations of X. In fact, no Huffman code for the second distribution can produce the string. To see this note that any codeword of any Huffman code for the second distribution has length at least 2. Since the string corresponds to 6 symbols, the string would have to be of length at least 12 to be compatible with such a code. Since the string has length 11 this concludes the argument. □