ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 11 — Information Theory and Coding
Homework 5 — Oct. 13, 2015

Problem 1. Let $X_1, X_2, \dots$ be i.i.d. random variables with distribution $p(x)$ taking values in a finite set $\mathcal{X}$. Thus, $p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i)$.

(a) Use the law of large numbers to show that

$$-\frac{1}{n} \log p(X_1, \dots, X_n) \to H(X)$$

in probability.

Let $q(x_1, \dots, x_n) = \prod_{i=1}^n q(x_i)$, where $q(x)$ is another probability distribution on $\mathcal{X}$.

(b) Evaluate

$$\lim_{n\to\infty} -\frac{1}{n} \log q(X_1, \dots, X_n).$$

(c) Now evaluate the limit of the log-likelihood-ratio

$$\frac{1}{n} \log \frac{q(X_1, \dots, X_n)}{p(X_1, \dots, X_n)}.$$

Problem 2. Assume $\{X_n\}_{-\infty}^{\infty}$ and $\{Y_n\}_{-\infty}^{\infty}$ are two i.i.d. processes (individually) with the same alphabet, with the same entropy rate $H(X_0) = H(Y_0) = 1$ and independent from each other. We construct two processes Z and W as follows:

- To construct the process Z, we flip a fair coin and depending on the result $\Theta \in \{0, 1\}$ we select one of the processes. In other words, $Z_n = \Theta X_n + (1 - \Theta) Y_n$.
- To construct the process W, we do the coin flip at every time n. In other words, at every time n we flip a coin and depending on the result $\Theta_n \in \{0, 1\}$ we select X_n or Y_n as follows $W_n = \Theta_n X_n + (1 - \Theta_n) Y_n$.

(a) Are Z and W stationary processes? Are they i.i.d. processes?

(b) Find the entropy rate of Z and W. How do they compare? When are they equal?

Recall that the entropy rate of the process U (if exists) is $\lim_{n\to\infty} \frac{1}{n} H(U_1, \cdots, U_n)$.

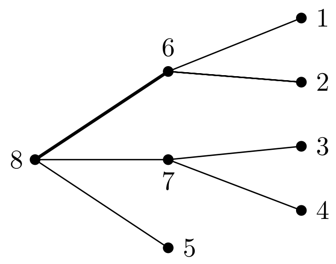
Problem 3. Construct a Tunstall code with $M = 8$ words in the dictionary for a binary memoryless source with $P(0) = 0.9$, $P(1) = 0.1$.

Problem 4. Consider a valid, prefix-free dictionary of words from a source of alphabet size D. Show that the set of lengths $L_1, \dots, L_M$ of the dictionary words satisfy the Kraft inequality

$$\sum_j D^{-L_j} \leq 1$$

with equality. Show that if the dictionary is valid, but not prefix-free, then the Kraft inequality is violated.

Problem 5. Consider a tree with M leaves $n_1, \dots, n_M$ with probabilities $P(n_1), \dots, P(n_M)$. Each intermediate node n of the tree is then assigned a probability $P(n)$ which is equal to the sum of the probabilities of the leaves that descend from it. Label each branch of the tree with the label of the node that is on that end of the branch further away from the root. Let $d(n)$ be a "distance" associated with the branch labelled n. The distance to a leaf is the sum of the branch distances on the path to from root to leaf.



For example, in the tree shown above, nodes 1, 2, 3, 4, 5 are leaves, the probability of node 6 is given by $P(1) + P(2)$, the probability of node 7 by $P(3) + P(4)$, of node 8 (root) by $P(1) + P(2) + P(3) + P(4) + P(5) = 1$. The branch indicated by the heavy line would be labelled 6. The distance to leaf 2 is given by $d(6) + d(2)$.

(a) Show that the expected distance to a leaf is given by $\sum_n P(n)d(n)$ where the sum is over all nodes other than the root. Recall that we showed this in the class for $d(n) = 1$.

(b) Let $Q(n) = P(n)/P(n')$ where n' is the parent of n, and define the entropy of an intermediate node n' as

$$H_{n'} = \sum_{n\colon n \text{ is a child of } n'} -Q_n \log Q_n.$$

Show that the entropy of the leaves

$$H(\text{leaves}) = -\sum_{j=1}^{M} P(n_j) \log P(n_j)$$

is equal to $\sum_{n\in I} P(n)H_n$ where the sum is over all intermediate nodes including the root. Hint: use part (a) with $d(n) = -\log Q(n)$.

(c) Let X be a memoryless source with entropy H. Consider some valid prefix-free dictionary for this source and consider the tree where leaf nodes corresponds to dictionary words. Show that $H_n = H$ for each intermediate node in the tree, and show that

$$H(\text{leaves}) = E[L]H$$

where $E[L]$ is the expected word length of the dictionary. Note that we proved this result in class by a different technique.

PROBLEM 6. 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 $\sum_x q(x)\,\text{length}[C(x)]$.

(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 $\text{length}[C(x)]$ 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.