ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 9	Information Theory and Coding
Homework 5	Oct. 15, 2013

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

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

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

(a) Evaluate

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

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

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

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? Hint: The entropy rate of the process X (if exists) is $\lim_{n\to\infty} \frac{1}{n} H(X_1, \cdots, X_n)$.

PROBLEM 3. Define the *type* $P_{\mathbf{x}}$ (or empirical probability distribution) of a sequence $x_1, \cdots, x_n$ be the relative proportion of occurrences of each symbol of $\mathcal{X}$; i.e., $P_{\mathbf{x}}(a) = N(a|\mathbf{x})/n$ for all $a \in \mathcal{X}$, where $N(a|\mathbf{x})$ is the number of times the symbol a occurs in the sequence $\mathbf{x} \in \mathcal{X}^n$.

(a) Show that if $X_1, \cdots, X_n$ are drawn i.i.d. according to $Q(\mathbf{x})$, the probability of $\mathbf{x}$ depends only on its type and is given by

$$Q^n(\mathbf{x}) = 2^{-n(H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}||Q))}.$$

Hint: Start by showing the following:

$$Q^n(\mathbf{x}) = \prod_{i=1}^n Q(x_i) = \prod_{a\in\mathcal{X}} Q(a)^{N(a|\mathbf{x})} = \prod_{a\in\mathcal{X}} Q(a)^{nP_{\mathbf{x}}(a)}$$

Define the type class $T(P)$ as the set of sequences of length n and type P:

$$T(P) = \{\mathbf{x} \in \mathcal{X}^n : P_{\mathbf{x}} = P\}.$$

For example, if we consider binary alphabet, the type is defined by the number of 1's in the sequence and the size of the type class is therefore $\binom{n}{k}$.

(b) Show for a binary alphabet that

$$|T(P)| \doteq 2^{nH(P)}.$$

We say that $a_n \doteq b_n$, if $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

Hint: Prove that

$$\frac{1}{n+1} 2^{nh_2(\frac{k}{n})} \leq \binom{n}{k} \leq 2^{nh_2(\frac{k}{n})}$$

$h_2(\cdot)$ denotes the binary entropy function. To derive the upper bound, start by proving

$$1 \geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} = \binom{n}{k} 2^{n\left(\frac{k}{n}\log\frac{k}{n} + \frac{n-k}{n}\log\frac{n-k}{n}\right)}$$

To derive the lower bound, start by proving

$$1 = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} \leq (n+1) \max_j \binom{n}{j} p^j (1-p)^j,$$

then take $p = k/n$ and show that the maximum occurs for $j = k$.

(c) Use (a) and (b) to show that

$$Q^n(T(P)) \doteq 2^{-nD(P||Q)}.$$

Note: $D(P||Q)$ is the informational divergence (or Kullback-Leibler divergence) between two probability distributions P and Q on a common alphabet $\mathcal{X}$ and is defined as

$$D(P||Q) = \sum_{a\in\mathcal{X}} P(a) \log \frac{P(a)}{Q(a)}.$$

Recall that we have already seen the non-negativity of this quantity in the class.

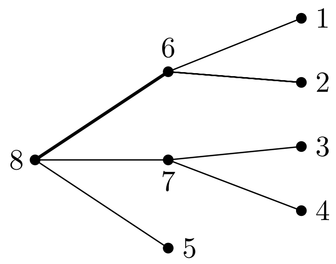
Problem 4. Construct a Tunstall code with $M = 8$ words in the dictionary for a binary memoryless source with $P(0) = 0.9$, $P(1) = 0.1$.

Problem 5. Consider a valid, prefix-free dictionary of words from a source of alphabet size D. Show that the set of lengths $L_1, \ldots, L_M$ 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 6. Consider a tree with M leaves $n_1, \ldots, n_M$ with probabilities $P(n_1), \ldots, P(n_M)$. 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.