

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 8
Homework 4

Information Theory and Coding
Oct. 6, 2015

PROBLEM 1.

- (a) Let U be a random variable taking values in the alphabet \mathcal{U} , and let f be a mapping from \mathcal{U} to \mathcal{V} . Show that $H(f(U)) \leq H(U)$.
- (b) Let U and V be two random variables taking values in the alphabets \mathcal{U} and \mathcal{V} respectively, and let f be a mapping from \mathcal{V} to \mathcal{W} . Show that $H(U|V) \leq H(U|f(V))$.

PROBLEM 2.

- (a) Let U and \hat{U} be two random variables taking values in the same alphabet \mathcal{U} , and let $p_e = \mathbb{P}[U \neq \hat{U}]$. Show that $H(U|\hat{U}) \leq h(p_e) + p_e \log(|\mathcal{U}| - 1)$, where $h(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}$.

Hint: use the random variable $W \in \{0, 1\}$ defined by

$$W = \begin{cases} 1 & \text{if } U \neq \hat{U}, \\ 0 & \text{otherwise.} \end{cases}$$

- (b) Let U and V be two random variables taking values in the alphabets \mathcal{U} and \mathcal{V} respectively, and let f be a mapping from \mathcal{V} to \mathcal{U} . Define $p_e = \mathbb{P}[U \neq f(V)]$. Show that $H(U|V) \leq h(p_e) + p_e \log(|\mathcal{U}| - 1)$.

PROBLEM 3. The entropy $H(U)$ of a random variable U is a function of the distribution p_U of the random variable. Denote by $h(p)$ the entropy of a random variable with distribution p , i.e., $h(p) = \sum_{u \in \mathcal{U}} p(u) \log \frac{1}{p(u)}$. Let p and q be two probability distributions on the same alphabet \mathcal{U} , and, for $\theta \in [0, 1]$ let r be the probability distribution on \mathcal{U} defined by

$$r(u) = \theta p(u) + (1 - \theta) q(u)$$

for every $u \in \mathcal{U}$. We are going to show that

$$H(r) \geq \theta H(p) + (1 - \theta) H(q).$$

- (a) Let U_1 and U_2 be random variables with distributions p and q respectively. Let $Z \in \{1, 2\}$ be a binary random variable with $P(Z = 1) = \theta$. Finally define the random variable U as

$$U = \begin{cases} U_1 & \text{if } Z = 1, \\ U_2 & \text{if } Z = 2. \end{cases}$$

What is the distribution of U ?

- (b) Compute $H(U)$ and $H(U|Z)$. What can you conclude?

PROBLEM 4. Consider a source U with alphabet \mathcal{U} and suppose that we know that the true distribution of U is either P_1 or P_2 . Define $S = \sum_{u \in \mathcal{U}} \max\{P_1(u), P_2(u)\}$.

- (a) Show that $S \leq 2$ and give a necessary and sufficient condition for equality.
- (b) Show that there exists a prefix-free code where the length of the codeword associated to each symbol $u \in \mathcal{U}$ is $l(u) = \left\lceil \log_2 \frac{S}{\max\{P_1(u), P_2(u)\}} \right\rceil$.
- (c) Show that the average length \bar{l} (using the true distribution) of the code constructed in (b) satisfies $H(U) \leq \bar{l} < H(U) + \log S + 1 \leq H(U) + 2$.

Now assume that the true distribution of U is one of k distributions P_1, \dots, P_k .

- (d) Show that there exists a prefix-free code satisfying $H(U) \leq \bar{l} < H(U) + \log_2 S + 1 \leq H(U) + \log_2 k + 1$, where $S = \sum_{u \in \mathcal{U}} \max\{P_1(u), \dots, P_k(u)\}$.

PROBLEM 5. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n pairs of random variables which may or may not be independent. For every $i \geq 1$ and $j \leq n$, define X_i^j to be the sequence X_i, \dots, X_j if $i \leq j$, and to be \emptyset if $i > j$. Define Y_i^j similarly. Therefore, since $X_{n+1}^0 = Y_1^0 = \emptyset$ we have $I(X_{n+1}^n; Y_n) = I(Y_1^0; X_1) = 0$ and $I(Y_1^{n-1}; X_n | X_{n+1}^n) = I(Y_1^{n-1}; X_n)$.

- (a) Show that $I(Y_1^{n-1}; X_n) = \sum_{i=1}^{n-1} I(X_n; Y_i | Y_1^{i-1})$.

- (b) Show that $\sum_{i=1}^n I(X_{i+1}^n; Y_i | Y_1^{i-1}) = \sum_{i=1}^n I(Y_1^{i-1}; X_i | X_{i+1}^n)$.

PROBLEM 6. Recall that s is a prefix of t if t is of the form $t = sv$, the concatenation of s and v for some string v . Similarly we say s is a *suffix* of t if $t = vs$. E.g., the suffixes of “banana” are “a”, “na”, “ana”, “nana”, “anana” and “banana”.

A code \mathcal{C} is said to be a *fix-free code* if and only if no codeword is the prefix or the suffix of any other codeword. Let l_1, \dots, l_k be k integers satisfying $l_1 \leq \dots \leq l_k$. Consider the following algorithm:

```

- Initialize  $A_i = \{0, 1\}^{l_i}$  as the set of available codewords of length  $l_i$  for every  $1 \leq i \leq k$ .
for  $i = 1 \dots k$  do
  if  $A_i \neq \emptyset$  then
    - Pick  $\mathcal{C}(i) \in A_i$ .
    for  $j = i \dots k$  do
      | - (*) Remove from  $A_j$  all the words which start with  $\mathcal{C}(i)$ .
      | - (**) Remove from  $A_j$  all the words which end with  $\mathcal{C}(i)$ .
    end
  else
    | - Algorithm failure.
  end
end
- Return  $\mathcal{C} = \{\mathcal{C}(i) : 1 \leq i \leq k\}$ .

```

- (a) For every $1 \leq i \leq k$ and every $i \leq j \leq k$, show that the number of words in A_j that start with $\mathcal{C}(i)$ is $2^{l_j - l_i}$, and that the number of words in A_j that end with $\mathcal{C}(i)$ is $2^{l_j - l_i}$.
- (b) Show that the number of words that are removed from A_j in (*) and (**) is at most $2^{l_j - l_i + 1}$.

- (c) Show that if $\sum_{i=1}^k 2^{-l_i} \leq \frac{1}{2}$, then the algorithm does not fail.
- (d) Show that if $\sum_{i=1}^k 2^{-l_i} \leq \frac{1}{2}$, then the returned code \mathcal{C} is fix-free and that the lengths of its codewords are l_1, \dots, l_k .
- (e) Let U be a random variable taking values in an alphabet \mathcal{U} . Show that there exists a fix-free code $\mathcal{C} : \mathcal{U} \rightarrow \{0, 1\}^*$ such that $H(U) \leq \mathbb{E}[\text{length}(\mathcal{C}(U))] \leq H(U) + 2$.

PROBLEM 7. Define the *type* $P_{\mathbf{x}}$ (or empirical probability distribution) of a sequence $\mathbf{x} = x_1, \dots, x_n$ 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, \dots, X_n are drawn i.i.d. according to $Q(x)$, the probability of \mathbf{x} depends only on its type and is given by

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

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 $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

- (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.