ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 3 Information Theory and Coding

Homework 2 September 28, 2010, SG1 – 15:15-17:00

Problem 1 (Weak Law of Large Numbers)**.** The weak law of large numbers is used to derive many interesting results in information theory. Here we see how this law is derived.

(a) (Markov's Inequality) If X is a non-negative continuous random variable, show that

$$\Pr\{X \geq a\} \leq \frac{E[X]}{a}, \qquad \forall\, a > 0.$$

[Hint: Split $E[X]$ into two integrals, one from 0 to a, and the other from a to ∞.]

(b) (Chebyshev's Inequality) If Y is a random variable with mean μ and variance σ^2, show using Markov's inequality that

$$\Pr\{|Y-\mu| \geq b\} \leq \frac{\sigma^2}{b^2}, \qquad \forall\, b > 0.$$

[Hint: Set $X = (Y-\mu)^2$ and use Markov's inequality for X.]

(c) (Weak Law of Large Numbers) Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and indentically distributed (IID) random variables with mean μ and variance σ^2. Let the sample mean be $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Show that

$$\Pr\{|\overline{X}_n - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2}, \qquad \forall\, \epsilon > 0.$$

This means that the probability that the sample mean differs from the actual mean by more than some given number ϵ goes to zero as n tends to infinity. This is the weak law of large numbers.

Problem 2 (Jensen's Inequality)**.** A function $f(x)$ is said to be convex over an interval (a,b) if for every $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

If f is a convex function and X is a discrete random variable that takes values $x_1, x_2, \cdots, x_n$ with probabilities $p_1, p_2, \cdots, p_n$, such that $\sum_{i=1}^n p_i = 1$, then show that

$$E[f(X)] \geq f(E[X]).$$

This is known as Jensen's inequality and is important in deriving many of the inequalities we will encounter in information theory. What is the condition for equality? [Hint: When $n = 2$, $E[f(X)] = p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 f(x_1) + p_2 f(x_2)) = f(E[X])$, because f is convex. Use induction.]

Problem 3 (Huffman Coding)**.** The joint probability mass distribution of random variables X and Y is given below:

$Y \setminus X$	x_1	x_2	x_3	x_4	x_5	x_6
y_1	$1/4$	$1/4$	$1/16$	$1/16$	$1/16$	$1/16$
y_2	$1/24$	$1/24$	$1/24$	$1/24$	$1/24$	$1/24$

(i) Given $Y = y_1$, what is the expected word length W_1 of a binary Huffman code for X?

(ii) Given $Y = y_2$, what is the expected word length W_2 of a binary Huffman code for X?

(iii) When Y is random, what is the expected word length W of a binary Huffman code for X? Is $W = \Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2$?

(iv) Is (iii) true in general, that is, given two random variables X and Y that take values in $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_m)$ respectively and their joint probability mass function, if W_i is defined as the expected word length of a binary Huffman code for X given $Y = y_i$ for $i = 1, \cdots, n$, and W is the expected word length of a binary Huffman code for X when Y is random, then is $W = \sum_{i=1}^{n} \Pr\{Y = y_i\}W_i$?

(v) What is the expected word length of a binary Huffman code for the pair (X, Y)?

Problem 4 (Code Mismatch)**.** A Shannon code is a code that assigns a codeword of length l_i, such that $\log(1/q_i) \le l_i \le \log(1/q_i) + 1$, to a symbol with probability q_i. Suppose we have such a code C_q for the probability distribution q_i, $i = 1, \cdots, n$. If however the actual input has a distribution p_i, $i = 1, \cdots, n$, but we use the code C_q for encoding anyway, then show that the expected Shannon codeword length L will satisfy

$$H(p) + D(p||q) \le L \le H(p) + D(p||q) + 1,$$

where the entropy $H(p) := -\sum_{i=1}^{n} p_i \log p_i$, and the Kullback-Leibler divergence $D(p||q) := \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$.