INFORMATION THEORY
Master of Logic, University of Amsterdam, 2016
TEACHER: Christian Schaffner, TA: Yfke Dulek

Practice problem set 2

This week's exercises deal with ...

You do not have to hand in these exercises, they are for practicing only. Problems marked with a ★ are generally a bit harder. If you have questions about any of the exercises, please post them in the discussion forum on Moodle, and try to help each other. We will also keep an eye on the forum.

Problem 1: Deriving the weak law of large numbers

(homework)

(a) **(3pt)** (Markov's inequality) For any real non-negative random variable X and any $t > 0$, show that

$$P_X(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Exhibit a random variable (which can depend on t) that achieves this inequality with equality.

(b) **(2pt)** (Chebyshev's inequality.) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y - \mu)^2$, show that for any $\varepsilon > 0$,

$$P(|Y - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

(c) **(2pt)** (The weak law of large numbers.) Let $Z_1, Z_2, ..., Z_n$ real i.i.d. random variables with mean $\mu = \mathbb{E}[Z_i]$ and variance $\sigma^2 = \mathbb{E}[(X_i - \mu)^2] < \infty$. Define the random variables $S_n = \frac{1}{n}\sum_{i=1}^n Z_i$. Show that

$$P(|S_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Thus, $P(|S_n - \mu| > \varepsilon) \to 0$ as $n \to \infty$. This is known as the weak law of large numbers (Theorem 2.6.1 in the lecture notes).

Problem 2: Maximal conditional entropy implies independence

(remove)

(a) Prove that if $H(X|Y) = \log(|\mathcal{X}|)$ then X and Y are independent.

(b) Give a joint distribution P_{XY} where $H(X) = \log(|\mathcal{X}|)$, but X and Y are dependent.

Problem 3: Inefficiency when using the wrong code

(homework with hints, see ex. 6 of 2015) Prove that when designing a code with length $\ell(X)$, believing that the distribution is Q_X when the true distribution is P_X incurs a penalty of $D(P_X||Q_X)$ in the average description length. More formally, prove that

$$H(P_X) + D(P_X||Q_X) \leq \mathbb{E}_{P_X}[\ell(X)] \leq H(P_X) + D(P_X||Q_X) + 1.$$

Problem 4: Huffman Coding

(a) **(4pt)** For a binary source P_X with $P_X(0) = \frac{1}{8}$ and $P_X(1) = \frac{7}{8}$, design a Huffman code for blocks of $N = 1, 2$ and 3 bits. For each of the three codes, compute the average codeword length and divide it by N, in order to compare it to the optimal length, i.e. the entropy of the source. What do you observe?

(b) **(1pt)** If you were asked in (a) to design a Huffman code for a block of $N = 100$ bits, what problem would you run into?

(c) **(2pt)** Consider the random variable Z with

z	1	2	3	4	5	6
$P_Z(z)$	1/10	3/10	2/10	2/10	1/10	1/10

Find an optimal *ternary* Huffman encoding for Z (i.e. using an alphabet with three symbols).

Problem 5: Kraft's inequality

Below, six binary codes are shown for the source symbols $x_1, ..., x_4$.

	Code A	Code B	Code C	Code D	Code E	Code F
x_1	00	0	0	0	1	1
x_2	01	10	11	100	01	10
x_3	10	11	100	110	001	100
x_4	11	110	110	111	0001	1000

(a) **(1pt)** Which codes fulfill the Kraft inequality?

(b) **(1pt)** Is a code that satisfies this inequality always uniquely decodable?

(c) **(1pt)** Which codes are prefix-free?

(d) **(1pt)** Which codes are uniquely decodable?

Problem 6

Something with binary entropy

Problem 7: An optimal code

(Practice) (MacKay 5.25) Let X be a random variable. Show that if for all $x \in \mathcal{X}$, there is some $n \in \mathbb{N}$ such that $P_X(x) = \frac{1}{2^n}$, then there exists a source code whose expected length equals the entropy.

Problem 8: Geometric distribution

(practice) Compute the entropy of the geometric distribution. (do with trees)

Problem 9: Mackay, example 2.13

(practice) see homework 1 from 2014