Information Theory Exercise Sheet #2 (Data Compression)

University of Amsterdam, Master of Logic, Fall 2015
Lecturer: Christian Schaffner
TA: Mathias Madsen

Out: Wednesday, 4 November 2015
(due: Wednesday, 11 November 2015, 9:00)

1. **Toy Language** A random variable X is distributed according to the point probabilities in the following table:

x	a	b	c	d	e
$P_X(x)$	$1/3$	$1/4$	$1/6$	$1/6$	$1/12$

   (a) Compute $H(X)$.

   (b) Construct a Huffman code for the variable, and compute its expected code length.

   (c) Decode the message `00101100001` according to your code.

2. **Dice Throws** Let X be a random variable distributed uniformly on the set $\{1, 2, 3, 4, 5, 6\}$.

   (a) Construct a Huffman code for the variable.

   (b) What is the average codeword length for your code? How does that compare with the entropy?

   (c) If you interpret a codeword of length k as a probability of 2^{-k}, what is then the implicit distribution expressed by your code?

3. **Maximum Equivocation** Suppose X and Y are random variables ranging over n different values.

   (a) Prove that if $H(X \mid Y) = \log n$, then X and Y are independent.

   (b) Provide an example showing that X and Y need not be independent even though $H(X) = \log n$.

4. **Typical Sequences** Suppose we have a bent coin with $P_X(1) = 1/3$. We flip this coin five times, producing a random sequence $S = (X_1, \dots, X_5)$.

   (a) Compute the entropy $H(X)$ of each individual coin flip, and the entropy $H(S)$ of the sequence.

(b) Draw up a table of the surprisal values $-\log P_S(s)$ and their probabilities.

(c) Compute the probability that S is typical at level $\varepsilon = 0.1$, that is,

$$P\left(\left|\frac{1}{5}\log\frac{1}{P_S(S)} \;-\; H(X)\right| < 0.1\right).$$

5. **High-Probability Sets (CT, Ex. 3.5)** Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with a common entropy $H(X)$. Let P_S be the probability distribution of the sequence $S = (X_1, \ldots, X_n)$, and define the set $C(\tau)$ as

$$C(\tau) \;:=\; \left\{s : P_S(s) \geq 2^{-n\tau}\right\}.$$

(a) For a fixed τ, what's the largest number of tuples such a set can contain?

(b) Sketch a graph of the probability

$$P(S \in C(\tau)),$$

as a function of τ, both for large n, and for extremely large n.

6. **Error Penalty** Suppose that an engineer believes that a source X can be described by the distribution Q_X given by the following table:

x	a	b	c
$Q_X(x)$	1/2	1/4	1/4

In fact, however, the source follows the distribution P_X:

x	a	b	c
$P_X(x)$	1/4	1/2	1/4

(a) Design a code for X based on the wrong distribution Q_X.

(b) Design a code for X based on the correct distribution P_X.

(c) Compute the expected number of bits per symbol used by each of these codes when X is sampled from P_X. How big is the difference?

(d) Explain how this number relates to the Kullback-Leibler divergence

$$D(P_X \,||\, Q_X) \;=\; \sum_x P_X(x)\log\frac{P_X(x)}{Q_X(x)}.$$

7. **Sparse File (CT, Ex. 5.7)** A source creates a black and white image by sampling 100 independent pixels $X_1, \ldots, X_{100}$ from a distribution with

$$P_{X_i}(0) \;=\; 0.995$$
$$P_{X_i}(1) \;=\; 0.005$$

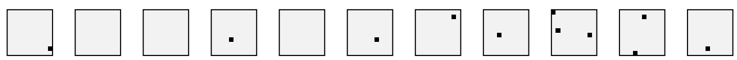
You decide to brute-force encode these images by means of a table of equally long codewords. You assign a codeword to any image containing three or fewer black pixels (three or fewer 1s), and accept that there will be an error in the remaining cases.

(a) Compute the probability of encountering an untabulated sequence.

(b) Compute the number of codewords needed for this code, and the number of bits required to express that many codewords.

(c) How does that number compare to the theoretical minimum?

(d) What are your options for improving this performance, theoretically and practically?

 $\cdots$

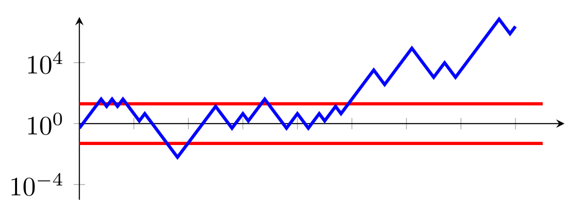
8. **Sequential Analysis*** Two scientists try to predict a binary sequence $X_1, \ldots, X_n$. Both model the sequence as an i.i.d. sample from a Bernoulli distribution, but they disagree which parameter to use:

$$P_X^1(1) = 0.6 \qquad \text{vs.} \qquad P_X^2(1) = 0.2.$$

In fact, the data is sampled from a fair coin flipping sequence, $P_X^*(1) = 0.5$. We measure the relative performance of the two scientists by looking at the likelihood ratio

$$\frac{P_X^1(X_1) \cdot P_X^1(X_2) \cdots P_X^1(X_n)}{P_X^2(X_1) \cdot P_X^2(X_2) \cdots P_X^2(X_n)},$$

which is itself a random variable when we sample $X_i \sim P_X^*$, $i = 1, \ldots, n$. We consider one scientist better than the other if this ratio exceeds 20/1 or drops below 1/20. Roughly how many coin flips will it take before this happens?



Homework is exercises 1, 4, and 7.