Information Theory

Exercise Sheet 2

Exercise 1 (AEP and mutual information)**.** *Let* $(\mathbf{X}_i, \mathbf{Y}_i)_i$ *be i.i.d. and distributed as* $p(x, y)$*. We form the log likelihood ratio of the hypothesis that* $\mathbf{X}_i$ *and* $\mathbf{Y}_i$ *are independent versus the hypothesis that* $\mathbf{X}_i$ *and* $\mathbf{Y}_i$ *are dependent. What is the limit of*

$$\frac{1}{n} \log \frac{p(\mathbf{X}_1), \ldots, p(\mathbf{X}_n) p(\mathbf{Y}_1), \ldots, p(\mathbf{Y}_n)}{p(\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{Y}_1, \ldots, \mathbf{Y}_n)}$$

Exercise 2 (Piece of cake)**.** *A cake is sliced roughly in half, the largest piece being chosen each time, the other pieces discarded. We will assume that a random cut creates pieces of proportions*

$$P = \begin{cases} \left(\frac{2}{3}, \frac{1}{3}\right) & \textit{with probability } \frac{3}{4} \\ \left(\frac{2}{5}, \frac{3}{5}\right) & \textit{with probability } \frac{1}{4} \end{cases}$$

Thus, for example, the first cut (and choice of largest piece) may result in a piece of size $\frac{3}{5}$*. Cutting and choosing from this piece might reduce it to size* $\frac{3}{5} \cdot \frac{2}{3}$*, and so on. How large, to first order in exponent, is the piece of cake after n cuts?*

Exercise 3 (Guess the integer, the game sixty-three)**.** *What's the smallest number (don't prove it, be informal) of yes/no questions needed to identify a random integer* x *between* 0 *and* $2^6 - 1$*? What do you conclude?*

Exercise 4 (Entropy versus optimal strategy)**.** *You are given* 12 *balls, all equal in weight except for one that is either heavier or lighter. You are also given a two-pan balance to use. In each use of the balance you may put any number of the* 12 *balls on the left pan, and the same number on the right pan, and push a button to initiate the weighing; there are three possible outcomes: either the weights are equal, or the balls on the left are heavier, or the balls on the left are lighter.*

Your task is to design a strategy to determine which is the odd ball and whether it is heavier or lighter than the others in as few uses of the balance as possible.

While thinking about this problem, you may find it helpful to consider the following questions:

(*a*) *When you have identified the odd ball and whether it is heavy or light, how much information have you gained?*

(*b*) *Once you have designed a strategy, draw a tree showing, for each of the possible outcomes of a weighing, what weighing you perform next. At each node in the tree, how much information have the outcomes so far given you, and how much information remains to be gained?*

(*c*) *How much information is gained when you learn (i) the state of a flipped coin; (ii) the states of two flipped coins; (iii) the outcome when a four-sided die is rolled?*

(*d*) *How much information is gained on the first step of the weighing problem if* 6 *balls are weighed against the other* 6*? How much is gained if* 4 *are weighed against* 4 *on the first step, leaving out* 4 *balls?*

1. *Is it possible to answer the question with two weighings?*

2. *Give a strategy to answer the question in three weighings.*

Exercise 5 (Huffman questions)**.** *Consider a random number* $\mathbf{X}$ *between* 1 *and* n*. Let* $p_i \overset{def}{=} \mathbb{P}(\mathbf{X} = i)$*. We are asked to determine the value of* $\mathbf{X}$ *as quickly as possible by asking questions. Any yes-no question you can think of is admissible.*

1. *Give a maximum lower bound on the minimum average number of required questions.*

2. *Give an upper bound on the minimum average number of required questions.*

3. *Find an optimal set of questions in the following case,*

p_1	p_2	p_3	p_4	p_5	p_6
$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Exercise 6 (Kraft inequality)**.** *Prove the following,*

1. *There exists a prefix code with codewords of length* $n_1, \ldots, n_K$ *if and only if*

$$\sum_{k=1}^{K} \frac{1}{2^{n_i}} \leq 1$$

2. *There exists a uniquely decodable code with codewords of length $n_1, \dots, n_K$ if and only if*

$$\sum_{k=1}^{K} \frac{1}{2^{n_i}} \leq 1$$

Exercise 7 (The cost of using the wrong codelengths to compress)**.** *Recall that Shannon's theorem states that for any distribution $\mathbf{X} : \Omega \to \mathcal{X}$, it exists a prefix code φ with expected length satisfying*

$$H(\mathbf{X}) \leq L(\varphi, \mathcal{X}) < H(\mathbf{X}) + 1$$

The crucial idea to prove this theorem is to use the following code assignment $\ell(x) \overset{def}{=} \lceil \log_2 \frac{1}{p(x)} \rceil$. Instead, suppose that we use

$$\ell'(x) \overset{def}{=} \left\lceil \log_2 \frac{1}{q(x)} \right\rceil$$

for another distribution $q(x) \overset{def}{=} \mathbb{P}(\mathbf{Y} = x)$. Show that

$$H(\mathbf{X}) + D_{\mathrm{KL}}(\mathbf{X}||\mathbf{Y}) \leq L(\mathbf{Y}) \leq H(\mathbf{X}) + D_{\mathrm{KL}}(\mathbf{X}||\mathbf{Y}) + 1$$

where $L(\mathbf{Y})$ denote the average length compression, i.e.,

$$L(\mathbf{Y}) = \sum_{x \in \mathcal{X}} \ell'(x) p(x)$$

Exercise 8 (Some Huffman codes)**.** *Given a random variable $\mathbf{X}$, then $\mathbf{X}^{\otimes i}$ denotes the distribution $(\mathbf{X}_1, \dots, \mathbf{X}_i)$ where the $\mathbf{X}_i$'s are i.i.d with the same distribution than $\mathbf{X}$.*

Make Huffman codes for $\mathbf{X}^{\otimes 2}$ and $\mathbf{X}^{\otimes 3}$ where $\mathcal{X} = \{0, 1\}$ and

$$\mathbb{P}(\mathbf{X} = 0) = \frac{9}{10}$$

Compute their expected lengths $L(\{0,1\}, \varphi)$'s and compare them with the entropies $H(\mathbf{X}^{\otimes 2})$ and $H(\mathbf{X}^{\otimes 3})$ as well as the $L(\{0,1\}, \varphi)/H(\mathbf{X}^{\otimes i})$'s.

Exercise 9 (Huffman codes are optimal prefix symbol codes)**.** *Our aim is to show that Huffman coding φ_{H} is an optimal prefix code. By optimal we mean that there are no uniquely prefix code φ such that,*

$$L(\mathcal{X}, \varphi) < L(\mathcal{X}, \varphi_{\mathrm{H}})$$

1. *Show that there exists an optimal prefix code such that*
   (a) *If $p(x_j) > p(x_k)$, then $\ell(x_j) \leq \ell(x_k)$*
   (b) *The two longest codewords have the same length and correspond to the least likely symbols*
   (c) *These two codewords differ only in their last bit*
2. *Deduce that Huffman coding is an optimal prefix code.*

Hint: *use an induction on the size of the source alphabet $\mathcal{X}$*

Exercise 10 (How many randomness to generate a distribution?)**.**

1. *Consider the following distribution,*

x_i	$p(x_i)$
a	$\frac{1}{2}$
b	$\frac{1}{4}$
c	$\frac{1}{4}$

and a fair coin $\mathbf{X}$, i.e., $\mathbb{P}(\mathbf{X} = 1) = \mathbb{P}(\mathbf{X} = 0) = \frac{1}{2}$. Give strategies to generate letters according to the above distribution with in average

(*i*) 2 *uses of* $\mathbf{X}$,
(*ii*) 1.5 *uses of* $\mathbf{X}$.

Compute the entropy of the distribution over a, b, c*. What do you conclude? How do you interpret this result?*

Our aim now is to solve the following problem. We are given a distribution $(p(x))_{x \in \mathcal{X}}$ and sequences of i.i.d. random variables $(\mathbf{Z}_i)_{i \geq 0}$ where

$$\mathbb{P}(\mathbf{Z}_i = 1) = \mathbb{P}(\mathbf{Z}_i = 0) = \frac{1}{2}.$$

Our goal is to draw $x \in \mathcal{X}$ according to $p(x)$ by using the $\mathbf{Z}_i$'s.

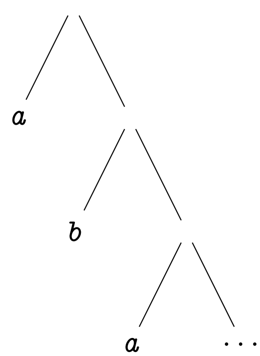
Drawing $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_m, \ldots$ can be viewed as a walk in a binary tree where the leaves have label $x \in \mathcal{X}$. When a leaf is encountered, output the corresponding letter.

The problem is to construct the optimal tree such that the average walk length is the smallest possible and such that the leaves are output with the right probability.

2. *Build the trees according to the strategies you gave in Question* 1.

3. *Let $\mathcal{X} = \{\mathtt{a}, \mathtt{b}\}$ with $p(a) = 2/3$ and $p(b) = 1/3$. Show that the following infinite tree enables to generate this distribution. What is the average number of coin flips enabling to generate $\mathbf{X}$? Compute $H(\mathbf{X})$. What do you conclude?*

4. *Let $\mathcal{X} = \{\mathtt{a}, \mathtt{b}\}$ with $p(a) = 2/3$ and $p(b) = 1/3$. Show that the following infinite tree enables to generate this distribution. What is the average number of coin flips enabling to generate $\mathbf{X}$? Compute $H(\mathbf{X})$. What do you conclude?*



Show that any algorithm generating $\mathbf{X}$ from fair bits, the expected number of fair bits used is greater than $H(\mathbf{X})$

5. *Let $\mathbf{X}$ be a random variable with dyadic distribution, i.e., $\mathbb{P}(\mathbf{X} = \boldsymbol{x}) = 2^{\ell_{\boldsymbol{x}}}$ where $\ell_{\boldsymbol{x}} \in \mathbb{N}$. Show that there exists a way to toss fair coins such that the average number of coin tosses is equal to $H(\mathbf{X})$.*

6. *Sow that the average number of coin tosses $\overline{T}$ to simulate a random variable $\mathbf{X}$ with associated distribution $(p_i)_i$ satisfies,*

$$H(\mathbf{X}) \leq \overline{T} \leq H(\mathbf{X}) + 2$$

Hint: *use the binary expansions of the probabilities.*