«Calcul formel avancé et application». Very brief lecture notes.

14.09.2023. Lecture 1.

1. The game guess a number : one players chooses an integer number between 1 and n, another player should find this number by asking questions with answers *yes* or *no*. There is a simple strategy that allows to find the chosen number in $\lceil \log n \rceil$ questions (bisection). Moreover, there is a non-adaptive strategy with the same number of questions (the second player asks bits of the binary expansion of the chosen number ; all questions are formulated in advance, before the first response).

These strategies are optimal : no strategy helps to reveal the chosen number in less than $\lceil \log n \rceil$ questions (in the worst case). Indeed, every guessing strategy can be represented as a binary tree (with questions in the internal nodes and the guessed numbers in a leave). Since such a tree must have at least n leaves (one leave for each possible answer), the depth of the tree must be at least $\lceil \log n \rceil$.

Exercise 1.1 (difficult)**.** How many questions do we need to guess an integer number between 1 and 100 (asking questions with answers *yes* or *no*) if one of the answers may be false.

2. Sorting algorithms. We are given n objects ("stones") and balance scales ; in one operation we can compare weights of two stones. To sort n stones by their weights, we have to do in the worst case $\log(n!)$ pairwise comparisons (no algorithm can guarantee the right answer with less than $\log(n!)$ weighings).

Exercise 1.2 (Sorting algorithms)**.** Find the number of comparisons needed in the worst case
(a) to sort an array of size 4 ;
(b) to sort an array of size 5.

3. Shannon's entropy. For a random variable α with n possible values $a_1, \ldots, a_n$ such that for $i = 1 \ldots n$ $\text{Prob}[\alpha = a_i] = p_i$, we define its Shannon's entropy as

$$H(\alpha) := \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$$

(with the usual convention $0 \cdot \log \frac{1}{0} = 0$).Reminder :

Proposition 1. *For every random variable α distributed on a set of n values*

$$0 \le H(\alpha) \le \log n.$$

Moreover, $H(\alpha) = 0$ if and only if the distribution is concentrated at one point (one probability p_i is equal to 1, and the other p_j for $j \ne i$ are equal to 0), and $H(\alpha) = \log n$ if and only if the distribution is uniform ($p_1 = \ldots = p_n = \frac{1}{n}$).

4. The game guess a number revisited : one players chooses an integer number between 1 and n with known probabilities $p_1, \ldots, p_n$, another player should find this number by asking questions with answers *yes* or *no*. We need to estimate the *average* number of questions needed to identify the number.

In the class we discussed a bisection strategy that requires *on average* $\approx \sum\limits_{i=1}^{n} p_i \log \frac{1}{p_i}$ questions (but a more precise statement and a formal proof is postponed). In the class we proved only the lower bound :

Proposition 2. Every strategy of guessing a number with yes-or-no questions requires *at least* $\sum\limits_{i=1}^{n} p_i \log \frac{1}{p_i}$ on average.

Idea of the proof : concavity of the logarithm and Jensen's inequality (more details in the class).