Multi-armed bandits strategies
PROJECT SUMMARY

The goal of the project is to compare the UCB algorithm to the following strategy, for instance for $K = \{1, 2\}$ (only two possible actions), for a given horizon $T = 4T'$:

- — from $t = 1$ to $t = T'$: choose action 1. Compute $C_1 = \sum_{t=1}^{T'} X_{1,t}$
- — from $t = T' + 1$ to $t = 2T'$: choose action 2. Compute $C_2 = \sum_{t=T'+1}^{2T'} X_{2,t}$
- — choose k such that action k minimizes C_k and use it systematically between $2T' + 1$ et T

The implementation has moreover generalized the former to the case of n arms, and UCB regret bounds have also been computed :

$$R_T \leq 8 \sum_{k:\Delta_k>0} \frac{\log(T)}{\Delta_k} + K\frac{\pi^2}{3}$$

$$R_T \leq \sqrt{KT\left(8\log(T) + \frac{\pi^2}{3}\right)}$$

Eventually, regret

$$R_t = \mathbb{E}\left[\sum_{j=1}^{t} X_{k(t),j}\right] - \min_{1\ldots K} \mathbb{E}\left[\sum_{j=1}^{t} X_{k,j}\right]$$

has been plotted for the different strategies in the case of i.i.d data, and strategies have been challenged experimentally in the case of non i.i.d data (arms distributions switching at half horizon).