Lab 1. EE – Multi-armed bandit

Multi-armed bandit (MAB) framework has attracted a lot of attention in various applications, from recommender systems and information retrieval to healthcare and finance, due to its stellar performance combined with certain attractive properties, such as learning from less feedback. The multi-armed bandit field is currently flourishing, as novel problem settings and algorithms motivated by various practical applications are being introduced.

If you are not familiar with Google Colab, follow first this tutorial.

1. Copy the Python source code provided in Atenea into a Google Drive folder and open it in Google Colab. Read the code in the notebook, run it, and analyze the results.
2. Use the Python code provided as a basis to include your own code and solve the next questions:

   a) For ε-greedy algorithm, check the reward on a single run for smaller values of the variance of the Gaussian $f(r|a)$ in example 2.1 in slides. Derive conclusions.
   b) Think of a practical application that can be modeled with an m-armed bandit. Use a sensible $f(r|a)$ $\forall a$ (Gaussian, binary Bernoulli, exponential, etc.) for that application. You may get inspiration from this paper. Assume **stationarity** over time.
   c) Program it using the base code provided. Note that Gaussian rewards have been assumed, you'll have to do the proper modifications.
   d) Program the UCB technique. Check the average reward obtained in convergence for several values of the parameter. Check also the evolution of the number of correct decisions.
   e) Plot the estimated value of $Q(a)$ for every action on a single run and comment.
   f) Plot the evolution of regret for ε-greedy and UCB algorithms and comment.
   g) Program another MAB procedure among those reviewed in the theory lectures (gradient bandit, probability matching).
   h) Make the environment in b) **non-stationary**, and include the appropriate changes to allow the algorithms track the best decisions.
   i) Is the capacity of exploration of UCB affected in the non-stationary case? Why? Propose a modification that performs better in these conditions.

3. Prepare a report in Colab containing a description and interpretation of results.
4. Hand in a link to your report in Atenea.