PROGRAMMING ASSIGNMENT #2
JOHN C.S. LUI
DEADLINE: NOVEMBER 4, 2024. 11:59 PM.

In class, we have discussed (a) the random arm selection policy, and (b) the Explore-then-Exploit (ETE) selection strategy. (c) the epsilon-greedy algorithm (both basic and the adaptive version with ϵ_t) and (d) the upper confidence bound (UCB) algorithm. The objectives of this assignment are:

- Make sure students understand the above arm selection policies from the programming perspective.
- Experiment with simulation and performance evaluation of the above arm selection policies.

Assume you have $N = 10$ arms, arm i, where $i \in [N]$, will have a reward distribution over $[0, 10]$ and they can have different uniform probability distributions over $[0, 10]$. For example, arm 1 may be uniformly distributed in $[0, 5]$ while arm 2 may be uniformly distributed in $[2, 7]$. In short, you can specify the uniform distribution of reward for all these 10 arms. In this programming assignment, you have to do the following:

- For each arm, specify their *average reward* based on your given distributions on these 10 arms.
- Assume T, the total number of time slots, is a given input parameter. You need to:
  - Calculate the *average cumulative reward* and the *average regret* for the random arm selection policy for $T = 10,000$.
  - Calculate the *average cumulative reward* and *average regret* for the ETE policy when $m = 1, 10, 20, 30, 40$ and 50 for $T = 10,000$.
  - Calculate the *average cumulative reward* and the *average regret* for the basic epsilon greedy policy for $T = 10,000$ when $\epsilon = 0.1, 0.2, 0.3, 0.4$ and 0.5.
  - Calculate the *average cumulative reward* and the *average regret* for the adaptive epsilon greedy arm selection policy for $T = 10,000$ when ϵ is a function of time t (as stated in the lecture).
  - Calculate the *average cumulative reward* and *average regret* for the UCB algorithm.
  - Plot the average cumulative reward vs. T for the above cases in a single figure.
  - Plot the average regret vs. T for the above cases in a single figure.
  - Repeat all questions above when we set $T = 1,000$ and $T = 1,500$ and $T = 2,000$.

When you submit your assignment, you need to include all the following in a **SINGLE ZIPFILE**.

- Your source code
- Instruction on how to compile and run your code in a file "README.txt"
- A report to indicate the average cumulative reward and average regret for each case list above.
- Plots for the average cumulative reward and average regret for each case above.

Just upload your zipfile on the blackboard on or before the deadline.