Probability and Mathematical Statistics: Final Project

Due on December 31, 2022 at 11:59am

Name: **Ren Hui, Wanchen Su**
Student ID: xxxxxxx, 2021533067

Problem 1 (Part 1: classical Bandit Algorithms)

Solution

1,2. See jupyter notebook.

3. Accorrding to python simulation, the results are repectively:
   results for epsilon-greedy Algorithm:
   [0.69956534 0.50167567 0.39846044] with parameter: 0.1
   [0.70064482 0.49688743 0.39858856] with parameter: 0.5
   [0.70009067 0.49985083 0.39811366] with parameter: 0.9
   results for UCB Algorithm:
   [0.70088294 0.49274932 0.38448981] with parameter: 1
   [0.7006747 0.49900798 0.39886189] with parameter: 5
   [0.70135645 0.49877856 0.40023188] with parameter: 10
   results for TS Algorithm:
   [0.6995964 0.45887541 0.37554019] with parameter: [[1, 1], [1, 1], [1, 1]]
   [0.68296885 0.4001996 0.362563] with parameter: [[601, 401], [401, 601], [2, 3]]

4. In the ϵ-greedy algorithm, ϵ decides the probability of choosing the max estimate of all the arms or choosing a random arm. Which means the larger ϵ is, the more evenly spread are the tests.
   In the UCB algorithm, the parameter c balances the estimation of θ_j, and considers the number of trials on that arm. Increasing c means considering more about the number of test on that arm, which will alow the decision maker choose some of the less experimented arms. This helps in instances where the number of tests on a arm is quite small and therefore the θ of that arm is way less than it should be.
   If the initial value of α_j, β_j we passed in is too small, then the result of the first few tests may influence the final result greatly, and only a few dozen tests are on the second and third arm. If this happens, the result will be frightfully small. And after 200 independent trials of $N = 5000$, the final estimation still has a great gap with the oracle value. On the other hand, if the initial value of α_j, β_j we passed in is relativly large, this value of a and b may influence the final result greatly. As in the trial with parameter [[601, 401], [401, 601], [2, 3]], we can see that the final estimation of arm two is very close to $\frac{401}{401+601}$. More impacts of ϵ, c and α_j, β_j concerning exploration-exploitation will be disccussed in 5.

5. In the case of all algorithms, more exploitation means that the total reward we get is larger, but if the estimate of each arm differs too much from the oracle value, after all tests, the total aggregated reward becomes smaller. More exploration means the estimated value of the reward of each arm is more accurate, it helps the decision maker choose the arm better, but the exploration means that the rewards during exploration is less.
   For the ϵ-greedy algorithm, the larger ϵ is, there is more chance of exploration than exploitation. So while ϵ grows, the gaps between the algorithm estimate and the oracle value decreases, but the sum of the reward of all trials is smaller because the DM spent too much time exploring when it is not that necessary.
   As for the UCB Algorithm, the exploration-exploitation trade-off depends on the value of c. Similar to the ϵ-greedy algorithm, the larger c is, there is more exploration and less exploitation. As can be deduced from the data in the python simulation, for larger c, the estimated value for the third arm is more accurate. This means that there are more times when the decision maker choose arm c. Consequently, the aggregated rewards over N time slots are less.
   The TS Algorithm is rather different from the previous ones. Its exploration is very limited. The arm selected is always the arm with the greatest current reward estimation. So arms with small oracle values may end up being pulled only a few times. Which makes the estimate value of arms 2 and 3

rather inaccurate. But the aggregated reward gained over the N trials should be the largest of all the algorithms.

6.