TP2 : Bandits

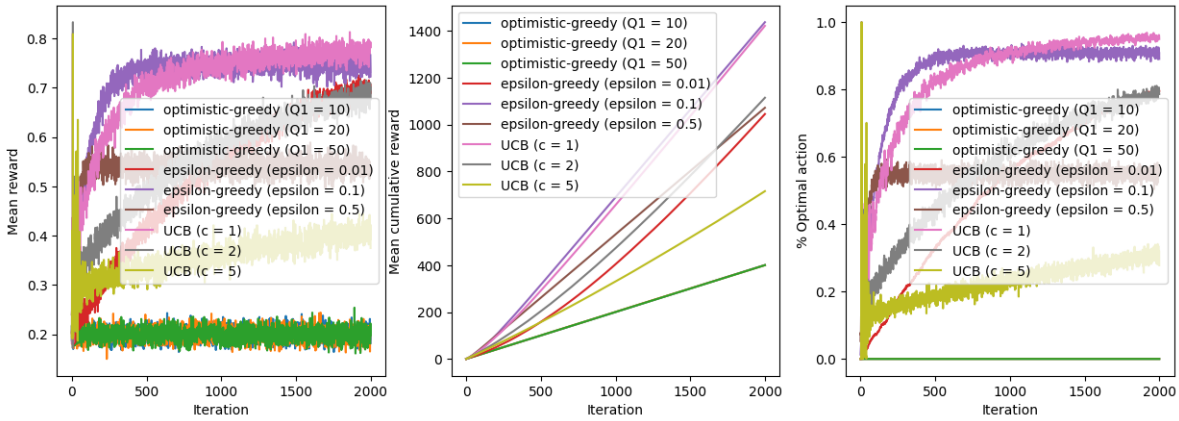
Jed Moutahir

October 2023

1 Basic Algorithms

1.1 Question 1

Plot the corresponding mean reward; the mean cumulative reward and the percentage of times the best arm was elected as time goes by. Play with the parameters (and c, and T) and study their effects.



1.2 Question 2

With ϵ-greedy, what is the asymptotic probability of taking the optimal action?

The asymptotic probability of taking the optimal action for ϵ-greedy algorithm is $\frac{1-\epsilon}{\epsilon/10}$.

1.3 Question 3

Which ϵ is better for a relatively small of T? and for large T

For a small T, the ϵ should be chosen small.
For a large T, the ϵ should be chosen large.

In the previous example for T = 2000, the optimal ϵ is around 0.1.

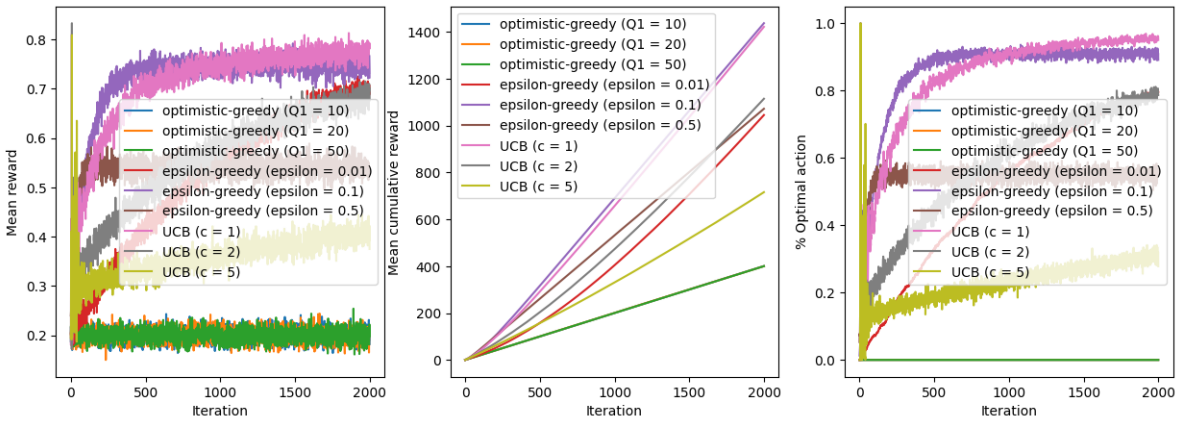
1.4 Question 4

Do you observe some spikes in the plot of average rewards? if yes, please provide an explanation.

The spikes observed are due to lucky initial guesses. They tend to disappear as the number of iterations increases for UCB since the algorithm explores more and become more confident about the rewards of each action.

1.5 Question 5

Plot the corresponding mean reward; the mean cumulative reward and the percentage of times the best arm was elected as time goes by. Interpret



Unfortunately the optimistic-greedy method is not the best one. Indeed, it is not able to explore the other arms and therefore it is not able to find the best arm. It uses the same arm all the time and randomly used the wrong one at first and now is stuck with it.

The epsilon-greedy method is better than the greedy one because it is able to explore the other arms. However, depending on the value of epsilon, it might take too long to find the best arm.

The UCB method is the best one because it is able to explore the other arms and exploit the best arm. Indeed, it is able to find the best arm and when it is confident enough, it will use it all the time. However, it is not perfect because it is not able to find the best arm at first. It needs to explore a bit before being able to exploit the best arm. This, is also why choosing the right value for c is important dependong on T.

1.6 Question 6

What are your conclusions in terms of methods? Give some intuition.

The UCB method is the best one because it is able to explore the other arms and exploit the best arm. Indeed, it is able to find the best arm and when it is confident enough, it will use it all the time. However, it is not perfect because it is not able to find the best arm at first. It needs to explore a bit before being able to exploit the best arm. This, is also why choosing the right value for c is important depending on T.

2 Gradient method

2.1 Question 1

Which method does work better, with or without baseline?