Assignment 2

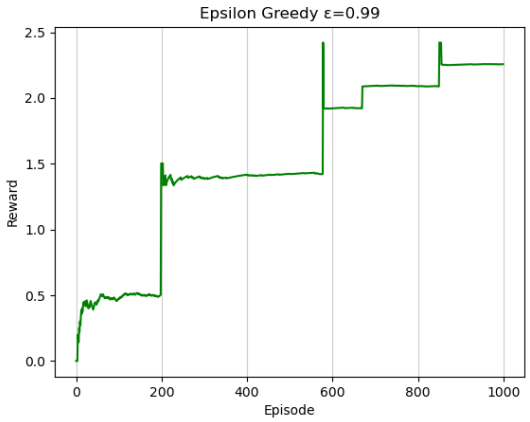
61075029H Jun Yu SHEN 沈峻宇

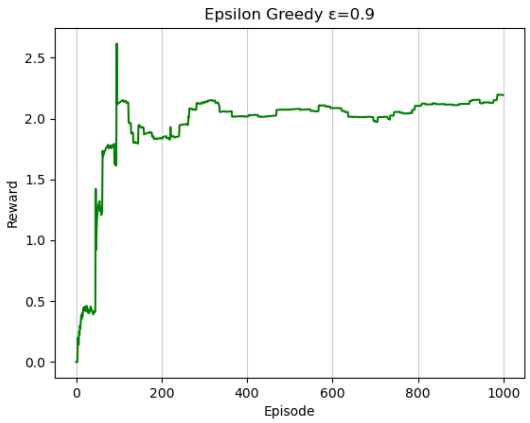
Q1. Compare MAB problem with three algorithms implementations (Thompson Sampling, UCB, and ε-greedy). Note: plot and compare mean total reward and time for three algorithms, 1000 iterations (develop given python example code that already has Thompson Sampling implementation)

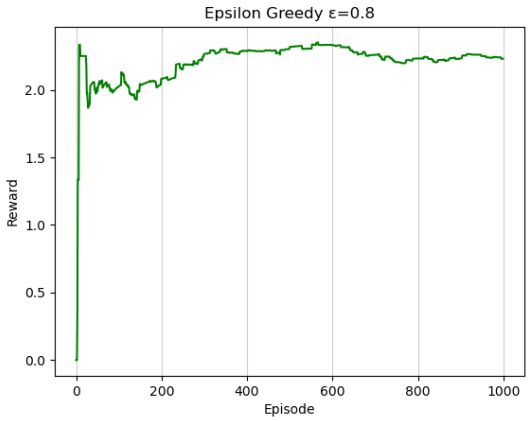
A1. The steps of **ε-greedy** are as follows.

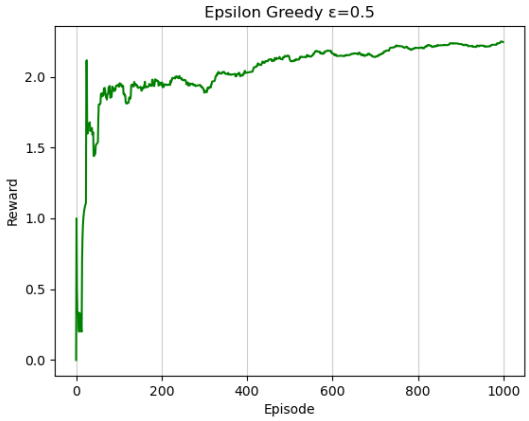
a. Generate a random number (rand ∈ [0, 1])
b. If rand < ε choose a greedy action (exploitation)
c. Otherwise choose a random action (exploration)

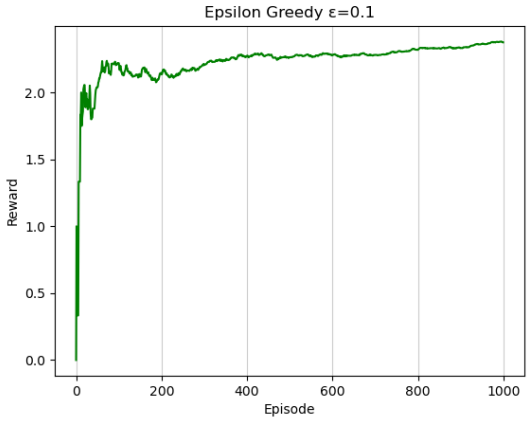
We can adjust "ε" at will to change the result.







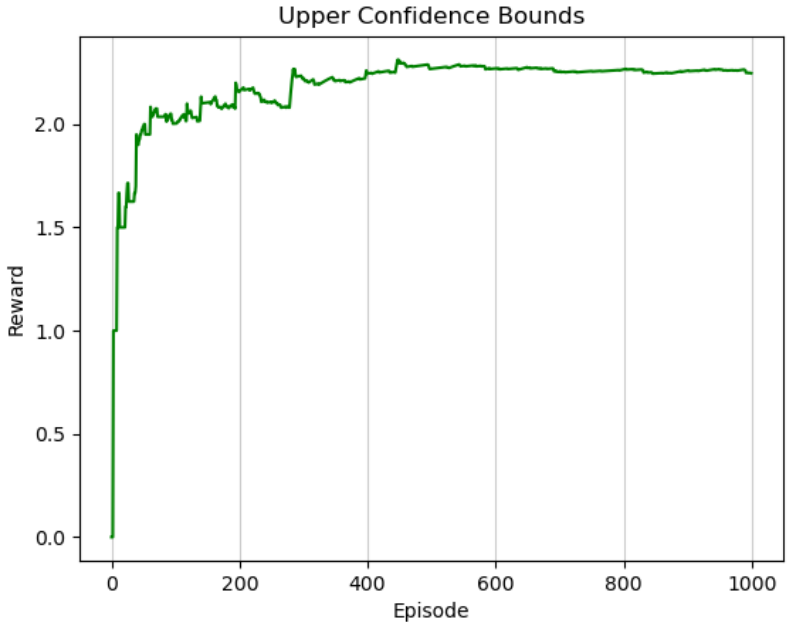




Set the appropriate parameters to effectively search for good bandits. Find a good balance between explore and exploit. **ε-greedy**'s average execution time is 0.014 seconds, very fast. **Upper Confidence Bounds (UCB)** algorithm Favor exploration of actions with a strong potential to have a optimal value! We have a bound per each action that indicates how confident we are about the Q-value of that action.

$$a_t^{UCB} = \arg\max_{a \in A} \quad Q_t(a) + U_t(a)$$

$$U_t(a) = \sqrt{\frac{2\log t}{N_t(a)}}, a \in A$$

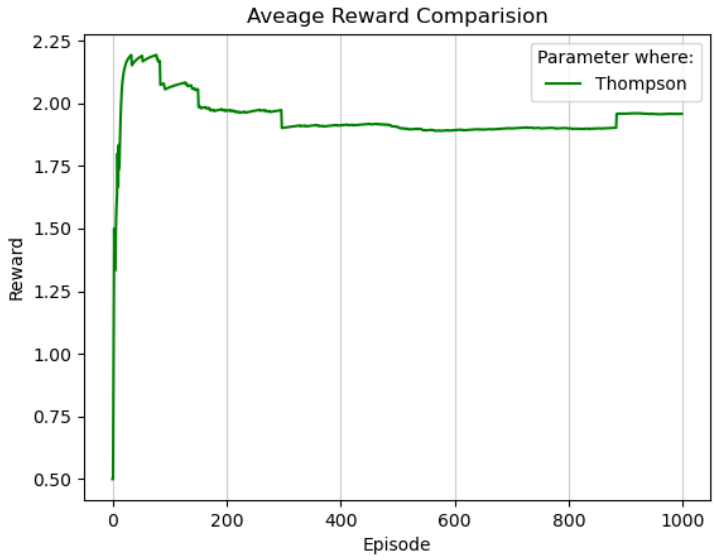


Its average execution time is 0.08 seconds.

Thompson Sampling algorithm. At each time step, we want to select action a according to the probability that a is optimal.

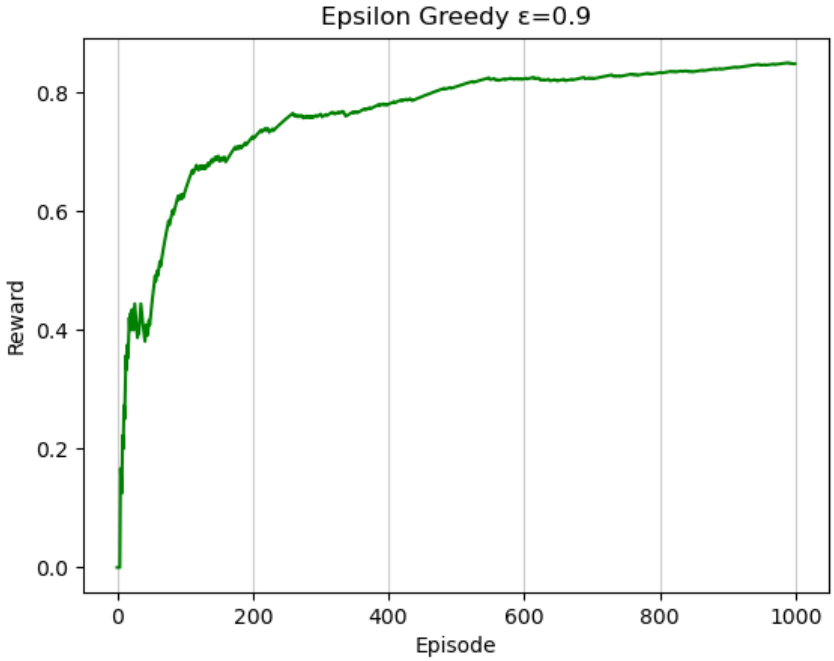
Beta distribution can be considered as $Q(a)$, which is essentially the success probability θ.

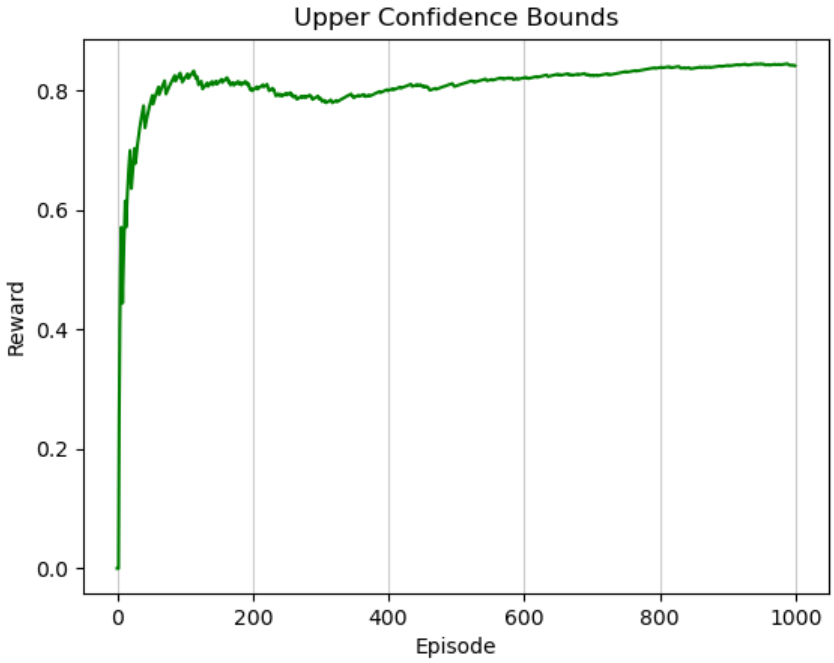
The α and β correspond to the counts when we succeeded or failed to get a reward respectively.

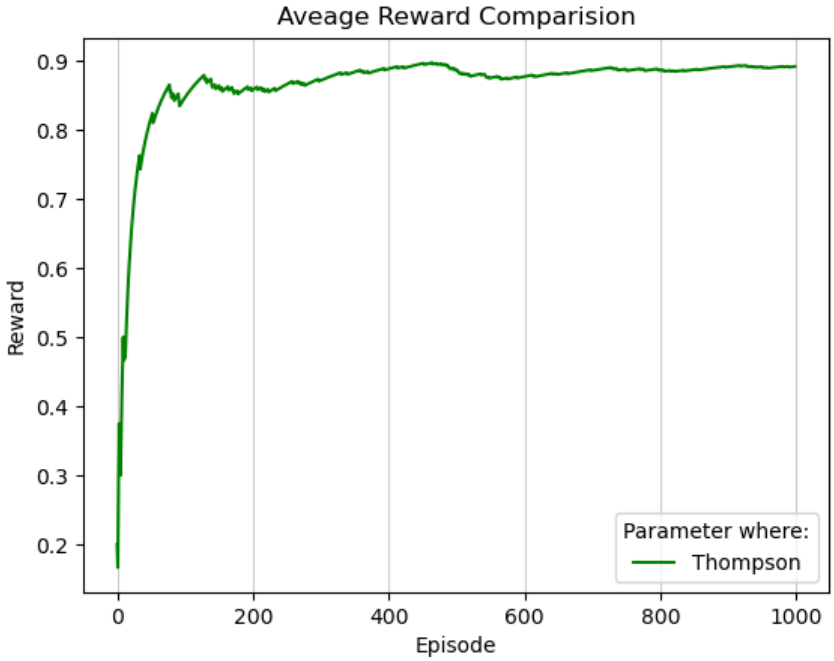


Its average exection time is 2 seconds.

The following cumulative odds are used. Calculate the odds from the beginning of the game to the last sum







From the above diagram, we can see that the Thompson sampling algorithm works very well. From the beginning to two hundred epsiode when the odds have reached 0.9, and the other two in the chart show only up to about 0.8 at the end.

According to the above results I think Thompson sampling algorithm's effect is the best.