1. Which hyperparameters are important for Thompson Sampling, e-greedy, UCB, and random sampling? Show that they are important (15 *Points)*

Thompson Sampling: The most important hyperparameter for Thompson Sampling algorithm is the prior expectation(beta distribution), which represents an initial probability estimate and the number of time-step, the more times the test is, the average reward value is closer to the theoretical reward.

e-greedy: epsilon value, this hyperparameter is a probability, important for e-greedy algorithms controls the degree of preference for exploit and explore, the lower the epsilon, the more conservative the exploration of “ “unknown”, the more inclined it is to choose the known local optimal solution.

UCB: At the beginning, the confidence of all arms is very low, so each arm will be selected several times, and the probability of selecting the optional arm is not high, when many times have been selected, the confidence of each arm is already quite high. The optional arm is always selected occasionally, the other arms with low confidence are selected. Determines the desire to explore through uncertainly, and measures are weighted “c” hyperparameter.

Random sampling: Due to the complete uncertainly of random sampling, these initial settings like time steps and the number of arms are important for the production of rewards.

1. How does the action space affect Thompson Sampling, e-greedy, UCC, and random sampling? Show why. (15 *Points)*

As mentioned in the article, increasing the action space, e-greedy did not change, because its working mode is to select those arms that are not the highest return with a certain probability, and as the number of times increases, the performance of Thompson sampling also Getting closer to steady.

On the other hand, UCB is drastically unstable and diverges. This behaviour is not only due to the increase in action space, but due to the fact that the Bernouilli reward probabilities are so close to each other. As for random sampling,because of the uncertainly attributes, will not have a noticeable impact.

1. How does stationary affect Thompson Sampling, e-greedy, UCB, and random sampling? Show why.

It can be found from the chart that the resample action will be performed every 100 tests, and the three algorithms will show periodic changes, but it may be due to the impact of the previous memory. The performance of the three algorithms tends to decline, while UCB also maintains instability.

1. When do Thompson Sampling, e-greedy, UCB, and random sampling stop exploring? Explain why. Explain the exploration-exploitation tradeoff.

Thompson Sampling UBC and random sampling will stop when approaches the optimal solution, but UCB, since the probability of exploration and exploitation is controlled by the e hyperparameter, it may stay on the local optimal solution, rather than the highest reward.

1. How long do Thompson Sampling, e-greedy, UBC, and random sampling remember the past actions? Explain your answer. (10 *Points)*

### Through theNon-stationarity test, it can be found that even if the resample action is executed, the previous data still affects the performance of the algorithm, so the memory of the past action may be retained until the end of the test.