**Name: John Stephen Gutam**

**Email: jgutam@gmu.edu**

**Problem 1:**

A white sheet of paper with writing on it

Description automatically generated

A white paper with math equations

Description automatically generated

A piece of paper with math equations

Description automatically generated

A white paper with writing on it

Description automatically generated

A white paper with writing on it

Description automatically generated

A white paper with math equations

Description automatically generated

A white board with math equations

Description automatically generated

**Problem 2:**

**A screenshot of a computer

Description automatically generated**

**A screen shot of a graph

Description automatically generated**

Based on the graphs and results, we can analyze the properties of the three different algorithms: Greedy, UCB1, and Thompson Sampling.

1. Average Regret vs. Time:
   * For the eleven-armed bandit setting, the Greedy algorithm performs poorly, with a constant high regret throughout the time steps. UCB1 and Thompson Sampling algorithms have significantly lower regret, with Thompson Sampling having the lowest regret overall.
   * For the five-armed bandit setting, the Greedy algorithm still performs poorly compared to UCB1 and Thompson Sampling. However, the difference in regret between UCB1 and Thompson Sampling is smaller compared to the eleven-armed bandit case.
2. Action Selection Over Time:
   * For the eleven-armed bandit setting, the Greedy algorithm quickly converges to selecting the arm with the highest probability (arm 10), but it takes a long time to explore and identify the optimal arm.
   * UCB1 and Thompson Sampling explore more efficiently and converge faster to the optimal arm (arm 10) compared to the Greedy algorithm.
   * For the five-armed bandit setting, all three algorithms converge to the optimal arm (arm 4) relatively quickly, but UCB1 and Thompson Sampling still explore more efficiently and converge faster than the Greedy algorithm.

Interesting Insights:

1. The Greedy algorithm performs poorly in both settings, as it lacks an exploration mechanism and can get stuck on sub-optimal arms, leading to high regret.
2. UCB1 and Thompson Sampling outperform the Greedy algorithm by balancing exploration and exploitation effectively. They have lower regret and converge faster to the optimal arm.
3. Thompson Sampling generally performs better than UCB1, especially in the eleven-armed bandit setting, where the number of arms is larger. This suggests that Thompson Sampling is more efficient in exploring and identifying the optimal arm in complex environments with more choices.
4. The difference in performance between UCB1 and Thompson Sampling is more pronounced in the eleven-armed bandit setting compared to the five-armed bandit setting. This indicates that as the number of arms increases, the advantage of Thompson Sampling over UCB1 becomes more significant.
5. The shape of the average regret curves for UCB1 and Thompson Sampling suggests that they have a logarithmic regret bound, which is a desirable property for bandit algorithms.

Overall, the results demonstrate the superiority of UCB1 and Thompson Sampling over the Greedy algorithm in multi-armed bandit problems, with Thompson Sampling having a slight edge, especially in more complex environments with a larger number of arms.