DDA 4230 Assignment 1

Name: Xiang Fei ; ID: 120090414

P1.

1. π_1 : $S_0 \to a_1$, $S_1 \to a_0$, $S_2 \to a_0$, $S_3 \to a_0$

   π_2 : $S_0 \to a_2$, $S_1 \to a_0$, $S_2 \to a_0$, $S_3 \to a_0$

2. $V^*(S_0) = \max\limits_{a\in\{a_1,a_2\}} 0 + \gamma \sum\limits_{s'} P(s'|s,a) V^*(s') = \gamma \max\{V^*(S_1), V^*(S_2)\}$

   $V^*(S_1) = \max\limits_{a\in\{a_0\}} 0 + \gamma \sum\limits_{s'} P(s'|s,a) V^*(s') = \gamma[(1-p)V^*(S_1) + pV^*(S_3)]$

   $V^*(S_2) = \max\limits_{a\in\{a_0\}} 1 + \gamma \sum\limits_{s'} P(s'|s,a) V^*(s') = 1 + \gamma[(1-q)V^*(S_0) + qV^*(S_3)]$

   $V^*(S_3) = \max\limits_{a\in\{a_0\}} 0 + \gamma \sum\limits_{s'} P(s'|s,a) V^*(s') = 0 + \gamma V^*(S_0)$

3. No.

   if $\gamma = 0$. since $\pi^*(S_0) = \arg\max\limits_{a\in\{a_1,a_2\}} \gamma \sum\limits_{s'} P(s'|s,a) V^*(s')$

   so π^* is not unique , then $\pi^*(S_0) = a_2$ can not be sure.

4. No.

   $V^*(S_2) = 1 + \gamma[(1-q)V^*(S_0) + qV^*(S_3)] \geq 1$

   $\Rightarrow$ iff $V^*(S_1) > V^*(S_2) \geq 1$, $\forall \gamma \in [0,1)$ and $p > 0$ $\to \pi^*(S_0) = a_1$

   $\Rightarrow V^*(S_1) = \gamma[(1-p)V^*(S_1) + pV^*(S_3)] = \dfrac{\gamma p V^*(S_3)}{1-\gamma(1-p)}$

   $\Rightarrow \exists \gamma$ such ~~that~~ $V^*(S_1) < 1$.

P2.

1. $\|BV - BV'\|_\infty = \|\max_a [R(s,a) + \gamma \sum_{s_j \in S} P(s_j|s_i,a) V(s_j)] - \max_a [R(s,a) + \gamma \sum_{s_j \in S} P(s_j|s_i,a) V'(s_j)]\|_\infty$

$\leq \max_a \|[R(s,a) + \gamma \sum_{s_j \in S} P(s_j|s_i,a) V(s_j) - R(s,a) - \gamma \sum_{s_j \in S} P(s_j|s_i,a) V'(s_j)]\|_\infty$

$\leq \gamma \max_a \|\sum_{s_j \in S} P(s_j|s_i,a) V(s_j) - \sum_{s_j \in S} P(s_j|s_i,a) V'(s_j)\|_\infty$

$= \gamma \max_a \|[\sum_{s_j \in S} P(s_j|s_i,a)(V(s_j) - V'(s_j))]\|_\infty$

$\leq \gamma \max_{a, s_i} \sum_{s_j \in S} P(s_j|s_i,a) |V(s_j) - V'(s_j)|$

$\leq \gamma \max_{a, s_i} \sum_{s_j \in S} P(s_j|s_i,a) |V - V'| = \gamma \|V - V'\|$, Q.E.D.

2.1. basic case: $\|V_2 - V_1\|_\infty = \|BV_1 - BV_0\|_\infty \leq \gamma \|V_1 - V_0\|_\infty$

inductive assumption: $\|V_{k+1} - V_k\|_\infty \leq \gamma^k \|V_1 - V_0\|_\infty$

inductive step: $\|V_{k+2} - V_{k+1}\|_\infty = \|BV_{k+1} - BV_k\|_\infty \leq \gamma \|V_{k+1} - V_k\|_\infty \leq \gamma^{k+1} \|V_1 - V_0\|_\infty$

$\Rightarrow \|V_{n+1} - V_n\|_\infty \leq \gamma^n \|V_1 - V_0\|_\infty$, Q.E.D.

2.2. $\|V_{n+c} - V_n\| \leq \|V_{n+c} - V_{n+c-1}\| + \|V_{n+c-1} - V_n\|$

$\leq \|V_{n+c} - V_{n+c-1}\| + \cdots + \|V_{n+1} - V_n\|$

$\leq \gamma^{n+c-1} \|V_1 - V_0\| + \gamma^{n+c-2} \|V_1 - V_0\| + \cdots + \gamma^n \|V_1 - V_0\|$

$= \frac{\gamma^n (1 - \gamma^c)}{1 - \gamma} \|V_1 - V_0\| \leq \frac{\gamma^n}{1 - \gamma} \|V_1 - V_0\|$, Q.E.D.

3. let $k < n$, $m = n + c$, $\varepsilon = \frac{\gamma^k}{1 - \gamma} \|V_1 - V_0\|_\infty$

$\|V_m - V_n\|_\infty \leq \frac{\gamma^n}{1 - \gamma} \|V_1 - V_0\|_\infty$

$< \frac{\gamma^k}{1 - \gamma} \|V_1 - V_0\|_\infty = \varepsilon$

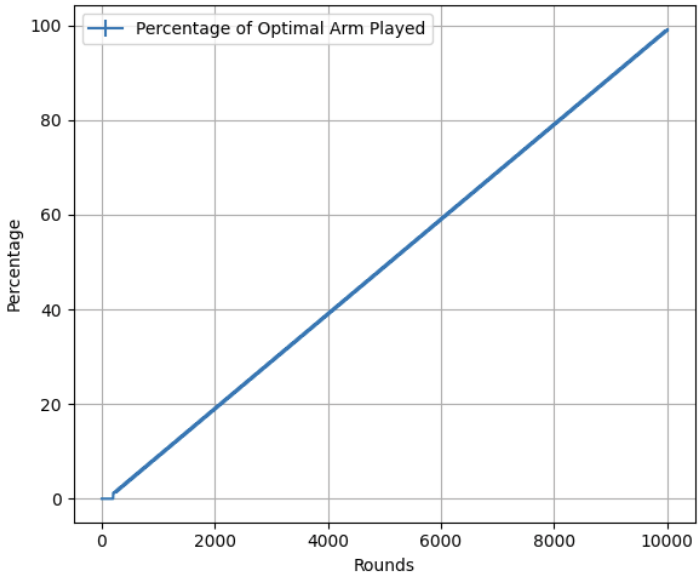
4. Assume two fixed points V and V'

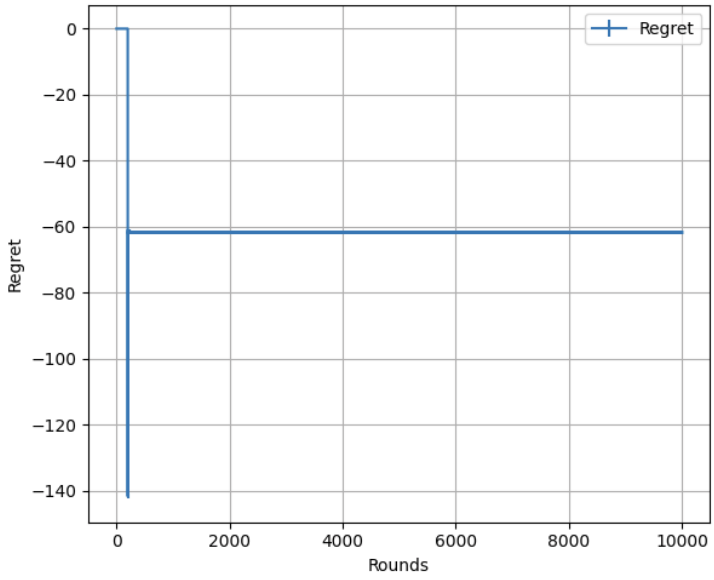
$\|V - V'\|_\infty = \|BV - BV'\|_\infty \leq \gamma \|V - V'\|_\infty$

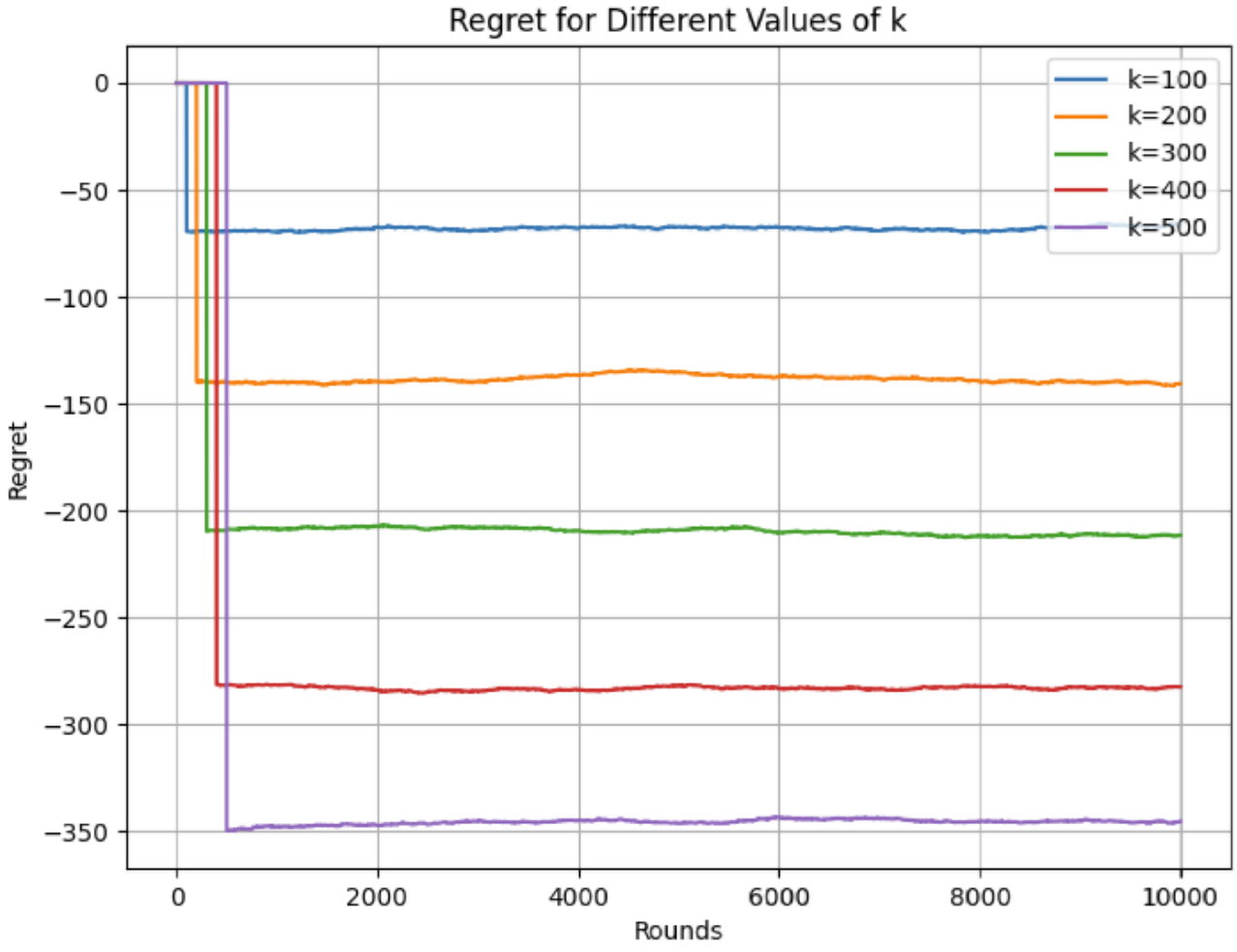
since $0 < \gamma < 1 \Rightarrow \|V - V'\|_\infty = 0 \Rightarrow V = V'$

therefore, only one fixed point

P3


Percentage of Optimal Arm Played
Percentage
Rounds
0
20
40
60
80
100
0
2000
4000
6000
8000
10000


Regret
Regret
Rounds
0
−20
−40
−60
−80
−100
−120
−140
0
2000
4000
6000
8000
10000


Regret for Different Values of k
k=100
k=200
k=300
k=400
k=500
Regret
Rounds
0
−50
−100
−150
−200
−250
−300
−350
0
2000
4000
6000
8000
10000

1. ETC with Optimal k:

- In the case of ETC with an optimal choice of k, we would ideally choose k such that it minimizes the gap-dependent regret. However, in practice, we often lack knowledge of the underlying gap, so this choice is challenging.
- Theoretical findings suggest that for the ETC algorithm with an optimal k, the regret should be minimized because we initially explore both arms for a limited number of rounds (k) and then commit to the arm that appears to be the best based on the collected data.
- In the implementation, we experimented with different values of k, and if we were able to choose an optimal k that closely matched the true underlying gap, we would expect to see low regret. The regret would decrease quickly in the commitment phase as we concentrate on the better arm.

2. ETC with Heuristic Choice for k:

- In practice, we often have to use heuristics or rules of thumb to select the value of k because we typically do not know the true gap.
- When we choose k heuristically, it may not necessarily be optimal. The performance of ETC with a suboptimal k can vary depending on how different the chosen k is from the optimal one.
- If the chosen k is too small, the exploration phase might be insufficient, leading to suboptimal arm selection in the commitment phase, resulting in higher regret. Conversely, if k is too large, it could lead to prolonged exploration and slower convergence to the optimal arm.
- In the implementation, we experimented with different values of k, and we can observe that as k deviates from the theoretically optimal value, the regret might increase. For example, if k is far from the optimal value, the algorithm may spend too much time in exploration or may commit to a suboptimal arm for too long, leading to higher regret.

Overall, the choice of k in the ETC algorithm is critical, and the performance depends on how well it aligns with the true gap between the arms. In practice, it's often challenging to determine the optimal value of k without prior knowledge of the gap, so heuristics are used. The results obtained through experimentation with different values of k help in understanding the trade-off between exploration and commitment and provide insights into the algorithm's performance in the absence of perfect information about the gap.