NAME: Andrew Caide

CSCI S-89c Deep Reinforcement Learning
Part I of Assignment 1

Suppose we run an ε-greedy algorithm for the k-armed Bandit problem, where $\varepsilon \in (0, 1)$. Assuming $q_*(a_1) \neq q_*(a_2)$ for all $a_1 \neq a_2$, where $a_1, a_2 \in \{1, 2, \ldots, k\}$, please express

$$\lim_{t \to \infty} \mathrm{E}[R_t]$$

in terms of ε and $q_*(a)$, $a \in \{1, 2, \ldots, k\}$.

SOLUTION:

$$E[R_t] = \sum \pi_t(a) \cdot a$$

where π_t $\begin{cases} \varepsilon/k & \text{- non greedy} \\ 1 - \varepsilon + (\varepsilon/k) & \text{- greedy} \end{cases}$

$$E[R_t] = \sum \left(\varepsilon/k + 1 - \varepsilon + \varepsilon/k\right) a$$

$$= \sum \left(\frac{2\varepsilon}{k} + 1 - \varepsilon\right) a$$

$$E[R_t \mid A_t = a] = \boxed{\sum \left(\frac{2\varepsilon}{k} + 1 - \varepsilon\right) = q_*(a)}$$