Name: Karma Tarap
CSCI E-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

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

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

SOLUTION:

$$\sum_{i=a}^{k} q_*(a) P(A_t = a)\varepsilon + (1 - \varepsilon) argmax(A) q_*(a)$$