Chapter 2 solutions **Hanna Gábor**

1. *In ϵ-greedy action selection, for the case of two actions and $\epsilon = 0.5$, what is the probability that the greedy action is selected?*

   $\mathbb{P}(\text{greedy action is selected}) = 0.5 + 0.5/2 = 0.75.$

2. *Bandit example. Consider a k-armed bandit problem with $k = 4$ actions, denoted 1, 2, 3, and 4. Consider applying to this problem a bandit algorithm using ϵ-greedy action selection, sample-average action-value estimates, and initial estimates of $Q_1(a) = 0$, for all a. Suppose the initial sequence of actions and rewards is $A_1 = 1$, $R_1 = 1$, $A_2 = 2$, $R_2 = 1$, $A_3 = 2$, $R_3 = 2$, $A_4 = 2$, $R_4 = 2$, $A_5 = 3$, $R_5 = 0$. On some of these time steps the ϵ case may have occurred, causing an action to be selected at random. On which time steps did this definitely occur? On which time steps could this possibly have occurred?*

   Any action can be an explorative move.

   What were the greedy options in different time steps?
   Step 1: all actions have 0 estimated values. Every action is a greedy choice.
   Step 2: $Q_1(1) = 1$. The greedy choice now is 1. $A_2 = 2$ must have been an explorative move.
   Step 3: $Q_2(2) = 1$. The greedy choice is either 1 or 2.
   Step 4: $Q_3(2) = 1.5$. The greedy choice is 2.
   Step 5: $Q_4(2) = 1.67$. The greedy choice is 2. $A_5 = 3$ must have been an explorative move.

   On time steps 2 and 5 a random action must have been selected.