Q3

Ex 2.3 : Which method will perform best in the long run in terms of cumulative reward & probability of selecting the best selecting the bet at. Do this for $\varepsilon = 0, 0.1, 0.01$

Solution: Let $Q_1(a) = 0 \quad \forall a$

Initially, it will choose an action with equal probability.

estimate. After exploring ~~on~~ choosing an action which gives +ve Reward., it won't explore further as with the probability $(1-\varepsilon)$ i.e 1, it will do $\arg\max_a Q_t(a)$!

When $\varepsilon = 0.01$, algorithm will choose greedy action $\left((1-\varepsilon) + \frac{\varepsilon}{k}\right) * 100$ % times

i.e $\left(1 - 0.01 + \frac{0.01}{10}\right) * 100\% = 99.1\%$

while $\varepsilon = 0.1$, algorithm will choose greedy action, $\left(1 - 0.1 + \frac{0.1}{10}\right) * 100\% = 91\%$ of the time.

Hence, $\varepsilon = 0.01$ will perform better in the long run. by 8.1%.

In case of $\varepsilon(t) = 1/t$,

Probability of choosing a greedy action $= 1 - \frac{1}{t} + \frac{1}{t}$

$= 1 - \frac{1}{t} + \frac{0.10}{t} = 1 - \frac{0.9}{t}$

as $t \to \infty$, P(choosing greedy action) $= 1$