

Assignment 1

STAT 497-H | Reinforcement Learning

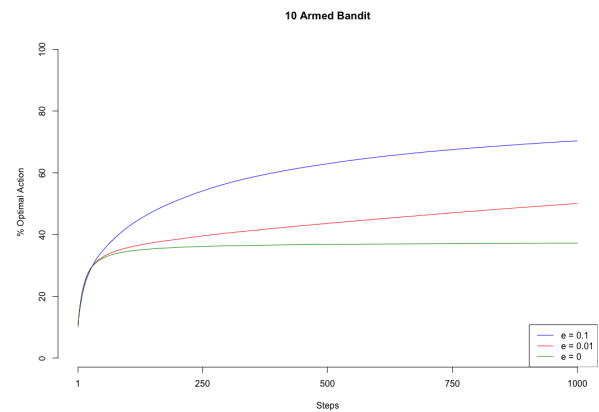
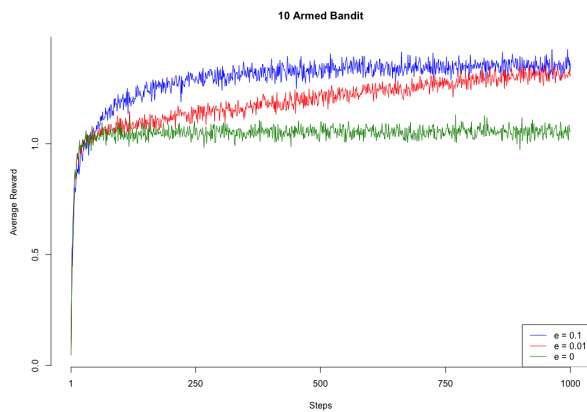
Bakr Abbas (40000333), Matteo Esposito (40024121), Spyros Orfanos (40032280),
Frederic Siino (40028348)

Concordia University

February 1st, 2019

Question 1

- a) See `main.R` code.
 b) See `main.R` code.
 c)

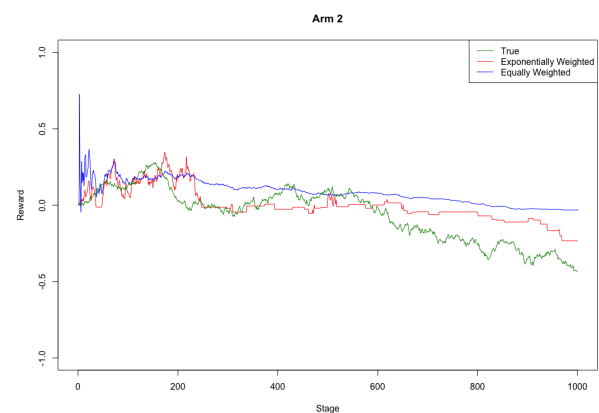
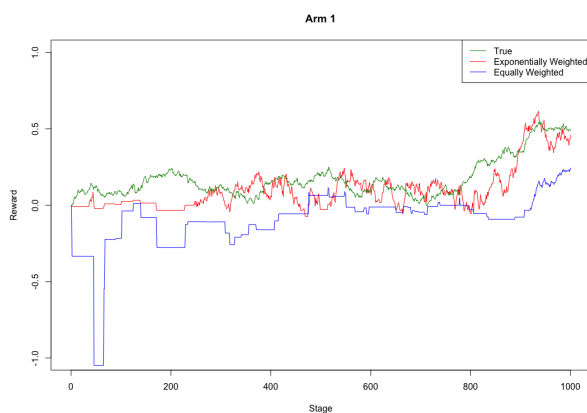


From the average reward plot we can observe that initially the greedy action ($\epsilon = 0$) is the best course of action, however this quickly changes and overtime the less greedy options are clearly better as they promote some level of exploration.

From the % optimal action graph we observe a gradual trend towards a value of $1 - \epsilon$ for each curve excluding the $\epsilon = 0$ curve as this is a property of the ϵ -greedy approach.

Question 2

- a) See `main.R` code.
 b) See `main.R` code.



- c) By observing the graphs of the action value for the two arms, it can be concluded that the exponentially weighted method yields more accurate predictions for the action value. Since the

rewards from both arms follow an autoregressive process, the current reward depends on the previous reward. Thus, by using an exponentially weighted step size, we are attributing a progressively decreasing weight to older action value estimates and giving more weight to more recent rewards, which is more suitable for autoregressive models. Since the equally weighted method assigns equal weight to all rewards, the more recent rewards have less of an impact than they should, while the older, less relevant rewards, have the same impact as the more recent rewards on the estimation.

Question 3

a) See `main.R` code.

b) We see that the probability of choosing Arm 3 approaches 1, which is expected since rewards from Arm 3 have the highest mean. Further, the action preference if Arm 2 is higher than that of Arm 1, which is also expected since its rewards have a higher mean than those of Arm 2.

