Learning Dynamics: Assignment 3
Multi-Armed Bandits

Hakim Boulahya

Université Libre de Bruxelles
hboulahy@ulb.ac.be - 000391737

December 16, 2017

Contents

1 N-Armed Bandit

1.1 Exercice 1

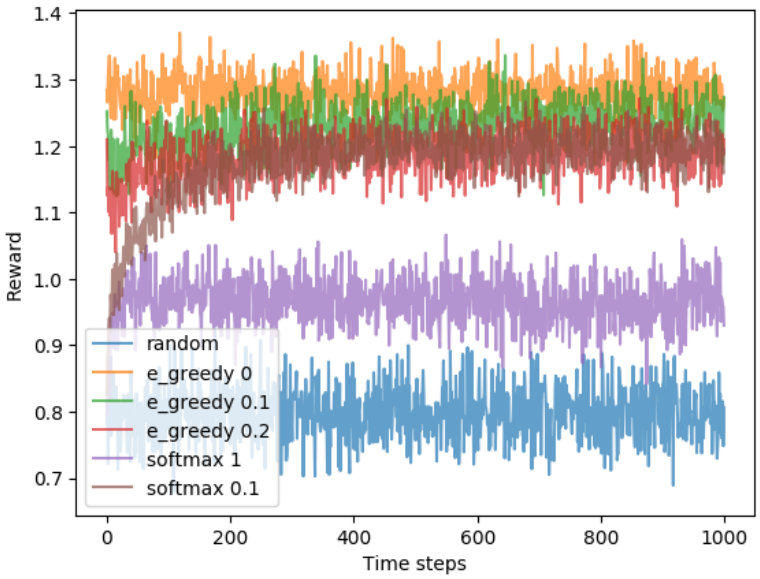
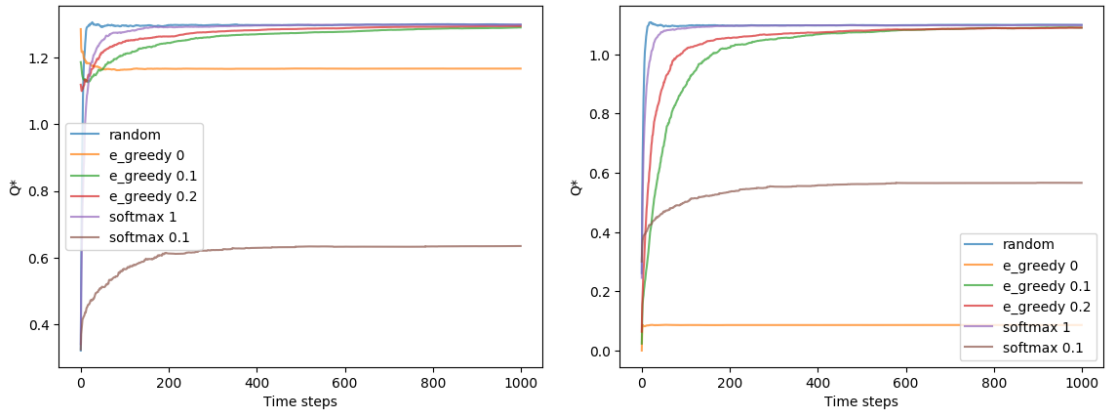
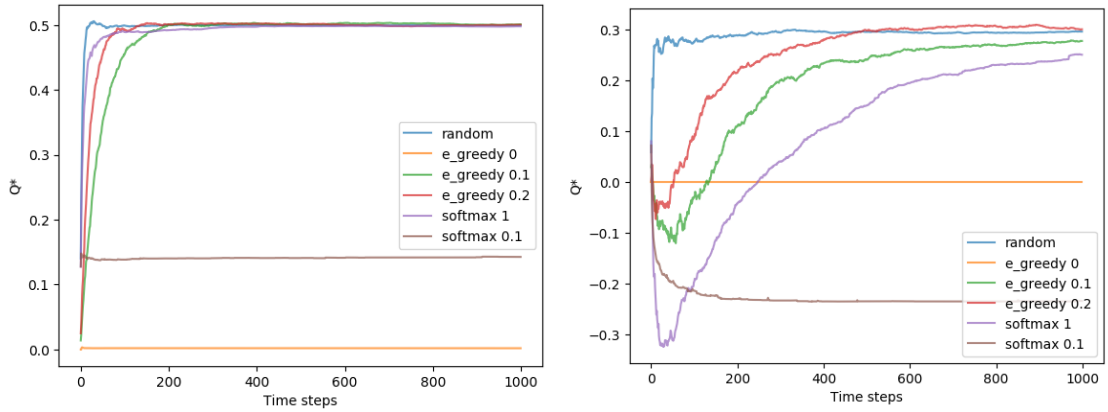


Figure 1: Average rewards for all algorithms

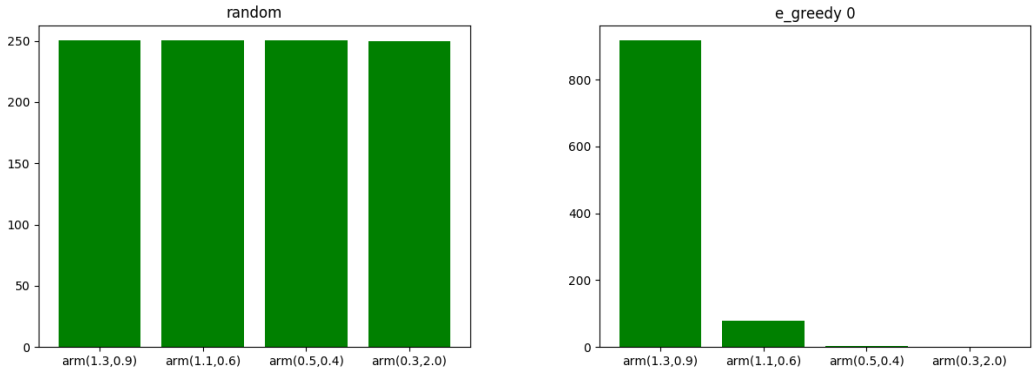


(a) $Q_{a_0}(t)$ (b) $Q_{a_1}(t)$



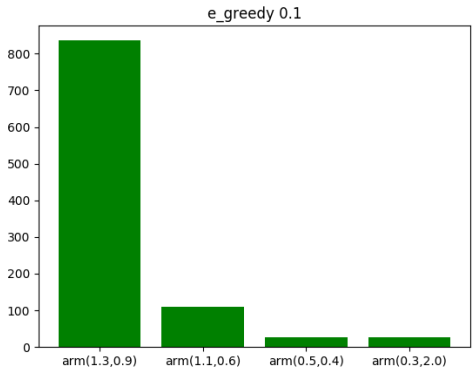
(c) $Q_{a_2}(t)$ (d) $Q_{a_3}(t)$

Figure 2: Plot per arm showing the $Q^*_{a_i}$ of that action along with the actual Q_{a_i} a i estimate over time with μ = (1.3, 1.1, 0.5, 0.3), σ = (0.9, 0.6, 0.4, 2.0)

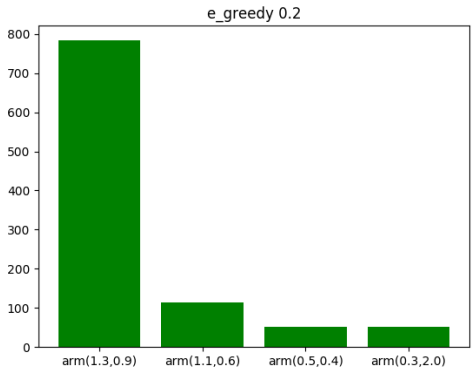


(a)

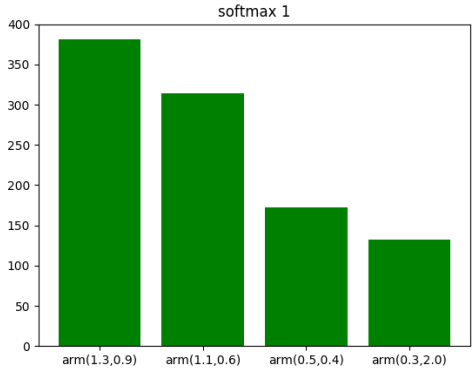
(b)



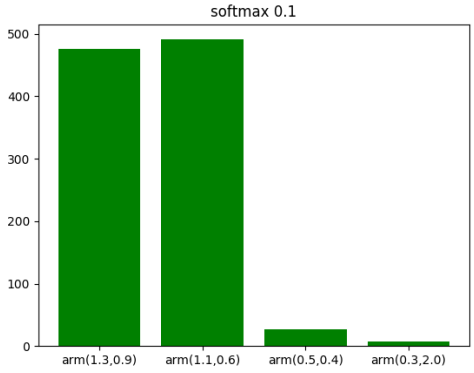
(c)



(d)



(e)



(f)

Figure 3: Histograms showing the number of times each action is selected per selection strategy with μ = (1.3, 1.1, 0.5, 0.3), σ = (0.9, 0.6, 0.4, 2.0)

1.2 Exercice 2

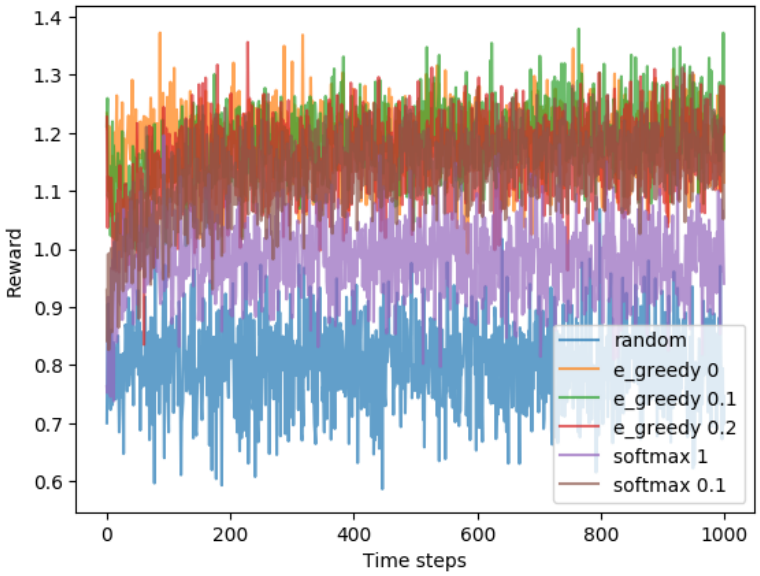
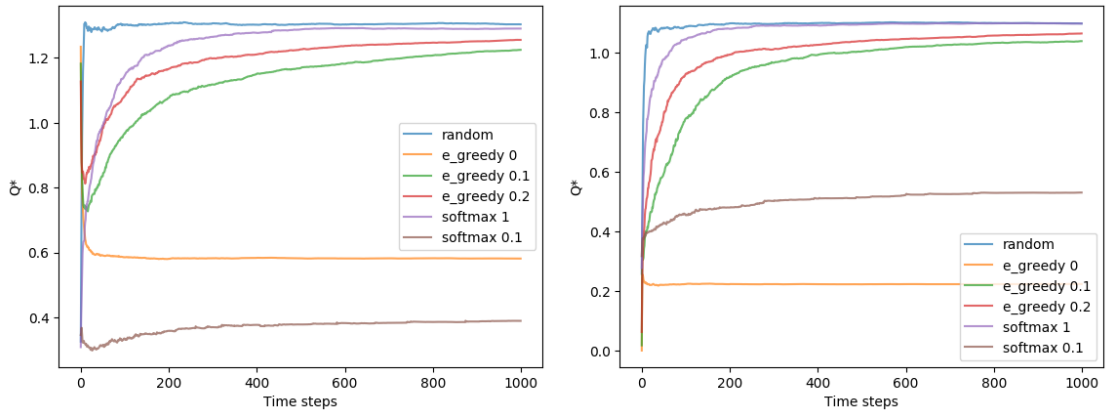
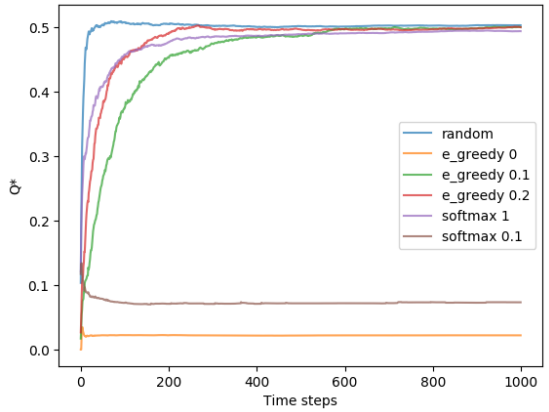


Figure 4: Average rewards for all algorithms

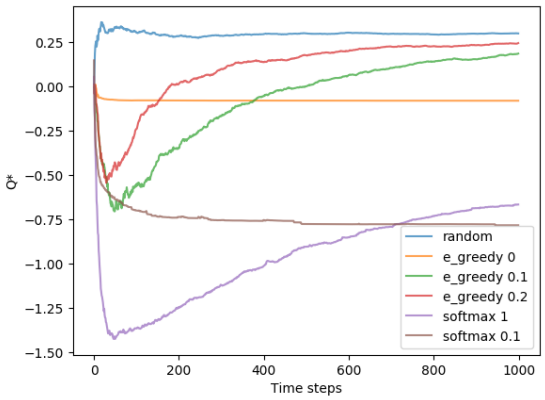


(a) $Q_{a_0}(t)$

(b) $Q_{a_1}(t)$

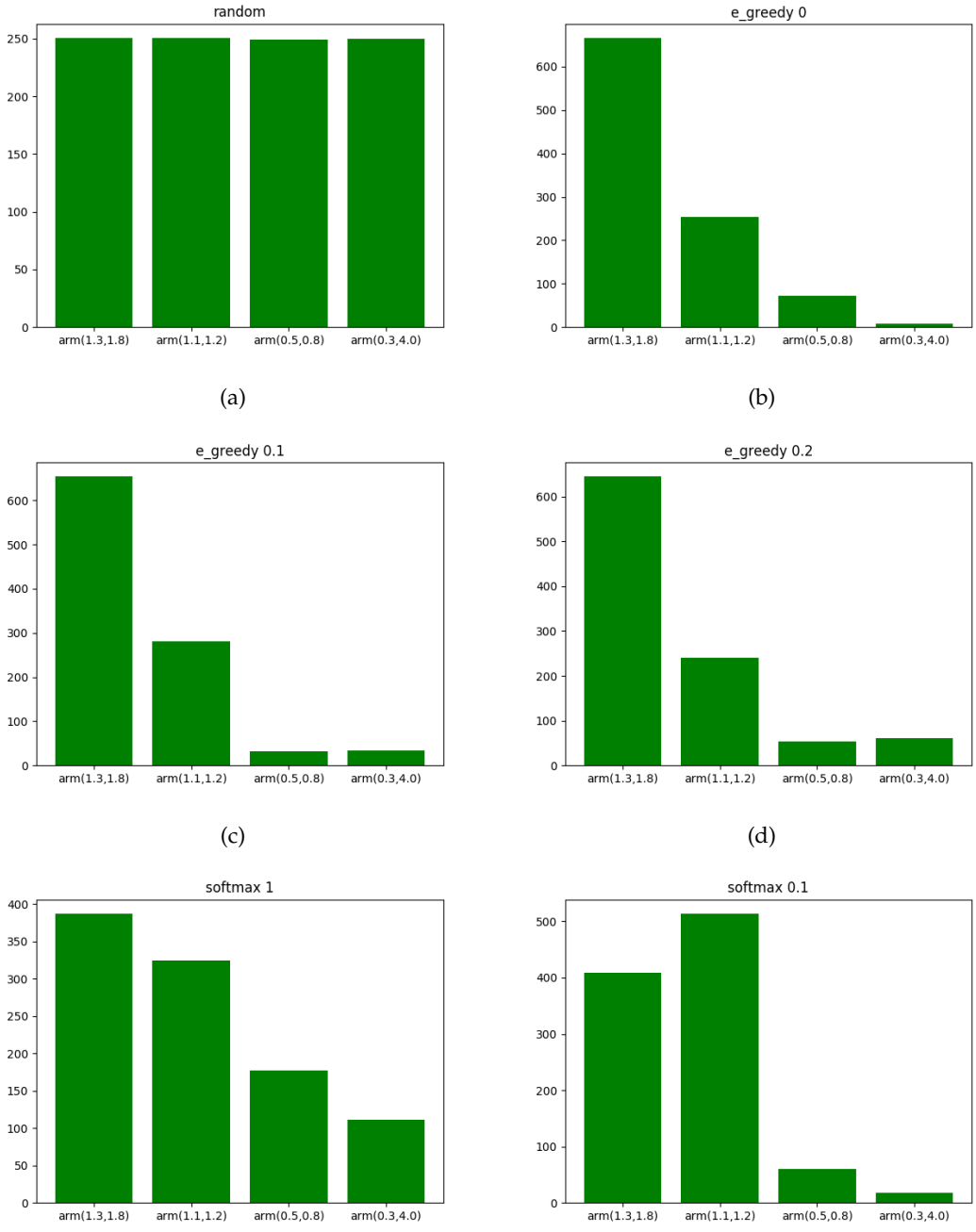


(c) $Q_{a_2}(t)$



(d) $Q_{a_3}(t)$

Figure 5: Plot per arm showing the $Q^*_{a_i}$ of that action along with the actual Q_{a_i} a i estimate over time with μ = (1.3, 1.1, 0.5, 0.3), σ = (1.8, 1.2, 0.8, 4.0)



(a) (b)

(c) (d)

(e) (f)

Figure 6: Histograms showing the number of times each action is selected per selection strategy with μ = (1.3, 1.1, 0.5, 0.3), σ = (1.8, 1.2, 0.8, 4.0)

1.3 Exercice 3

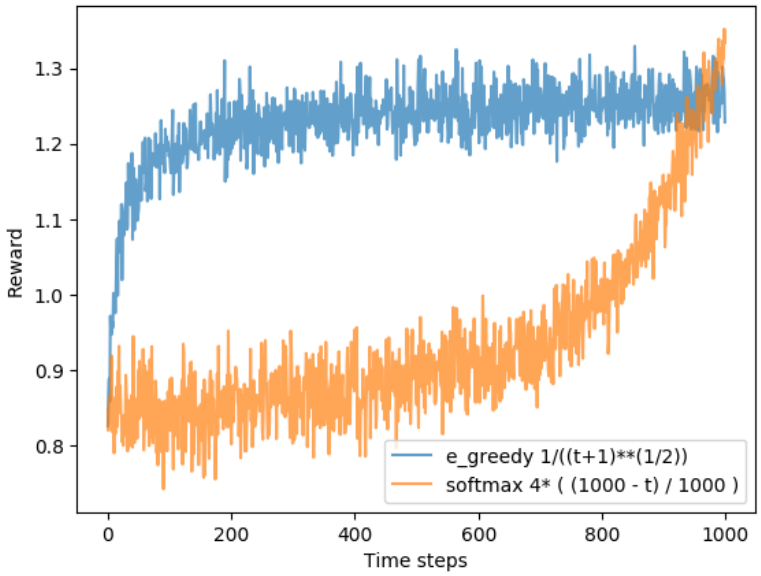
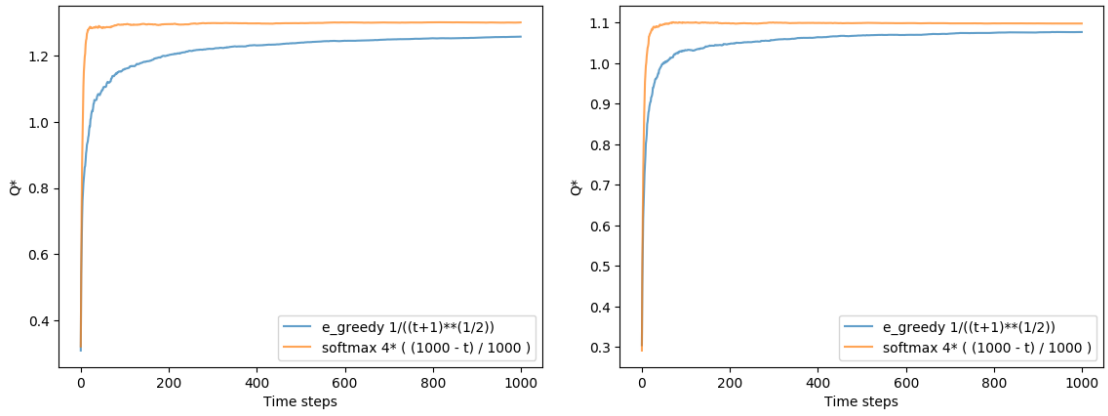
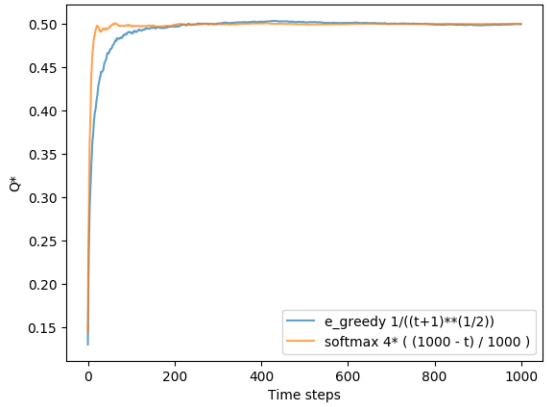


Figure 7: Average rewards for all algorithms

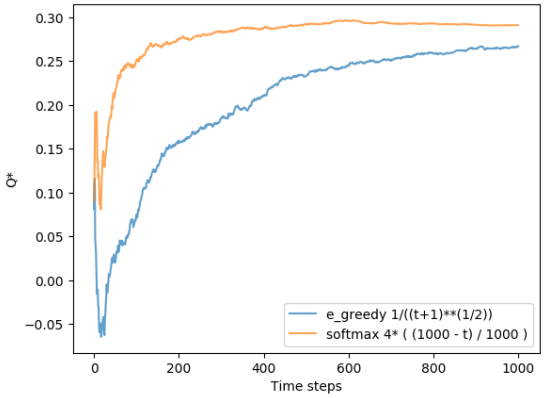


(a) $Q_{a_0}(t)$

(b) $Q_{a_1}(t)$

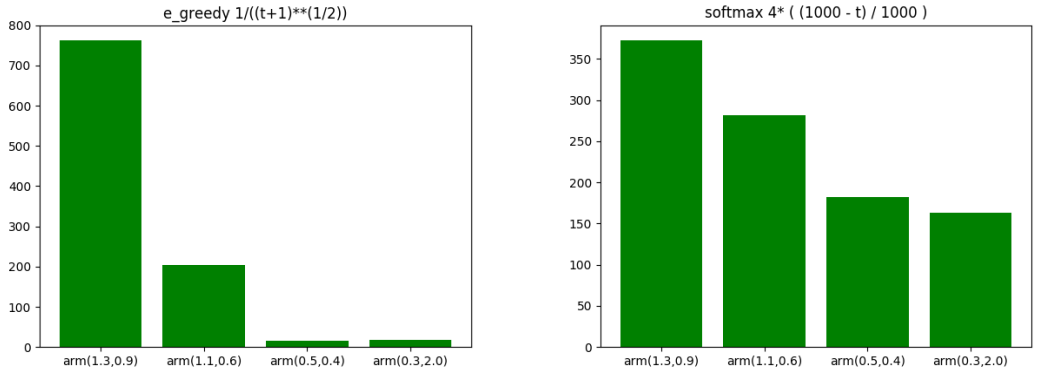


(c) $Q_{a_2}(t)$



(d) $Q_{a_3}(t)$

Figure 8: Plot per arm showing the $Q^*_{a_i}$ of that action along with the actual Q_{a_i} a i estimate over time



(a) (b)

Figure 9: Histograms showing the number of times each action is selected per selection strategy