**Aldar Saranov**

Aldar.Saranov@ulb.ac.be

Université libre de Bruxelles

Assignment 3

Multi-Armed Bandits

Discipline: INFO-F409 Learning dynamics (M-INFOS/F277)

Teachers: Tom Lenaerts, Roxana Rădulescu

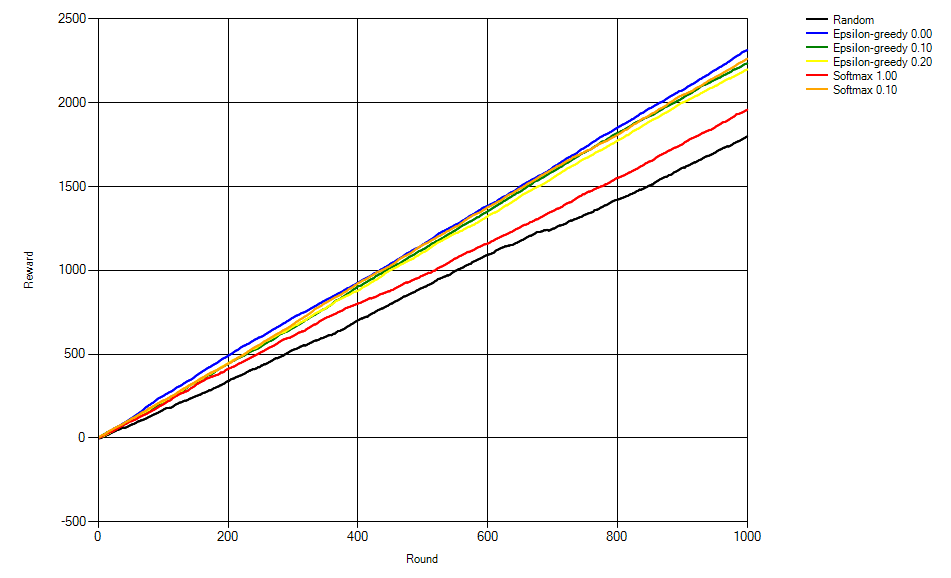
November 2016

# N-Armed Bandit.

The specified problem was simulated using an application (code is located in the same archive).

## Exercise 1

### Reward per round



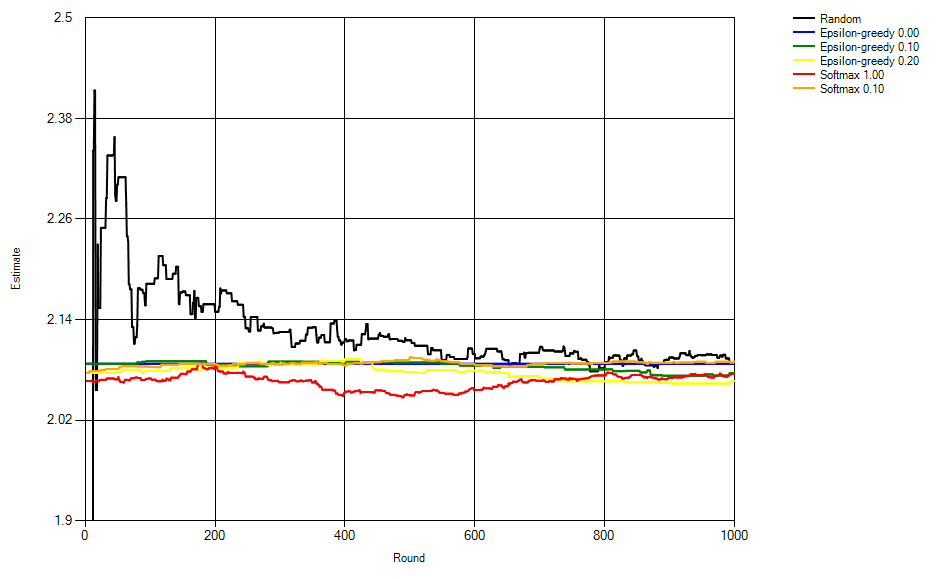
The algorithms achieve rewards in following descending order:

1. Epsilon greedy 0.0
2. Softmax 0.1
3. Epsilon-greedy 0.1
4. Epsilon-greedy 0.2
5. Softmax 1.0
6. Random

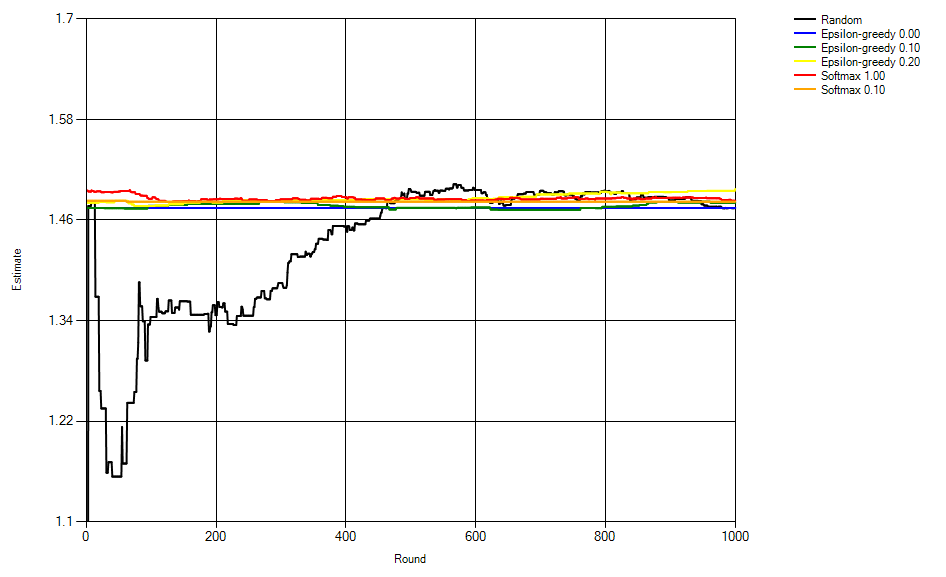
Random showed the worst result since it does not do much of exploiting. Epsilon greedy 0.0 showed the best result obviously because it stands for the best combination of exploration and exploiting.

### Arm 1

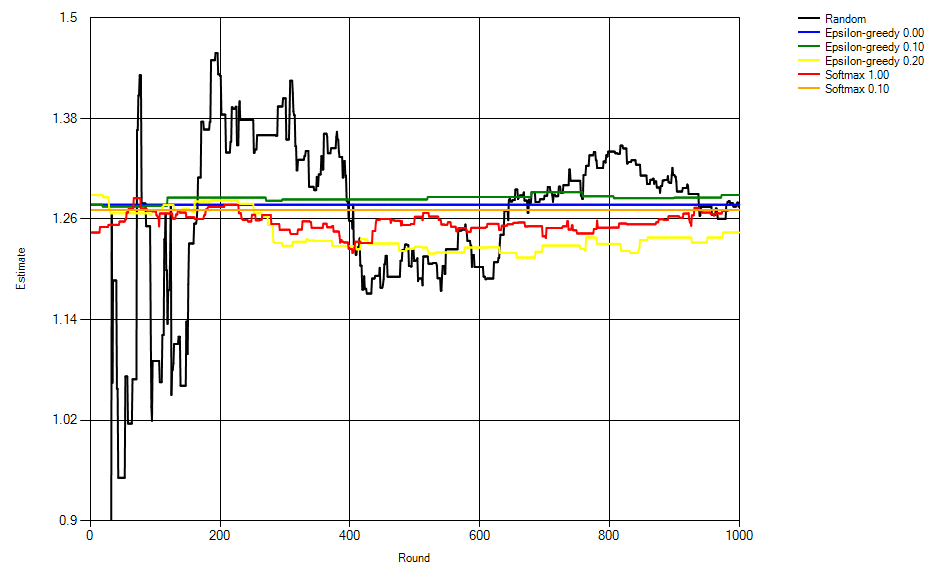
### Arm 2



### Arm 3



### Arm 4



Based on research of these 4 arms estimations we can establish a table of ranks of the action selection algorithms.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm rank\Arm number | 1 | 2 | 3 | 4 |
| 1 | Softmax 0.1 | Softmax 0.1 | Epsilon-greedy 0.2 | Epsilon-greedy 0.1 |
| 2 | Epsilon-greedy 0.2 | Epsilon-greedy 0.0 | Softmax 1.0 | Epsilon-greedy 0.0 |
| 3 | Softmax 1.0 | Random | Softmax 0.1 | Random |
| 4 | Random | Epsilon-greedy 0.1 | Epsilon-greedy 0.1 | Softmax 0.1 |
| 5 | Epsilon-greedy 0.0 | Softmax 1.0 | Epsilon-greedy 0.0 | Softmax 1.0 |
| 6 | Epsilon-greedy 0.1 | Epsilon-greedy 0.2 | Random | Epsilon-greedy 0.2 |

From this table we can infer that in average (in a general case) softmax 0.1 proved to be a more precise action selection algorithm.