# Project: Performance of Multi-armed Bandit Algorithms

Zhuyun YIN

18602298

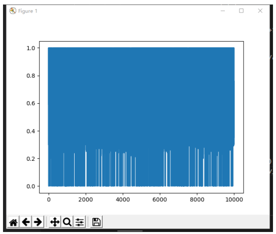
## 1 $\epsilon$-GREEDY ALGORITHM

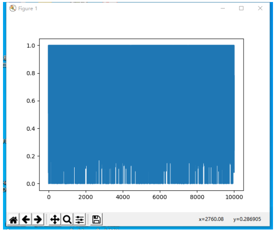
1. $\epsilon$ control the explore of the Algorithm.When $\epsilon$ increasing,it will more tend to choose other arms.So if $\epsilon$ is appropriate,the algorithm will have a good performance.In this project,I choose $\epsilon = 0.1$,0.5 and 0.9.

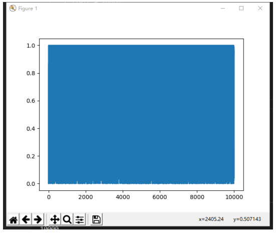
2. Result:

| $\epsilon$ | 0.1 | 0.5 | 0.9 |
|---|---|---|---|
| avg reword | 6181.13 | 6981.59 | 7775.53 |

It seems $\epsilon = 0.9$ is best.







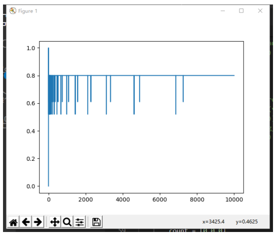
# 2 Upper Confidence Bound Algorithm

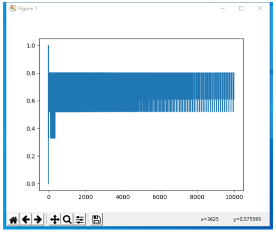
1. In UCB algorithm $I(t) = \sqrt{a^2 + b_0^2 + e^x}$

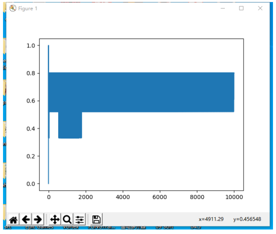
2. Result:

| C | 1 | 5 | 10 |
|---|---|---|---|
| avg reword | 7289.33 | 7133.51 | 6833.53 |

It seems c = 1 is best.







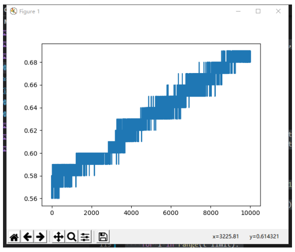
# 3 Thompson sampling (TS) Algorithm

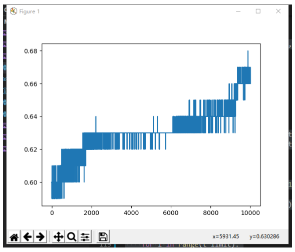
1. parameter a and b in Beta distribution control the PDF.Because the mean is $\frac{a}{a+b}$ and the variance is $\frac{(a*b)}{(a+b)^2*(a+b+1)}$
In this project,we use two set of parameter:(1,1),(1,1),(1,1) and (2,4),(3,6),(1,2)

2. Result:

| Beta | 1 | 2 |
|---|---|---|
| avg reword | 4999.96 | 4986.71 |

It seems second set is better.





## 4 EXPLORATION-EXPLOITATION TRADE-OF

1. Because we can't get accurate prediction of the value.If we don't explore in bandit algorithms,maybe we will never choose an indeterminate option though it can give us more reward.

   But if the exploration rate too high,it will close to a random choose.So we need to control it in a suitable area.Through our experience,we can set appropriate parameter to get better result.

## 5 ALGORITHM OF DEPENDENT CASE

1. If we assume the function f(x) perform relation of arms.Like at same time,there's only one arm has reward.Through testing previous algorithm,I get some results:

| dependent | $\epsilon = 0.1$ | $\epsilon = 0.5$ | $\epsilon = 0.9$ | c = 1 | c = 5 | c = 10 | Beta set1 | Beta set2 |
|---|---|---|---|---|---|---|---|---|
| avg reword | 2797.96 | 3318.02 | 3824.61 | 5190.94 | 5038.06 | 4743.74 | 3177.87 | 3150.22 |

   So I choose to improve UCB algorthm.