# Multi-Agent Systems Assignment 4

Konstantin Mihhailov, Haohui Zhang, Roel van de Laar, Henrik Kathmann

December 3, 2021

## 1 Question 1 - Monte Carlo Simulation

### 1.1 Uncertainty on the result.

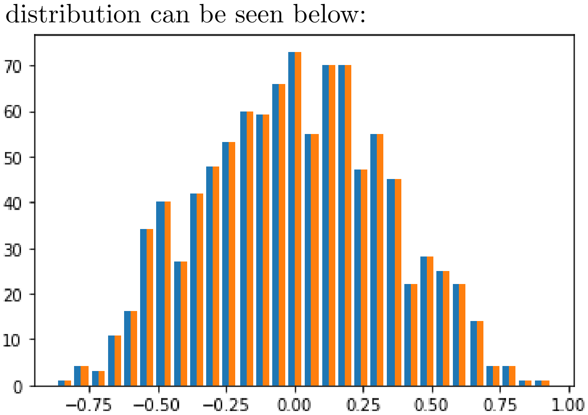
**Answer:** All the processes were done in Python. To get the Monte Carlo estimate we generated 10000 sample points from standard normal distribution. Then we computed the given function $E(\cos^2(X))$ for each sample point. The mean of the function values was then calculated: 0.56809. This mean is the Monte Carlo Estimate. From the estimate, the standard deviation was calculated using the Scipy librabry. The standard deviation was: 0.00346.

This indicates that $E(\cos^2(X)) \sim 0.56809$, gathered from the 10000 samples. With uncertainty of $\pm 0.00346$

### 1.2 Score and Hyper-parameter correlation

Since we can take it that there is no correlation, then the data will be subject to random fluctuation. Using python, random independent variables were generated to find the p-value. In this case, after seeing how often the correlations of 0.3 are found, we get the p-value of 0.37. This means that 37% the correlation coefficients are higher values than 0.3. This indicates that correlation is not significant.

The distribution can be seen below:




## 2 Importance Sampling

### 2.1 Part one

Let $X \sim N(0,1)$ be a standard normal stochastic variable. Use importance sampling to estimate $E(X^2)$ by sampling from a uniform distribution $q \sim U(-5,5)$ on the interval $[-5,5]$. What value do you expect (based on your knowledge of the normal distribution)? How accurate is your estimate based on importance sampling?

**Answer:** all the calculations were done using Python. Using 10.000 samples a mean varying between

0.99 and 1 was retrieved (when running the experiment multiple times) with a standard deviation of $\approx 0.01$.

### 2.2 Part two

Suppose some random process produces output ($1 \leq X \leq 1$) that is distributed according to the following continuous density:

$$f(x) = \frac{1 + cos(\pi x)}{2}$$

Again we are interested in estimation $E(X^2)$. However, as this is not a standard distribution it makes sense to use importance sampling to estimate this value. Quantify the uncertainty on your result.

**Answer:** after running the experiments multiple times, a mean of $\approx 0.13$ was retrieved with a standard deviation of $\approx 0.0028$. After quantifying the uncertainty of our result, we end up with an uncertainty 0.0084.

## 3 Kullback-Leibler Divergence

### 3.1 Part one

There is no symmetry given in the Kullback-Leibler Divergence

$$KL(f|g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx$$

$$= \int f(x) \log \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(x-\nu)^2}{2\tau^2}\right)} dx$$

$$= \frac{1}{2}\left[\log \frac{\tau^2}{\sigma^2} - 1 + \frac{\sigma^2 + (\mu-\nu)^2}{\tau^2}\right.$$

$$KL(g|f) = \int_{-\infty}^{\infty} g(x) \log \frac{g(x)}{f(x)} dx$$

$$= \int g(x) \log \frac{\frac{1}{\sqrt{2\pi\tau^2}} \exp\left(\frac{-(x-\nu)^2}{2\tau^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

$$= \frac{1}{2}\left[\log \frac{\sigma^2}{\tau^2} - 1 + \frac{\tau^2 + (\nu-\mu)^2}{\sigma^2}\right.$$

so $KL(f|g) \neq KL(g|f)$

### 3.2 Part 2

If we choose a function with a mean of 4 and a variance of 1 and another function with a mean of 5 and a variance of 2, we will obtain KL(f/g) = .76 and KL(g/f) = .99. Therefore, our outcome is in line with the theory as different values indicate asymmetry.

# 4 Question 4

## 4.1 Part 1

We restrict our experiments to normal distribution with the same standard deviation $\sigma$. We initialize we have 10 bandits with different mean $\mu$ separately 98, 82, 87, 87, 32, 29, 58, 62, 93, 88 and $\sigma = 23$. We calculate the the loss first according to:

$$l(t) = E(\sum_{i=1}^{t}(q^* - q(a_i)))$$

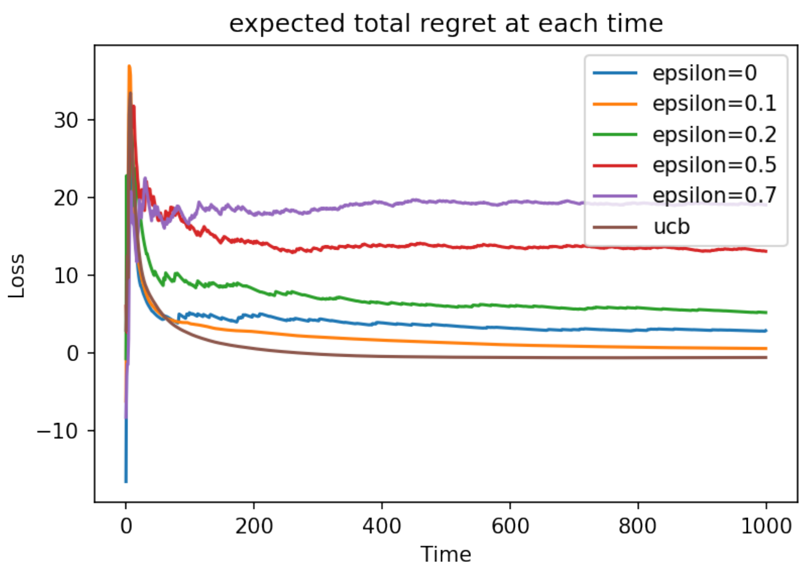



Figure 1: Expected total regret for $\epsilon$-greedy and UCB

Because the normal distribution, we can directly use the conclusion we got from the question 4. $\Delta = |q_* - q(a)|$ and KL divergence and be transformed to:

$$KL(f_1||f_2) = 0 + \frac{1+\Delta^2}{2} - \frac{1}{2} = \frac{\Delta^2}{2}$$

And then follows Lai-Robbins we can get:

$$A = \frac{\Delta}{\Delta^2/2} = \frac{2}{\Delta}$$

## 4.2 Part 2

In this part, we adjust the hyperparameter c and get the figure below. We can find that when c equals between 0.5 and 4, UCB will get a better result.

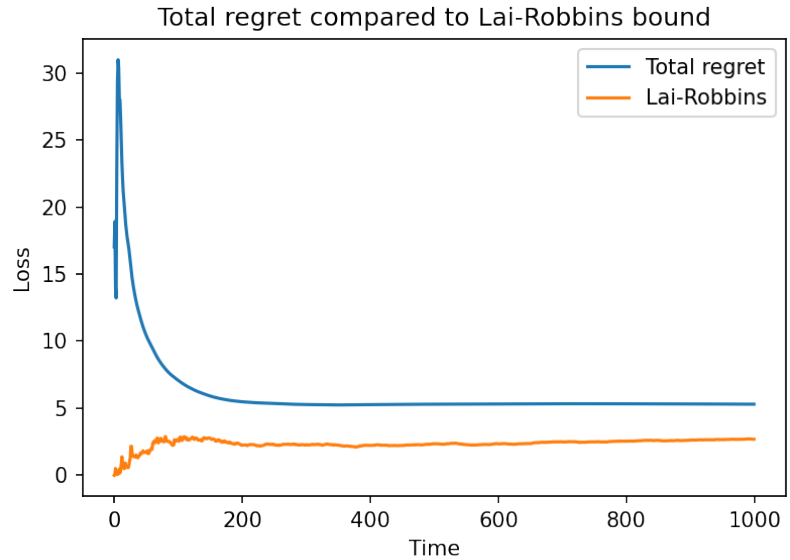



Figure 2: Comparison between expected total regret and Lai-Robbins for UCB

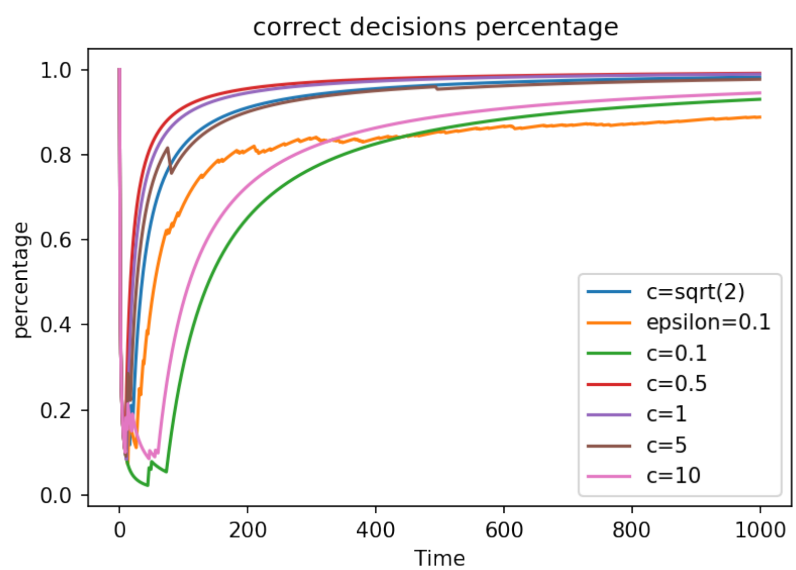



Figure 3: Percentage correct decisions for $\epsilon$-greedy and UCB

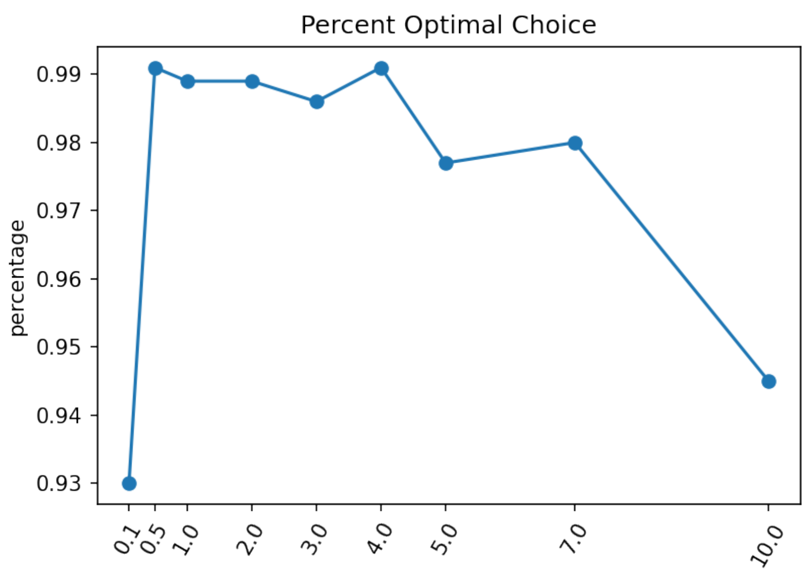



Figure 4: Optimal percentage correct decisions for different UCB hyper-parameter c