Stochastic models and optimization - Problem set 2

Hrvoje Stojic

May 3, 2021

This is a programming assignment in which you should work on your own. The goal is to get some hands on experience with bandit algorithms. You can choose between R and Python for the assignment. You are not allowed to use any of the functions from the packages that have already coded up these algorithms. I will evaluate your submission based on correctness of the implementation (and to some extent how well the code is written). Please make sure your solution is reproducible and operating system independent, I should be able to rerun your code and obtain your results/figures ideally with a single line command.

In this problem your task will be to reproduce a figure from an article by Chapelle and Li (2011) "An Empirical Evaluation of Thompson Sampling". Note that you don't need to read the whole article, pages 1 to 3 will suffice.

The figure that you will have to reproduce is Figure 1. The Bernoulli bandit problem for this figure has K arms where the best arm has a reward probability of 0.5 and the $K-1$ other arms have a probability of $0.5-\epsilon$. There will be four different computational experiments where you will apply algorithms in Bernoulli bandits with different K and ϵ parameters: $K \in \{10, 100\}$ and $\epsilon \in \{0.02, 0.1\}$.

You will have to program a UCB algorithm tuned for a Bernoulli bandit problem and a Thompson sampling algorithm. You will also compute an asymptotic lower bound for a Bernoulli problem (hint: there is an exact formula for computing KL divergence between two Bernoulli distributions). You can find these algorithms in the slides, but check the details in the article as well.

There should be four figures in total, one for each computational experiment, with cumulative regret on y-axis and log-transformed steps/trials on x-axis. You will need to simulate agents for 1000000 trials to get good trends on $K = 10$ and 10000000 on $K = 100$ and run at least 10 simulations to get more reliable estimates for each algorithm.

* If you feel adventurous and cannot get enough of bandits, add Exp3 algorithm to the mix and compare it with UCB and Thompson sampling.
* In case you find it computationally challenging to finish all four conditions, complete only the two conditions with $K = 10$ (without any penalty).

Don't hesitate to send me a message if anything is unclear.

Deadline is **May 14, 23:59 BCN time**