Stochastic models and optimization - Bandits - Problem set 1

Hrvoje Stojic

May 18, 2022

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 (partially). 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 $0.5-\epsilon$. Figure 1 displays four different computational experiments where UCB and Thompson sampling algorithms are applied in Bernoulli bandits with different K and ϵ parameters: $K \in \{10, 100\}$ and $\epsilon \in \{0.02, 0.1\}$. You should replicate only the condition with $K = 10$ and $\epsilon = 0.1$ to reduce the amount of computation.

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 should follow the paper for exact formulations of the algorithms.

To be clear, you should produce a figure 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 run at least 10 simulations to get more reliable estimates for each algorithm.

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

Deadline is **June 3, 23:59 BCN time**. Submit your code and figures by uploading them to Classroom problem set assignment.