In this assignment, you will be exploring algorithms for tackling the *multi-armed bandit problem* — a simple setting for studying the exploration-exploitation trade-off in reinforcement learning. Start by reading the introduction section of the following paper by Bubeck and Cesa-Bianchi, to learn more about the multi-armed bandit problem and its applications. You can stop reading once the content starts getting technical.

`http://homes.di.unimi.it/~cesabian/Pubblicazioni/banditSurvey.pdf`

In 2002, Auer et. al. published their seminal paper that introduced and analyzed several optimal algorithms for this problem.

`http://homes.di.unimi.it/~cesabian/Pubblicazioni/ml-02.pdf`

In this assignment, your task is to reproduce a subset of the experimental results presented in section 4 of the above paper. Specifically, here is the restricted experimental set-up for this assignment:

- You are only required to implement the UCB1 and ϵ_n-GREEDY algorithms. These algorithms are described in figures 1 and 3 in the paper, on pages 237 and 239 respectively. Note that Auer et. al. use a variant of UCB1 called UCB1-TUNED in their evaluations. You do *not* have to implement UCB1-TUNED — in your experiments, just use vanilla UCB1.

- Auer et. al. use 7 different reward distributions in their evaluations (see table on page 245). You only need to run experiments on 4 distributions — distributions 1, 3, 11, and 14 as defined in the paper.

You may implement one or more of the other algorithms detailed in the paper (UCB2, UCB1-TUNED, UCB1-NORMAL) for extra-credit.

The Write-Up

Remember the overarching writing rule for this course: *you need to be sufficiently precise with your writing and include enough details that a competent reader could reproduce your results.* Here are some specific things to address in your write-up, in no particular order. This is *not* meant to be an exhaustive list.

- Remember to introduce and define any new technical terms — for example, multi-armed bandit, regret, etc.

- Did you deviate from the experimental set-up described in the paper? Did you make any additional assumptions to account for any details missing from the paper?
- Don't just present your results, but analyze them. Did you encounter any surprises? Or do the results mostly align with those of Auer et. al.?
- Don't forget your citations!

Deliverables

Upload your code archive and your write-up as a single zip file to Moodle by 11:55pm on Friday, April 3. Due to the intervention of Easter Break, there will not be an opportunity for peer review — your April 3 submissions will be your final write-ups.