BIO 450/IBS 534
Spring 2025
Emory University

Reinforcement learning homework: 2-armed bandit (100 points total)

1. What is a reward prediction error? How is it calculated? (**10 points**)

2. In a Q-learning model, how do prediction errors relate to Q values? (**10 points**)

In the file Qlearning_driver.m, update line 7 with the location that you saved the folder BIO_450_RL. Qlearning_driver.m is the script that you will run to fit the model.

The RPEs, Q values, and probabilities are updated in FitSimpleQModel.m. **The script is incomplete!** Complete lines 21 and 22 to calculate the PE and update the Q value for the chosen option. Make sure the "subjects" (or files) that you want to include are listed (uncommented) in Qlearning_driver.m, and run the script to fit the model.

Note: Qlearning_driver.m, is set to clear your MATLAB workspace.

3. <u>Learning in a stable environment (Experiment 1)</u>. Each simulated data file represents a single subject with 200 trials of a 2-armed bandit task with binary outcomes (win $1, win nothing). Fit the model to the data from the 5 "subjects" from Experiment 1 (SampleDataExp1_1b-5b) and examine the output, displayed in the MATLAB window.

Include the best-fitting parameters and fit for each subject. (**15 points**)

Note for a-e. If you run the script with the option "`makeplots = 1;`" you will see 3 plots for each subject showing the RPE, Q value, and probability of choosing each option on each trial. The values for option/action 1 are blue, and the values for option/action 2 are red.

    a. Generally, how do RPE and Q values change over time? (**5 points**)

    b. What types of trials have the largest positive prediction error? Negative prediction error? (**5 points**)

    c. How would you describe the behavior of the subjects with the lowest learning rates? (**5 points**)

    d. How would you describe the behavior of the subjects with the highest learning rates? (**5 points**)

e. From the data, what would you estimate as the probability of reward for Option 1? Option 2? (**5 points**)

f. What would you expect behavior to look like with a learning rate of 1? A learning rate of 0? (**5 points**)

4. <u>Learning in a volatile environment (Experiment 2)</u>: Each data file contains 200 simulated trials from a 2-armed bandit task with binary outcomes (win $1, win nothing) with reversals (i.e. the option that more frequently gives rewards reverses throughout the task). Fit the model to the data from 5 subjects from Experiment 2 (SampleDataExp2_1b-5b) and examine the output.

Include the best-fitting parameters and fit for each subject. (**15 points**)

a. Looking at the data, how would you describe the timing of reversals? (i.e. when do you think they occur?) (**10 points**)

b. How does the environment affect optimality of learning rates? Is there a benefit to a lower learning rate? A higher learning rate? (**10 points**)