

Solving the n -armed bandit problem:

Recall that we had two competing properties by which we're judging the "goodness" of the algorithm:

- Get good estimates of the payoffs of the machines.
- Maximize our payoff (always choose the best machine).

Download the four files, `banditS.m`, `softmax.m`, `banditW.m`, `winstay.m`. The files `banditS` and `banditW` are the drivers for the softmax and "Win-Stay, Lose-Shift" algorithms (respectively). Here are some implementation details that we did not get a chance to discuss in class:

- We'll play 10 machines 1000 times (but these are modifiable).
- The drivers will set the actual machine payoffs to random numbers- Every time you run the driver, you'll have different results.
- We need to change either the parameter τ or the parameter β from some initial value to some final value. Here's a nice way to do that, that is implemented in the drivers: If you want a parameter to change from α_{initial} to α_{final} for t running from 0 to N , set the parameter value at time t to:

$$\alpha_t = \alpha_{\text{initial}} \left(\frac{\alpha_{\text{final}}}{\alpha_{\text{initial}}} \right)^{\frac{t}{N}}$$

You should verify that $\alpha_0 = \alpha_{\text{initial}}$ and that $\alpha_N = \alpha_{\text{final}}$. (NOTE: We'll start t and 1 instead of 0, but that's a small issue).

- When the driver program finishes, there are three things that are output:
 - The index of the machine with the (actual) best payoff (This is the machine we should've been playing most often).
 - The number of times each machine has been played.
 - A plot of the actual payoffs (black) and the algorithm's estimates of those payoffs (red).

Homework:

Run each of the two algorithms several times and comment on how each does with respect to our two "goodness" criteria. Next, try changing the parameter values τ and β to see if the algorithms produce the same types of results (qualitatively)- Comment on the *stability* of the algorithms (that is, an algorithm is stable if small changes of the parameters do not result in changes on the final result).