Multi-Armed Bandit

Colin Jemmott
DSC 96

Results from the widget factory

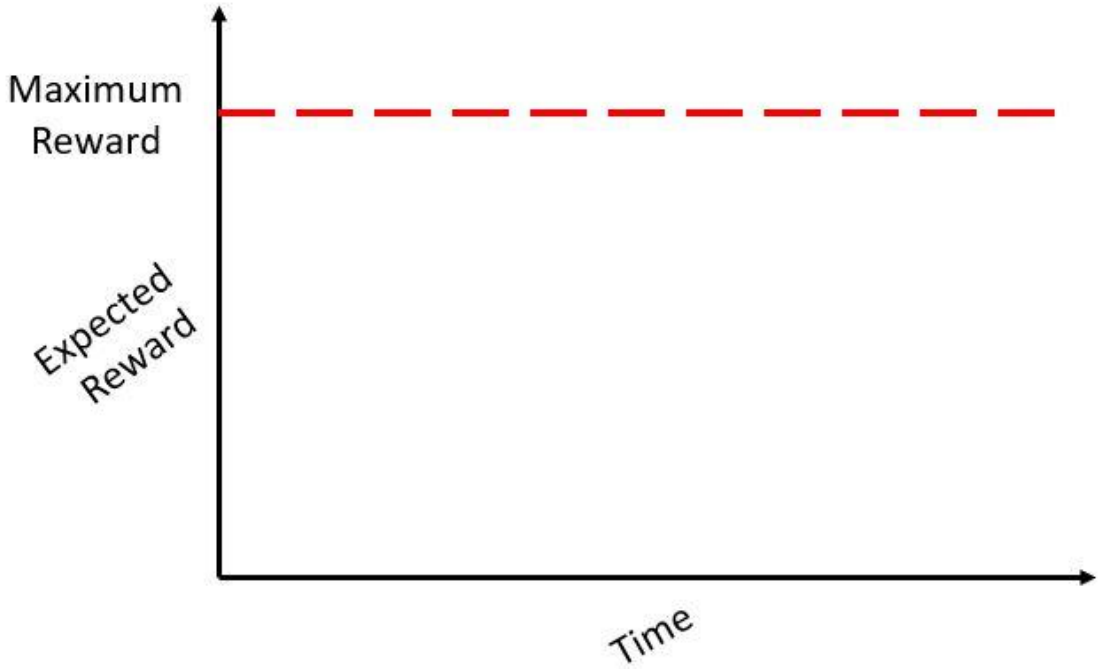
From the Jupyter notebook last time:

- Widget line B was 10% better than A. You might have found significance depending on:
  - The sensitivity you chose (which determined how long you ran the experiment)
  - Luck of random numbers
- Widget line C was terrible. But because the manager was *less* sure, you ran the experiment longer!

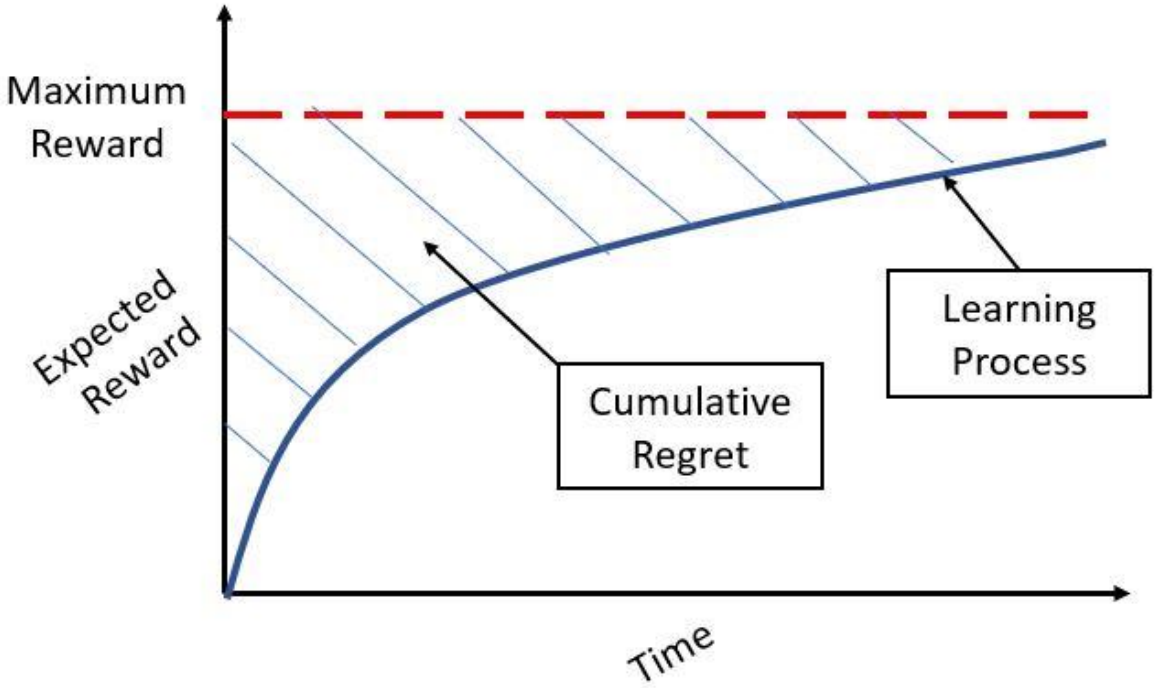
There must be a better way...

Multi-Armed Bandits

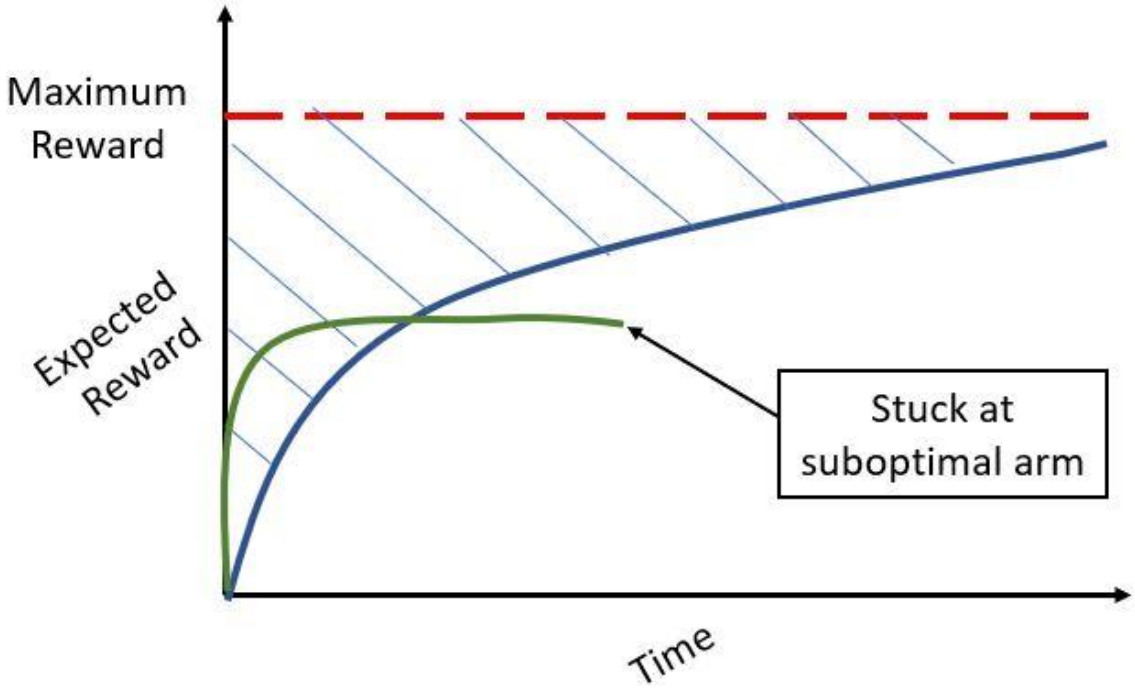
Perfect Knowledge



Exploring causes regret



Not exploring enough causes more regret



Explore then greedy

Hypothesis testing approach:

1. Run each experiment *N* times.
2. Choose a winner, and use that forever.

How to determine *N*?

What might go wrong with this approach?

- *N* too small and you might choose the wrong winner
- *N* too large and you spend too long experimenting

Upper Confidence Bound

j=1,...,K possible actions

Each action has a stationary random reward between 0 and 1

Choose the action that maximizes:

$$m_j + \sqrt{\frac{2 \ln N}{N_j}}$$

Other Approaches

- Epsilon Greedy
- Softmax Exploration
- Decayed Epsilon Greedy
- Thompson Sampling

Common Pitfalls of A/B Tests

1. Optimizing for the wrong metric
2. Failing to correctly randomize
3. Doing the math wrong
4. Stopping early
5. Unethical testing
6. Non-stationary problem