

Exploring the explore-exploit trade-off

Matthew Allcock - University of Sheffield

1 Plan

1.1 Details

This competition is open to Early Career Mathematicians (ECMs), with a prize of £250. We invite articles on any mathematical topic, including pure, applied, teaching, case studies, etc. between 1,000–2,000 words.

1.2 Possible titles

- Dora the exploit or explorer
- Dora the exploiter
- Dora the explorer or exploiter
- When should Dora explore?
- Exploring the explore-exploit trade-off
- To explore or to exploit

1.3 Structure

- Abstract
- Introduction: Examples of explore/exploit - Exploring possible partners, restaurants in a new city, career options to follow, reinforcement learning. We want a decision theory for when to explore and when to exploit

- modelling the trade-off - Multi-armed bandit problem.
- Lower bound - naive approach - see HAAISS notes. Expected value of information for next turn. Explain why this is a lower bound.
- Upper bound - Expected value of perfect information
- True expected value
- Markov chain monte carlo simulation - link to github repo?
- Conclusion

2 Abstract

3 Introduction

We are often faced with a decision between several options where we are uncertain as to how good each option is. We can *explore* the space of options to learn more about . However, exploring comes at a cost of time, money, energy, or all three. At some point, we will want to *exploit* the best option we know about, avoiding the costs of further exploration and reaping the rewards of what we believe to be the best option. Explore too long and you pay the opportunity cost of not exploiting the best known option. Exploit too early and you might be committing to a sub-optimal choice.

Let's say you move to a new city and want to find a good hair-dresser. One option is to try them all sequentially, then choose the one that gave you the best haircut to be your regular. This would work in a small town, where the option-space is small. You might only need to try three hair-dressers to exhaust all the options. You'd better hope that one of them did a good job. What if you move to a large city, where you have a hundred to choose from? It would take many years to explore them all. How should you choose when to stop exploring and to start getting your haircuts from the best known hair-dresser?

But it's not just a fun exercise in choosing who's going to make you look sharp. Understanding the explore-exploit trade-off is important for life biggest decisions. How long should you spend exploring different career options before going all-in on the one that think you can be most successful

in? How many people should you date before committing to a long-term partnership? How long should a mathematician explore the space of possible open research questions before tackling one? All of these examples involve a trade-off between exploring the space of options and exploiting the best known option.

Intuitively, we would expect that the optimum decision procedure is to explore the option-space until the expected value of exploring another option is less than the incurred cost. Let's put this trade-off on a mathematical footing.

4 Modelling the trade-off

5 When to draw: lower bound

6 When to draw: upper bound

7 Simulating the trade-off

8 Conclusion