Sequential decision making: modeling how we interact with the world

Alan Malek

April 9th, 2016

- Omer Atun: CEO of AgilOne Predictive Marketing
- Brienne Ghafouriar: CEO of Entefy
- Jeffrey Rothschild: founder Veritas Software and Mpath Interactive, Facebook VP of Infrastructure Software

- Omer Atun: CEO of AgilOne Predictive Marketing
- Brienne Ghafouriar: CEO of Entefy
- Jeffrey Rothschild: founder Veritas Software and Mpath Interactive, Facebook VP of Infrastructure Software

- Alan Malek: Graduate student

- Omer Atun: CEO of AgilOne Predictive Marketing
- Brienne Ghafouriar: CEO of Entefy
- Jeffrey Rothschild: founder Veritas Software and Mpath Interactive, Facebook VP of Infrastructure Software
- Alan Malek: Graduate student, Harker alumnus '05

Goals of this talk

- What is Grad school?
- Modern sequential decision making problems
- Whet your appetite with a cool problem
- Some advice: is grad school for you?

What is a Grad school?

What is a Grad school?

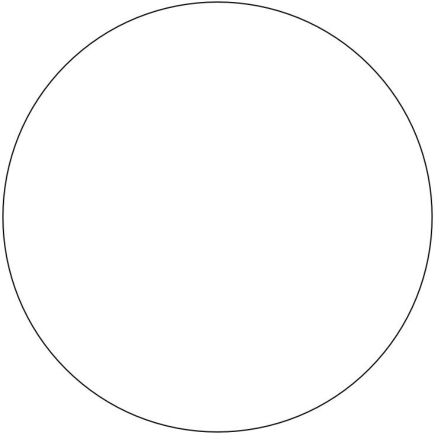
- A third the pay for a third the responsibilities

What is a Grad school?

- A third the pay for a third the responsibilities
- Take classes for a few years
- Be a TA
- Do research
  - In theory: think about how to formalize problems, prove theorems
  - Or more applied: engineer solutions

What is a PhD?

Imagine a circle that contains all of human knowledge:

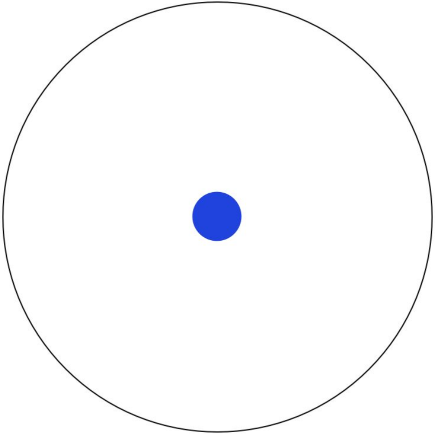


Credit: Matt Might

http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

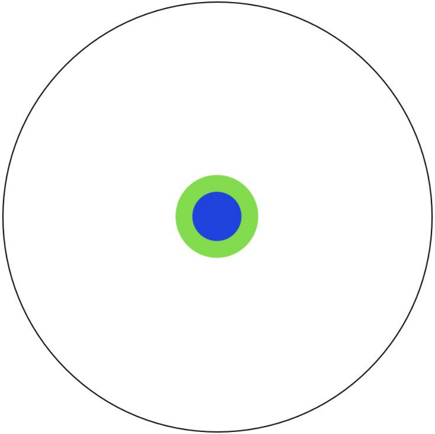
By the time you finish elementary school, you know a little:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

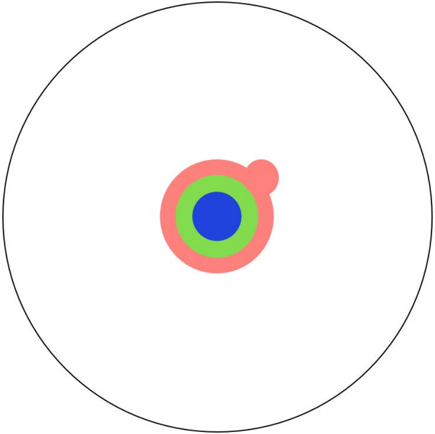
By the time you finish high school, you know a bit more:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

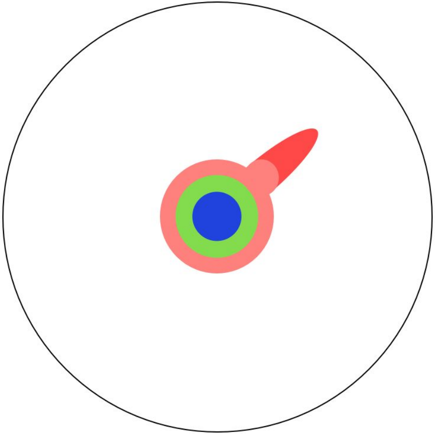
With a bachelor's degree, you gain a specialty:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

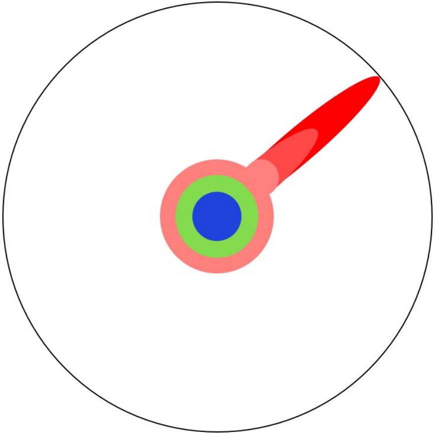
A master's degree deepens that specialty:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

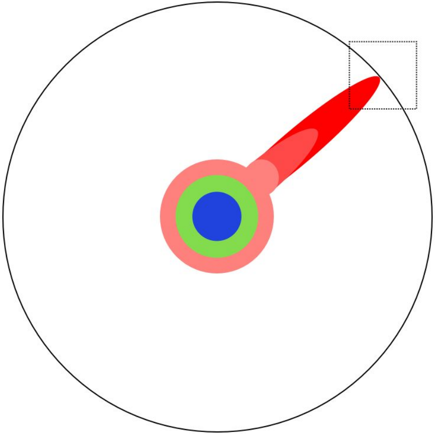
Reading research papers takes you to the edge of human knowledge:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

Once you're at the boundary, you focus:

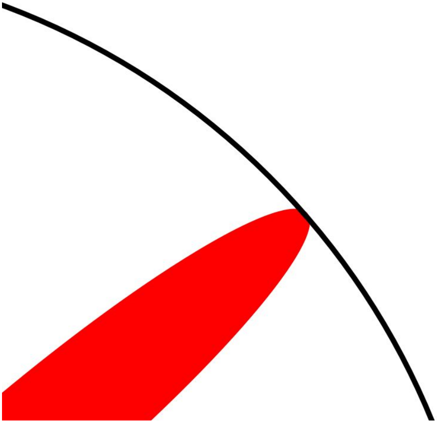


Credit: Matt Might

http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

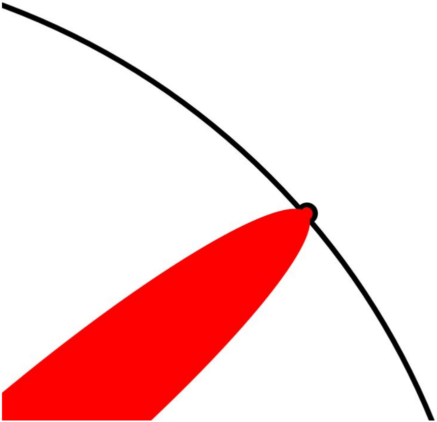
You push at the boundary for a few years:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

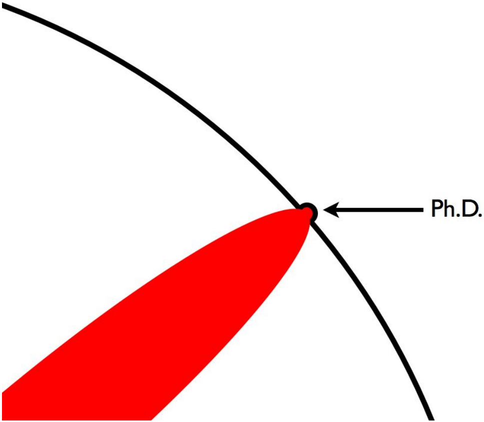
Until one day, the boundary gives way:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

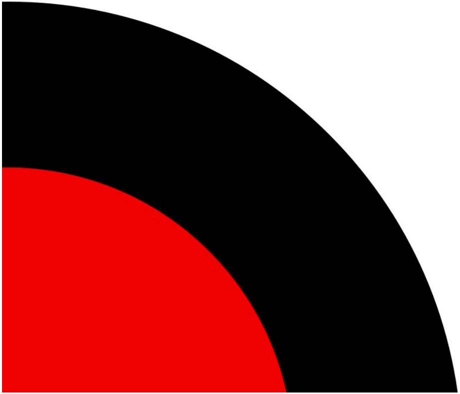
And, that dent you've made is called a Ph.D.:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

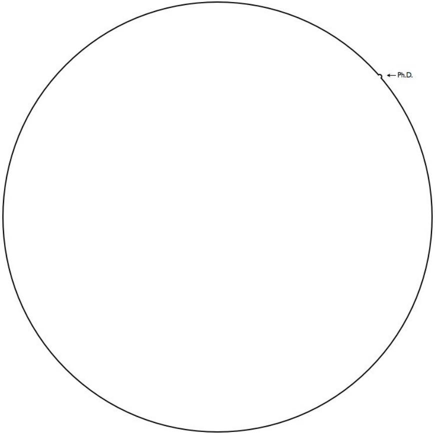
Of course, the world looks different to you now:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

What is a PhD?

So, don't forget the bigger picture:



Credit: Matt Might
http://matt.might.net/articles/phd-school-in-pictures/

Goals of this talk

- What is Grad school?
- Modern sequential decision making problems
- Whet your appetite with a cool problem
- Give some advice: how to prepare for grad school

Big Data

- Tons of data: 2.5×10^{18} bytes a day [1]
- 90% of the world's data created in the last 2 years [1]
- Google: 100 billion searches a month, half from mobile [2]

[1] http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[2] http://blogs.wsj.com/digits/2015/10/08/google-says-mobile-searches-surpass-those-on-pcs/

Big Data

- Tons of data: 2.5×10^{18} bytes a day [1]
- 90% of the world's data created in the last 2 years [1]
- Google: 100 billion searches a month, half from mobile [2]
- Personalization: Andrew Ng's Coursera example

[1] http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[2] http://blogs.wsj.com/digits/2015/10/08/google-says-mobile-searches-surpass-those-on-pcs/

Other side: Interactions

- Ubiquity of devices

[3] Pouchter, Jacob "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies, Pew Research Center". Pewglobal.org. Retrieved 2016-02-23.

Other side: Interactions

- Ubiquity of devices
- Smartphone adoption rates:[3]

Rank	Country	% adoption
1	South Korea	88
2	Australia	77
3	Israel	74
4	United States	72
5	Spain	71

[3] Pouchter, Jacob "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies, Pew Research Center". Pewglobal.org. Retrieved 2016-02-23.

Other side: Interactions

- Ubiquity of devices
- Smartphone adoption rates:[3]

Rank	Country	% adoption
1	South Korea	88
2	Australia	77
3	Israel	74
4	United States	72
5	Spain	71

- Sequential Interaction

[3] Pouchter, Jacob "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies, Pew Research Center". Pewglobal.org. Retrieved 2016-02-23.

Other side: Interactions

- Ubiquity of devices
- Smartphone adoption rates:[3]

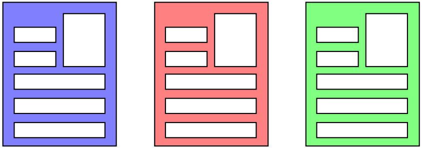
Rank	Country	% adoption
1	South Korea	88
2	Australia	77
3	Israel	74
4	United States	72
5	Spain	71

- Sequential Interaction
- Thought experiment: how did people arrange to meet before cell phones?

[3] Pouchter, Jacob "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies, Pew Research Center". Pewglobal.org. Retrieved 2016-02-23.

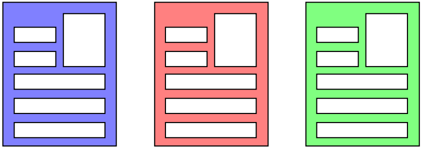
Sequential Decision Making - running a newspaper

- First problem: which headline to choose?



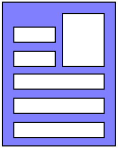
Sequential Decision Making - running a newspaper

- First problem: which headline to choose?



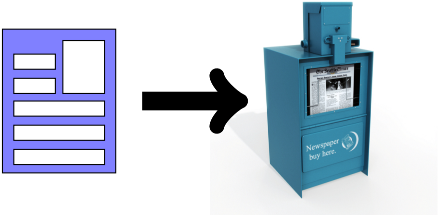
- Goal: pick the headline that sells the best

Print Newspapers



1. Choose headline

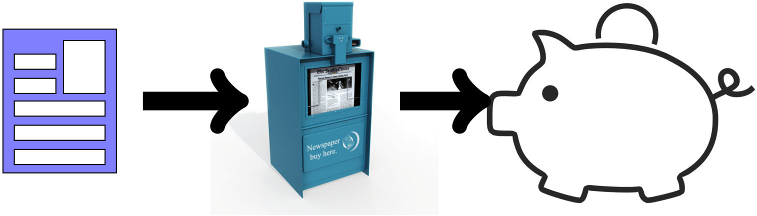
Print Newspapers



1. Choose headline
2. Sell paper

Too late: by the time feedback comes, your headlines are stale.

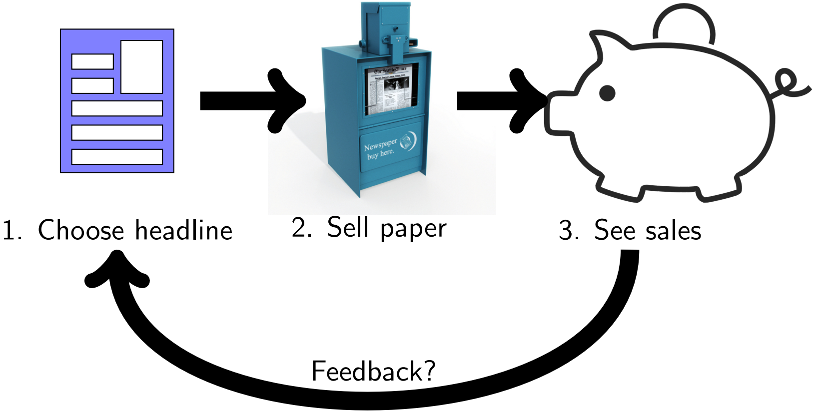
Print Newspapers



1. Choose headline
2. Sell paper
3. See sales

Too late: by the time feedback comes, your headlines are stale.

Print Newspapers

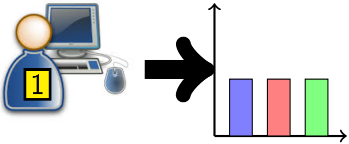


Too late: by the time feedback comes, your headlines are stale.

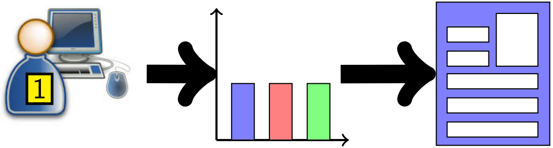
Internet

1. User arrives

Internet

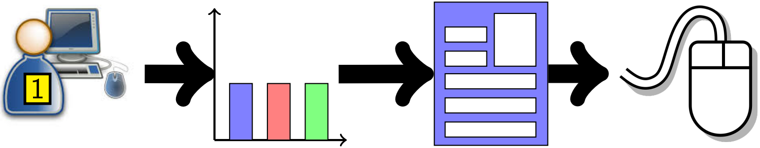


1. User arrives 2. Choose article

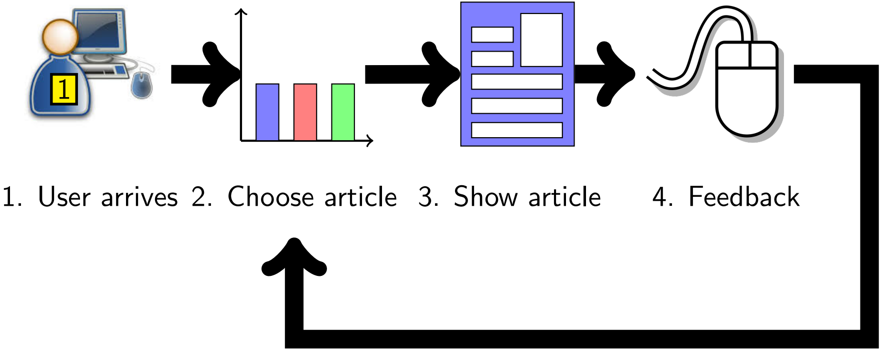


1. User arrives 2. Choose article 3. Show article

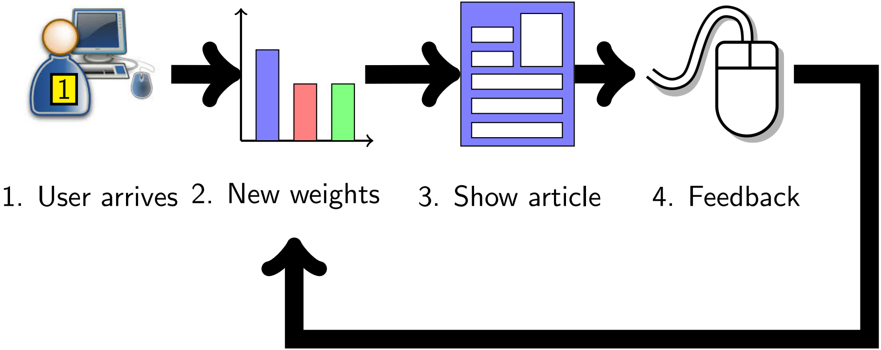
Internet



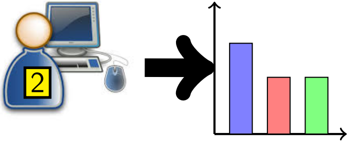
1. User arrives 2. Choose article 3. Show article 4. Feedback


1
1. User arrives
2. Choose article
3. Show article
4. Feedback

Internet

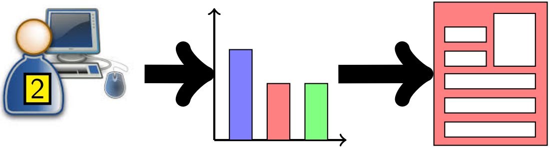


Internet



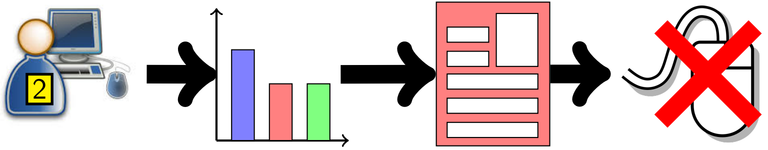
1. User arrives 2. New weights

Internet



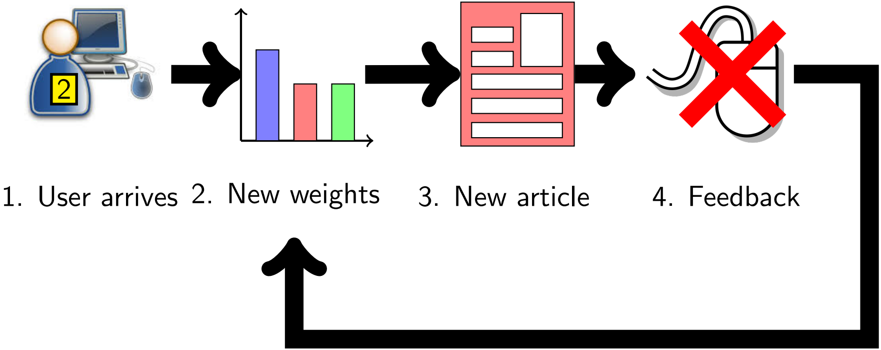
1. User arrives 2. New weights 3. New article

Internet

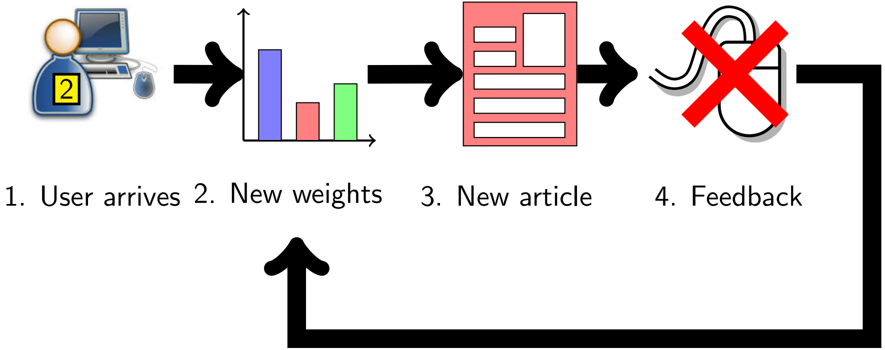


1. User arrives 2. New weights 3. New article 4. Feedback

Internet



Internet



Sequential problem

- Unlike newspaper stand, we get repeated feedback
- Can modify our choices in *real time*

exploration vs. **exploitation**

Sequential problem

- Unlike newspaper stand, we get repeated feedback
- Can modify our choices in *real time*

exploration vs. **exploitation**

Continue to explore headlines until we are very sure we haven't missed the best one.

Sequential problem

- Unlike newspaper stand, we get repeated feedback
- Can modify our choices in *real time*

exploration vs. **exploitation**

Continue to explore headlines until we are very sure we haven't missed the best one.

Use the headline that has been the best so far.

Sequential problem

- Unlike newspaper stand, we get repeated feedback
- Can modify our choices in *real time*

exploration vs. **exploitation**

Continue to explore headlines until we are very sure we haven't missed the best one.

Use the headline that has been the best so far.

- How can we formalize this?

Multi-Armed Bandit

Given: game length T, arms $1, \ldots, K$
For $t = 1, 2, \ldots, T$:

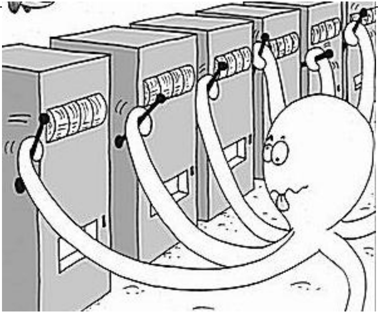
- Adversary chooses rewards $r(t, k) \in [0, 1]$
- Learner chooses an arm k_t
- Learner gets reward $r(t, k_t)$

Figure: game protocol

Multi-Armed Bandit

Given: game length T, arms $1, \ldots, K$
For $t = 1, 2, \ldots, T$:

- Adversary chooses rewards $r(t, k) \in [0, 1]$
- Learner chooses an arm k_t
- Learner gets reward $r(t, k_t)$



Multi-Armed Bandit

Given: game length T, arms $1, \ldots, K$
For $t = 1, 2, \ldots, T$:

- Adversary chooses rewards $r(t, k) \in [0, 1]$
- Learner chooses an arm k_t
- Learner gets reward $r(t, k_t)$

Learner does not see rewards for other actions

Figure: game protocol

Multi-Armed Bandit

Given: game length T, arms $1, \ldots, K$
For $t = 1, 2, \ldots, T$:

- Adversary chooses rewards $r(t, k) \in [0, 1]$
- Learner chooses an arm k_t
- Learner gets reward $r(t, k_t)$

Learner needs to randomize

Learner does not see rewards for other actions

Figure: game protocol

Multi-Armed Bandit

Given: game length T, arms $1, \ldots, K$
For $t = 1, 2, \ldots, T$:

- Adversary chooses rewards $r(t, k) \in [0, 1]$
- Learner chooses an arm k_t
- Learner gets reward $r(t, k_t)$

Learner needs to randomize

Learner does not see rewards for other actions

Figure: game protocol

$$\mathrm{Regret}_{\mathrm{T}} = \underbrace{\max_{k'} \sum_{t=1}^{T} r(t, k')}_{\text{reward of best arm}} - \underbrace{\sum_{t=1}^{T} r(t, k_t)}_{\text{Learner's reward}}$$

Simple problem, but already interesting

Given: game length T, arms $1, \ldots, K$
For $t = 1, 2, \ldots, T$:

- Adversary chooses rewards $r(t,k) \in [0,1]$
- Learner chooses an arm k_t
- Learner gets reward $r(t, k_t)$

Figure: game protocol

- Exploitation vs. exploration
- Stochastic vs. Adversarial data

Stochastic vs. Adversarial data

- Stochastic
  - Most of statistics, machine learning
  - Assume that data are generated from some random variable
  - E.g. arm k has mean μ_k and

$$r(t, k) = \begin{cases} 1 & \text{with probability } \mu_k \\ 0 & \text{with probability } 1 - \mu_k. \end{cases}$$

  - Past data and future data are similar

Stochastic vs. Adversarial data

- Stochastic
  - Most of statistics, machine learning
  - Assume that data are generated from some random variable
  - E.g. arm k has mean μ_k and

$$r(t,k) = \begin{cases} 1 & \text{with probability } \mu_k \\ 0 & \text{with probability } 1 - \mu_k. \end{cases}$$

  - Past data and future data are similar
- Adversarial
  - *No* assumptions on $r(t,k)$; much harder
  - Adversary could choose $r(t,k)$ based on your choices for time $1, \ldots, t-1$ to make you do badly
  - Need regret; minimizing loss is hopeless

$$\mathrm{Regret_T} = \underbrace{\max_{k'} \sum_{t=1}^{T} r(t,k')}_{\text{reward of best arm}} - \underbrace{\sum_{t=1}^{T} r(t,k_t)}_{\text{Learner's reward}}$$

Solutions to the bandit problem

- ϵ-greedy (warm-up)
- EXP3

ϵ-greedy

Given: parameter ϵ, arms $1, \ldots, K$

For $t = 1, 2, \ldots$:

- With probability ϵ, pick arm k_t uniform at random
- Otherwise, pick $k_t = \text{argmax}_k \sum_{s=1}^{t-1} r(s, k)$

ϵ-greedy

Given: parameter ϵ, arms $1, \ldots, K$
For $t = 1, 2, \ldots$:

- With probability ϵ, pick arm k_t uniform at random
- Otherwise, pick $k_t = \text{argmax}_k \sum_{s=1}^{t-1} r(s, k)$

- Easy to defeat with adversarial data
- Not even optimal for stochastic data

EXP3 algorithm

Given: parameter η, arms $1, \ldots, K$
Set: $w_i(1) = 1$ for $i = 1, \ldots, K$
For $t = 1, 2, \ldots$:

- Set $p_i(t) = \frac{w_i(t)}{\sum_{j=1}^K w_j(t)} \quad i = 1, \ldots, K$
- Draw k_t randomly proportional to $p_1(t), \ldots, p_K(t)$
- Get reward $r(t, k_t)$
- For $i = 1, \ldots, K$ set

$$\hat{r}(t,i) = \begin{cases} r(t,j)/p_j(t) & \text{if } i = k_t \\ 0 & otherwise, \end{cases}$$

$$w_j(t+1) = w_j(t)\exp(\eta\hat{r}(t,i)) = \exp\left(\eta\sum_{s=1}^{t}\hat{r}(s,i)\right)$$

EXP3 (weak) regret theorem

Theorem

The EXP3 algorithm has the following bound:

$$\mathbb{E}\left[\mathrm{Regret_T}\right] = \max_{k'} \sum_{t=1}^{T} r(t,k') - \sum_{t=1}^{T} \mathbb{E}\left[r(t,k_t)\right]$$
$$\leq \frac{\eta TK}{2} + \frac{\log(K)}{\eta}.$$

If we tune γ *for* T*,*

$$\mathbb{E}\left[\mathrm{Regret_T}\right] \leq \sqrt{2KT\log(K)}.$$

EXP3 (weak) regret theorem

Theorem

The EXP3 algorithm has the following bound:

$$\mathbb{E}\left[\mathrm{Regret_T}\right] = \max_{k'} \sum_{t=1}^{T} r(t,k') - \sum_{t=1}^{T} \mathbb{E}\left[r(t,k_t)\right]$$
$$\leq \frac{\eta TK}{2} + \frac{\log(K)}{\eta}.$$

If we tune γ for T,

Regret per round $\to 0$

$$\mathbb{E}\left[\mathrm{Regret_T}\right] \leq \sqrt{2KT\log(K)}.$$

EXP3 (weak) regret theorem

Theorem

The EXP3 algorithm has the following bound:

$$\mathbb{E}\left[\mathrm{Regret_T}\right] = \max_{k'} \sum_{t=1}^{T} r(t,k') - \sum_{t=1}^{T} \mathbb{E}\left[r(t,k_t)\right] \leq \frac{\eta TK}{2} + \frac{\log(K)}{\eta}.$$

If we tune γ for T,

Regret per round $\to 0$

$$\mathbb{E}\left[\mathrm{Regret_T}\right] \leq \sqrt{2KT\log(K)}.$$

lower bound of $\Omega(\sqrt{KT})$

Recap

- Started with real world problem
- Abstracted into Multi-Armed Bandit framework
- Proposed algorithms
- Proved upper bounds on their regret
- Compared to lower bounds

Goals of this talk

- What is Grad school?
- Modern sequential decision making problems
- Whet your appetite with a cool problem
- Give some advice: how to prepare for grad school

Things I wish I knew in college

Things I wish I knew in college

- Your professors want to talk to you and meet undergrads

Things I wish I knew in college

- Your professors want to talk to you and meet undergrads
- Go to their office hours, group meetings, ask about problems

Things I wish I knew in college

- Your professors want to talk to you and meet undergrads
- Go to their office hours, group meetings, ask about problems
- If research might be for you, get involved early (sophomore)

Things I wish I knew in college

- Your professors want to talk to you and meet undergrads
- Go to their office hours, group meetings, ask about problems
- If research might be for you, get involved early (sophomore)
- You will need three good letters to get a good grad school

Things I wish I knew in college

- Your professors want to talk to you and meet undergrads
- Go to their office hours, group meetings, ask about problems
- If research might be for you, get involved early (sophomore)
- You will need three good letters to get a good grad school
- Sometimes you will learn more from a research project than an extra class

Things I wish I knew in college

- Your professors want to talk to you and meet undergrads
- Go to their office hours, group meetings, ask about problems
- If research might be for you, get involved early (sophomore)
- You will need three good letters to get a good grad school
- Sometimes you will learn more from a research project than an extra class
- *Failing is part of it!*

Is a PhD for you?

Pros:

- Research is also rewarding; occasionally fun
- Very independent
- Become an expert in something!
- Exciting problems

Is a PhD for you?

Pros:

- Research is also rewarding; occasionally fun
- Very independent
- Become an expert in something!
- Exciting problems

Cons:

- Research is frustrating; many more failed attempts
- Long hours, little pay
- Few jobs require a PhD
- Don't escape politics

Thank you!