DATA 37200: Learning, Decisions, and Limits
(Winter 2025)

Lecture 1: Intro and the First Problem

Instructor: Haifeng Xu

Outline

- Course Overview
- Administrivia
- The First Problem

Recall: Classic Machine Learning Problems

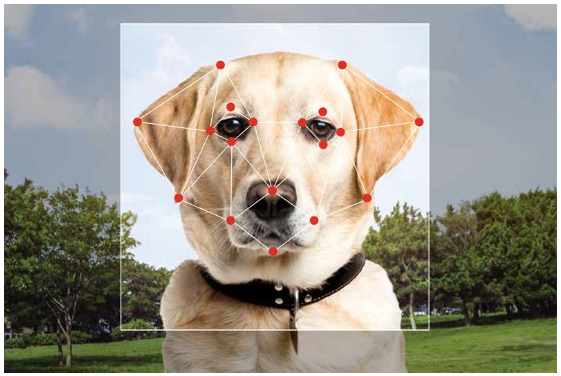
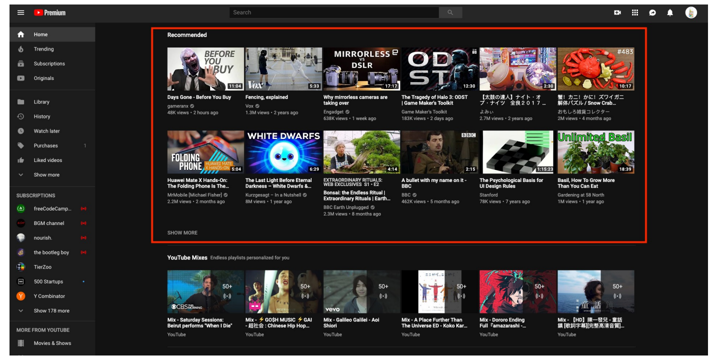
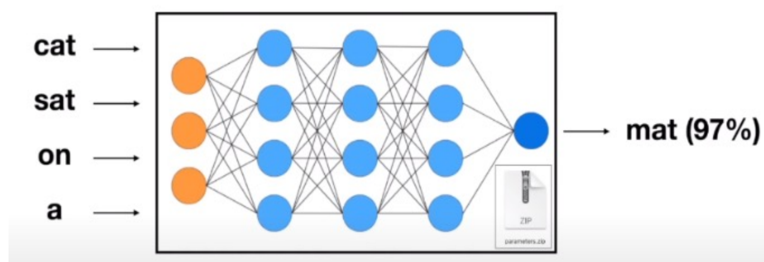


Image recognition



Preference learning for recommendations

Speech recognition



Next token/word prediction (for language models)

Recall: Classic Machine Learning Problems

These are recognition-based learning problems

- ➢ Task environments are usually static
- ➢ Often use supervised learning
- ➢ Relatively mature by now, and quite successful in both theory and applications

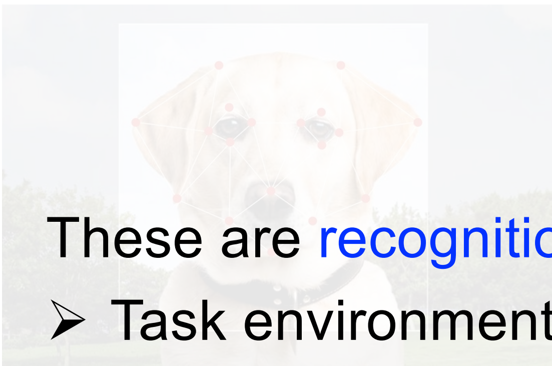
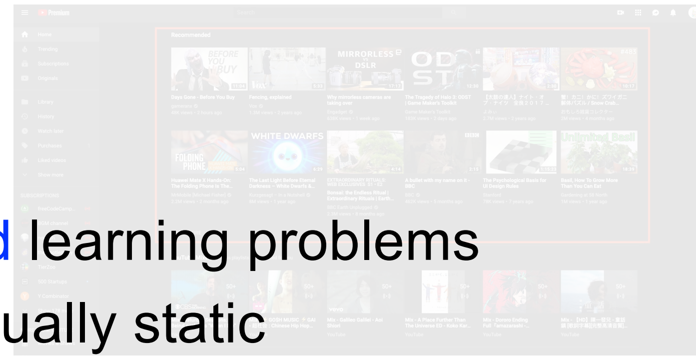
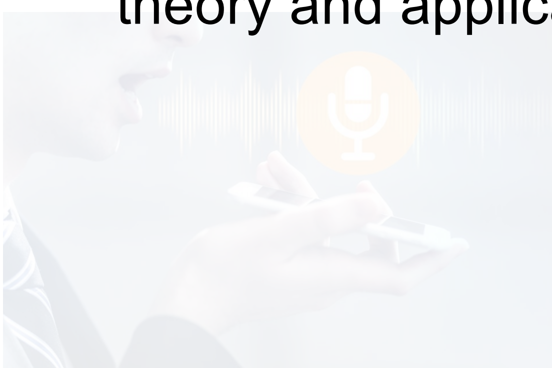


Image recognition



Preference learning for recommendations



Speech recognition

Next token/word prediction (for language models)

This Course: Decision-Based Learning Tasks

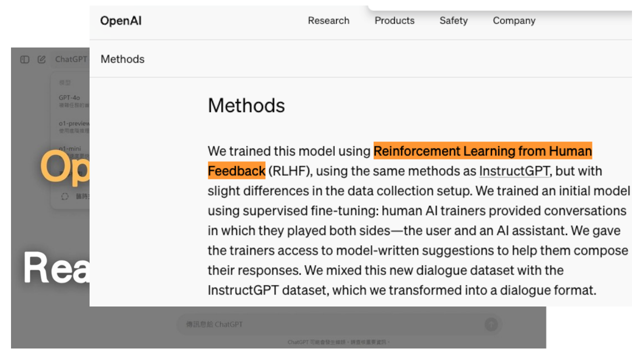
- Often use quite different design principles and learning techniques
  - Will see in our first learning problem why new design ideas are necessary
  - A well-known field studying this is reinforcement learning (RL), about which this course will cover a lot, though also beyond
  - Problems are often more complex

- Why more complex? To learn decisions, we have to consider many factors beyond just accuracy:
  - Rewards/payoffs/costs/utilities
  - Decision consequences – your learned decisions act on (hence change) the environments
  - Conflicting interests/incentives
  - Societal issues: fairness, alignment, welfare-efficient…

Why Important?

- Core *decision-based learning* techniques are underlying many breakthrough research

Deepmind's Alpha series



GPT-o1, even ChatGPT

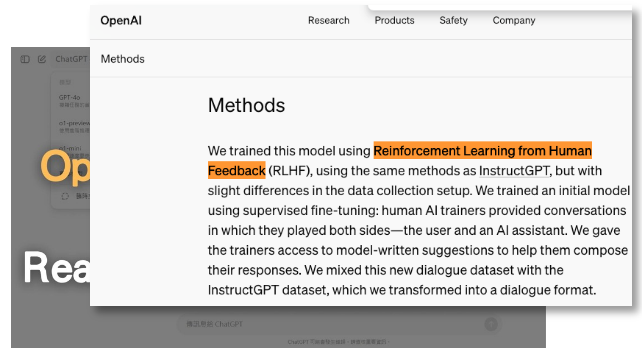
learn to decide next move, how to search, how to find next reasoning step

learn to find next reasoning step, to align with human's preferences/values/payoffs

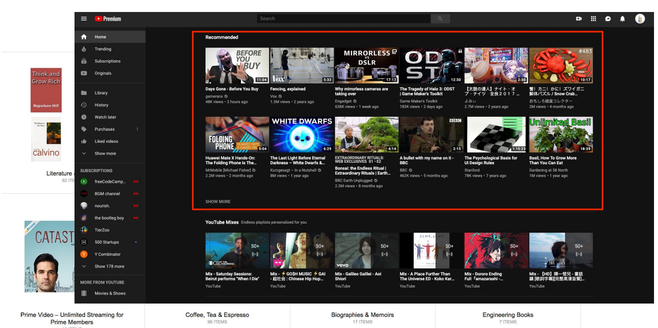
Why Important?

➢ Core *decision-based learning* techniques are underlying many breakthrough research and billions$-scale industrial applications

Deepmind's Alpha series



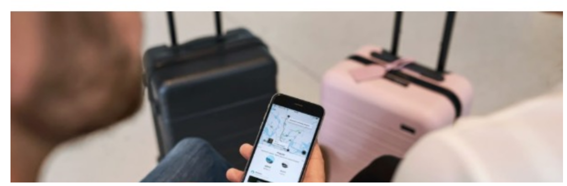
GPT-o1, even ChatGPT



Product/content recommendation

Products

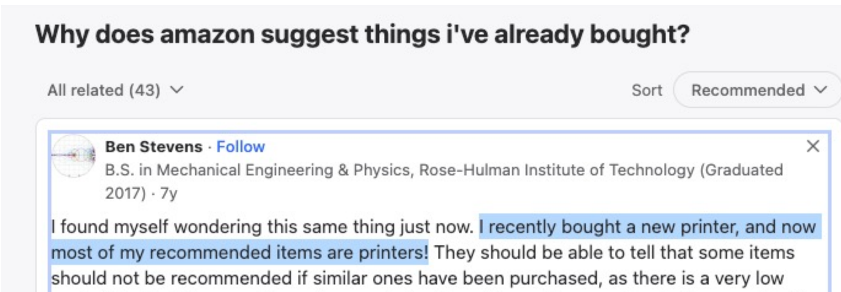
How Uber's dynamic pricing model works



Dynamic pricing based on traffic/supply/demand prediction

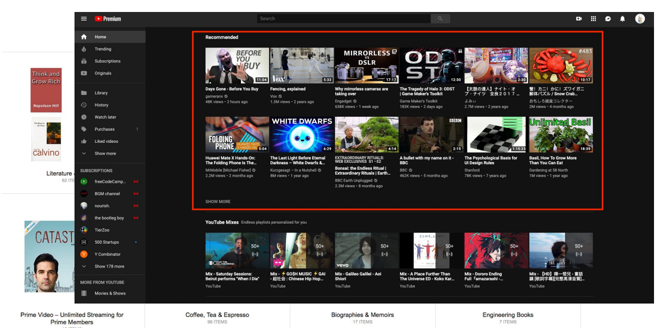
Why Important?

- Core *decision-based learning* techniques are underlying many breakthrough research and billions$-scale industrial applications

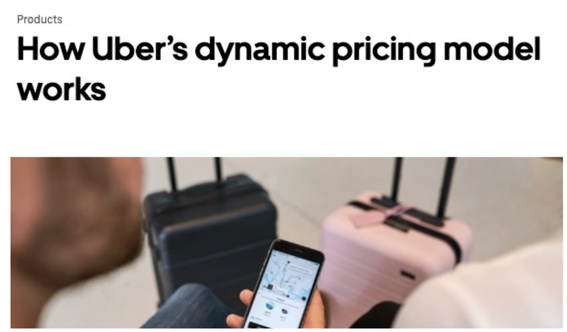


Recommendation needs consider its action consequences

Challenge: demand/supply → price → changes demand/supply



Product/content recommendation

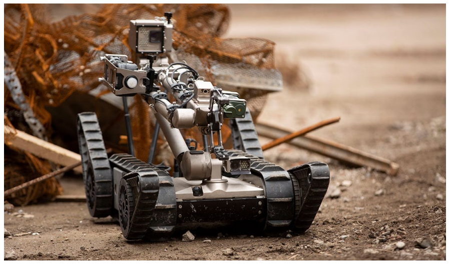


Dynamic pricing based on traffic/supply/demand prediction

Why Important?

➢ Core *decision-based learning* techniques are underlying many breakthrough research and billions$-scale industrial applications

Not to mention many data-driven policy/decision making problems in critical societal, health and security applications



Wow, cool! So… after this course, will I become the hero to work towards Nobel, or solving Google's/Amazon's problems?

- Not immediately…
  - Those are not easy problems to solve
  - This is designed to be a foundational (theory-focused) course
  - (Programming/implementation is also important, just not our focus)
- Goal of this course is to build your foundational understandings about
  - What key factors to consider while learning optimal decisions
  - Basic design principles of optimal learning algorithms
  - What is possible, and what is not possible
  - Along the way, also enrich your statistical and algorithmic toolkits

Learning Objectives

- Understand how to mathematically formulate and analyze models for interactive learning problems; learn how to apply core techniques from probability, statistics, optimization, etc.
- Understand key difficulties/challenges with solving RL problems
- Understand principles underlying relevant cutting-edge technologies, such as Reinforcement Learning from Human Feedback (RLHF) and AlphaGo training
- Be well-prepared to understand state-of-the-art papers about online learning, RL and data-driven decision making
- Have the foundations to work on relevant practical applications

Tentative Topics of the Course

- (week 1) Concentration bound, and UCB
- (week 2) Information-theoretic lower bound for KL and distribution testing
- (week 3-4) Elliptical potential lemma, and linear contextual bandits
- (week 5) Online learning, online gradient descent, reduction from contextual bandit to online learning
- (week 6) MDP, dynamic programming
- (week 6) Policy iteration and value iteration
- (week 7) Reinforcement learning and optimism principle
- (week 8) multi-agent RL, equilibria, counterfactual regret minimization, self-play
- (week 9) Sampled recent learning paradigms: RLHF, etc.

Targeted Audience of This Course

- Anyone planning to do research in machine learning (theoretical or empirical), particularly with human factors involved
  - The course is theory-focused, but we cover the very basics that even applied researcher should benefit from these basics
  - Even you do not work on interactive decision learning, it is still useful to see how it interplay with bandits, decisions.
- Anyone who wants to grasp the basics about how ML can be used for recommendation, preference alignment, dynamic pricing, etc.
- Anyone who want to see what other useful ML paradigms there are beyond supervised learning via large neural networks
  - Offer you a more comprehensive view about machine learning
  - Deep learning is super useful and powerful, but real industrial success also crucially hinges on other equally critical techniques

Outline

- Course Overview
- Administrivia
- The First Problem

Basic Information

- Course time: Tue/Thu, 12:30–1:50 pm at JCL 011
- Lecture: in person (unless further instruction)
- Instructor: Frederic Koehler (fkoehler@uchicago.edu) and Haifeng Xu (haifengxu@uchicago.edu)
  - Joint teaching due to new development
  - Office Hour: **Frederic (Tue 4:30-5:30 pm); Haifeng (Thur 4-5 pm)**
  - Can add more office hour, depending on demand
- TAs
  - Aditya Prasad; office hour: **Wed 2-3 pm**
- Couse website: https://frkoehle.github.io/data37200-w2025/
  - Easier way is to search our personal website and navigates to course
- References: linked papers/notes on website, no official textbooks
  - Slides will be posted *after* lectures

Prerequisites

- Mathematically mature: be comfortable with proofs
- Sufficient exposures to probabilities and algorithms/optimization
  - Algorithms (CMSC 27200/27220 or equivalent)
  - Linear algebra (CMSC 25300 or equivalent)
  - Probability (STATS 25100 or equivalent).
- If not sure, consult with the instructor. Note that no background on learning theory is required.

Requirements and Grading

- Part I (30%): 3~4 proof-based assignments

- Part II (45%): course project. Instructions will be posted on website later.
  - Team up: **up to 3** people per team
  - Make progress on a *research question* or *reproducing proofs* of existing papers, or a mixture
  - Deliverables: a presentation + a technical report in PDF
  - Grading is based on novelty + non-triviality

- Part III (25%): 3 in-class 30-mins quizzes
  - Not meant to be challenging
  - Just to check whether you are on top of key materials

Notes on Relevant Materials

- There are courses (and blogs) online that overlap with materials of this course
- These are great resources for extra reading, but it is still very useful for you to follow lectures as closely as you can because
  - Different instructors interpret the same knowledge differently
  - This will shape your *way of thinking* differently, which we think are the most valuable thing to learn from a course

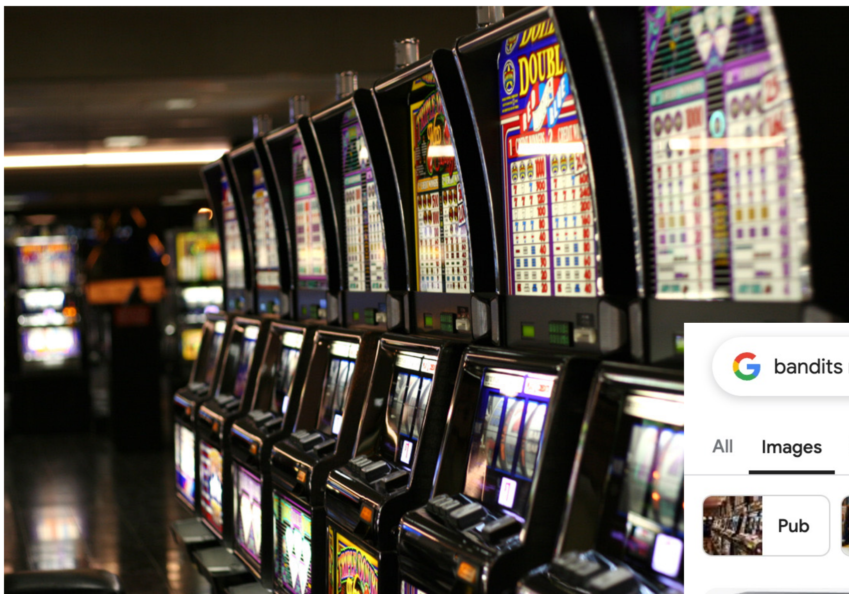
If you have any suggestions/comments/concerns, feel free to email us.

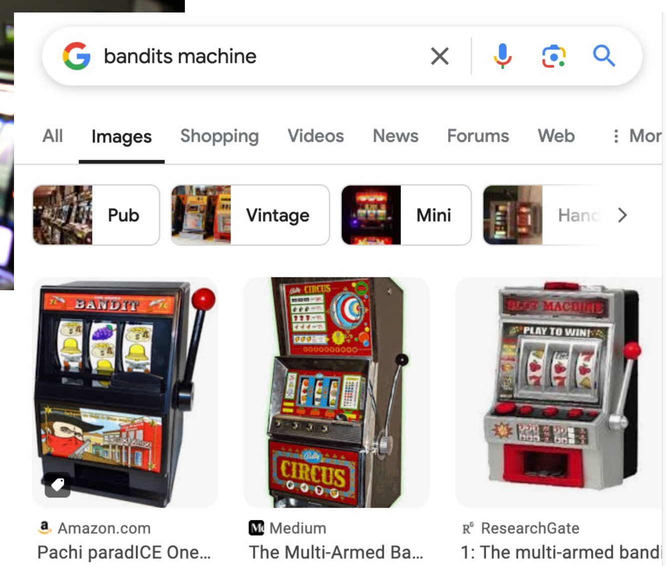
Outline

- Course Overview
- Administrivia
- The First Problem

The Stochastic Multi-Armed Bandit Problem

- Named after a gambling game
- A foundational RL problem with a simple and elegant formulation





The Stochastic Multi-Armed Bandit Problem

Formulation of the Multi-Armed Bandits (MAB)

$1 : r_1 \sim D_1$

$2 : r_2 \sim D_2$

. . .

$k : r_k \sim D_k$

- A set of k arms, denoted as $[k] = \{1,2, \cdots, k\}$
- Pulling arm i once generates a random reward r_i drawn from distribution D_i
  - Useful notations: let $\mu_i = \mathbb{E}[R_i]$ and $\mu^* = \max_{i \in [k]} \mu_i$
- As the algorithm designer, you decide which arm to pull to maximize your expected reward
  - This question is often asked in a "limited horizon" setting where you are allowed to play for T rounds
  - Assume 0 cost of pulling, which is without loss

The Stochastic Multi-Armed Bandit Problem

Formulation of the Multi-Armed Bandits (MAB)

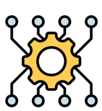
$1: r_1 \sim D_1$

$2: r_2 \sim D_2$

$\cdots$

$k: r_k \sim D_k$

- A set of k arms, denoted as $[k] = \{1,2,\cdots,k\}$
- Pulling arm i once generates a random reward r_i drawn from distribution D_i
  - Useful notations: let $\mu_i = \mathbb{E}[R_i]$ and $\mu^* = \max_{i\in[k]} \mu_i$
- As the algorithm designer, you decide which arm to pull to maximize your expected reward

Round	1	2	$\cdots$	t	$\cdots$	T
 Algorithm's choice	i^1	i^2		i^t		i^T

Goal: $\max_{i^1,\cdots,i^T} \mathbb{E}[\sum_{t=1}^{T} r_{i^t}]$

The Stochastic Multi-Armed Bandit Problem

Formulation of the Multi-Armed Bandits (MAB)

$1: r_1 \sim D_1$

$2: r_2 \sim D_2$

. . .

$k: r_k \sim D_k$

Question: if you know D_i (or even just $\mu_i = \mathbb{E}[R_i]$), what would be your optimal strategy?

Ans: always pull the $i^* = \arg\max_{i \in [k]} \mu_i$, with expected reward μ^*

👍 This achieves maximum possible expected reward $T\mu^*$

Round	1	2	$\cdots$	t	$\cdots$	T
Algorithm's choice	i^1	i^2		i^t		i^T

Goal:

$$\max_{i^1, \cdots, i^T} \mathbb{E}[\sum_{t=1}^{T} r_{i^t}]$$

The Stochastic Multi-Armed Bandit Problem

Formulation of the Multi-Armed Bandits (MAB)

$1: r_1 \sim D_1$

$2: r_2 \sim D_2$

$\cdots$

$k: r_k \sim D_k$

- Challenges arise when we do not know μ_i's, and need to learn from samples of D_i
- This leads to formulation of the MAB problem

Stochastic Multi-Armed Bandit (MAB)

Without knowing $\{\mu_i, D_i\}_{i=1}^{k}$, design a strategy/policy that chooses an arm sequence $i^1, i^2, \cdots, i^T$ to maximize $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$

The Stochastic Multi-Armed Bandit Problem

Why this is a learning problem?

- Do not know μ_i's in advance, hence need to learn them

Why this is not *just* a learning problem?

- Likely we need to learn μ_i's to some extent, but that's not final goal
- It is possible to achieve very high reward without needing to learn every μ_i well
- Btw, this makes a lot of sense in real life – we find effective ways to do things with needing to failing a lot at every other alternative

Stochastic Multi-Armed Bandit (MAB)

Without knowing $\{\mu_i, D_i\}_{i=1}^k$, design a strategy/policy that chooses an arm sequence $i^1, i^2, \cdots, i^T$ to maximize $\mathbb{E}[\sum_{t=1}^T R_{i^t}]$

Measuring Learning Performance

Stochastic Multi-Armed Bandit (MAB)

Without knowing $\{\mu_i, D_i\}_{i=1}^{k}$, design a strategy/policy that chooses an arm sequence $i^1, i^2, \cdots, i^T$ to maximize $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$

Q: how to measure performance, or how well an algorithm did?

- A natural first thought would be to calculate achieve rewards $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$
- In online learning, it is more conventional to measure its slight variant

$$\text{Regret} = \underbrace{T\mu^*}_{} - \mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$$

Best possible award in hindsight (i.e., with perfect knowledge so no need to learn)

Measuring Learning Performance

Stochastic Multi-Armed Bandit (MAB)

Without knowing $\{\mu_i, D_i\}_{i=1}^{k}$, design a strategy/policy that chooses an arm sequence $i^1, i^2, \cdots, i^T$ to maximize $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$

Q: how to measure performance, or how well an algorithm did?

- A natural first thought would be to calculate achieve rewards $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$
- In online learning, it is more conventional to measure its slight variant

$$\text{Regret} = T\mu^* - \mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$$

- Goal is to **minimize regret**
  - equivalent to maximize $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$, but analytically more convenient

A Little History of MAB

Stochastic Multi-Armed Bandit (MAB)

Without knowing $\{\mu_i, D_i\}_{i=1}^{k}$, design a strategy/policy that chooses an arm sequence $i^1, i^2, \cdots, i^T$ to maximize $\mathbb{E}[\sum_{t=1}^{T} R_{i^t}]$

- This is a very clean and elegant problem
- Despite "bandit" in its name, MAB was initially motivated by designing reward-maximizing clinic trials, where an arm = a medical treatment
    - Started by William R. Thompson in 1930s whose designed the first algorithm for MAB, now called "Thompson Sampling"
- Extensively studied in the past two decades, due to being the cornerstone of reinforcement learning
    - Many design principles for MAB naturally generalize to RL
- Has really a lot of applications, even in many of today's real systems

Next: Concentration Inequalities

Very useful technical lemmas for later lectures

Balancing Reward and Risk is Crucial in Decisions

- In many real decision-making problems, we only receive random rewards, but optimal decisions depends on underlying expected reward
  - MAB is such an example; so is buying stocks
- Samples' average (also called empirical mean) is a good proxy of true mean, but not always accurate → there is risk (i.e., chance that true mean is actually very different from empirical mean)
- Intuitively, the more samples, typically the closer empirical mean is to true mean (thus less risk)

Balancing Reward and Risk is Crucial in Decisions

- In many real decision-making problems, we only receive random rewards, but optimal decisions depends on underlying expected reward
    - MAB is such an example; so is buying stocks
- Samples' average (also called empirical mean) is a good proxy of true mean, but not always accurate → there is risk (i.e., chance that true mean is actually very different from empirical mean)
- Intuitively, the more samples, typically the closer empirical mean is to true mean (thus less risk)

We want a rigorous quantitative statement for the above intuition!

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \le \sqrt{\frac{\log 1/\delta}{n}}\right) \ge 1 - 2\delta$$

- Intuitively, the more samples, typically the closer empirical mean is to true mean (thus less risk)

We want a rigorous quantitative statement for the above intuition!

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left| \frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n} \right| \leq \sqrt{\frac{\log 1/\delta}{n}} \right) \geq 1 - 2\delta$$

- Intuitively, the more samples, typically the closer empirical mean is to true mean (thus less risk)

We want a rigorous quantitative statement for the above intuition!

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \le \sqrt{\frac{\log 1/\delta}{n}}\right) \ge 1 - 2\delta$$

- Intuitively, the more samples, typically the closer empirical mean is to true mean (thus less risk)

We want a rigorous quantitative statement for the above intuition!

Remark: the dependence on t, n are tight order-wise!

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left| \frac{\sum_{i=1}^n r_i}{n} - \frac{\sum_{i=1}^n \mu_i}{n} \right| \leq \sqrt{\frac{\log 1/\delta}{n}} \right) \geq 1 - 2\delta$$

Three important insights from the theorem

1. **The role of n (#samples)**: gap between empirical mean and true mean decays at $1/\sqrt{n}$ speed

Equivalently: sum of n independent random samples will be off from sum of their means roughly by $\sqrt{n}$ (ignoring effects of $t, \log t$)

$$\left| \frac{\sum_{i=1}^n r_i}{n} - \frac{\sum_{i=1}^n \mu_i}{n} \right| \leq \sqrt{\frac{\log t}{n}} \Leftrightarrow \left| \sum_{i=1}^n r_i - \sum_{i=1}^n \mu_i \right| \leq \sqrt{n \cdot \log t}$$

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left| \frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n} \right| \leq \sqrt{\frac{\log 1/\delta}{n}} \right) \geq 1 - 2\delta$$

Three important insights from the theorem

1. **The role of n (#samples)**: gap between empirical mean and true mean decays at $1/\sqrt{n}$ speed
   - ➢ Why should you should be amazed by this conclusion?
     - Intuitively, if each sample is off from mean by a small constant ϵ_i, then naively we expect $\sum_{i=1}^{n} \epsilon_i \approx \epsilon n$
     - This much sharper $\sqrt{n}$ bound is because summing up independent randomness hedges out uncertainties/risk, exactly at rate $\Theta(\sqrt{n})$
     - Mathematical reason: central limit theorem

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \leq \sqrt{\frac{\log 1/\delta}{n}}\right) \geq 1 - 2\delta$$

Three important insights from the theorem

2. Risk probability δ: gap between empirical mean and true mean amplifies at $\sqrt{\log(1/\delta)}$ speed as risk decreases

- Hence reducing probability of “bad” event has low cost
  - For example, reducing from $\delta = t^{-1}$ to $\delta = t^{-2}$, the $\log 1/\delta$ term changes from $\sqrt{\log t}$ to $\sqrt{2\log t}$
  - We will heavily rely on this property in algorithm design since it makes high probability guarantees “low cost”

Balancing Reward and Risk is Crucial in Decisions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \leq \sqrt{\frac{\log 1/\delta}{n}}\right) \geq 1 - 2\delta$$

Three important insights from the theorem

3. r_i's do not need to be from the same distribution – only independence and boundedness are needed

Corollary: for the special case when D_i's are the same, with mean μ, we say r_i's are *independent and identically distributed* (I.I.D.), and we have

$$\Pr\left(|\bar{\mu} - \mu| \leq \sqrt{\frac{\log 1/\delta}{n}}\right) \geq 1 - 2\delta$$

Balancing Reward and Risk is Crucial in Decisions

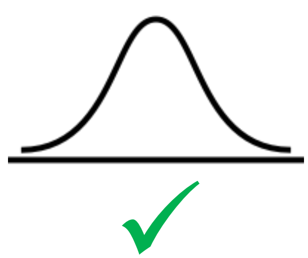
Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

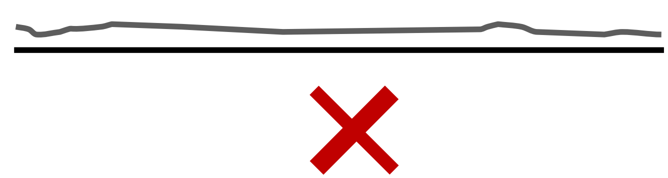
$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \leq \sqrt{\frac{\log 1/\delta}{n}}\right) \geq 1 - 2\delta$$

Three important insights from the theorem

3. r_i's do not need to be from the same distribution – only independence and boundedness are needed

- Boundedness can be easily generalized -- what is intrinsic is that distributions cannot be too spread out (i.e., having "heavy tails")





Generalized Versions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \le \sqrt{\frac{\log 1/\delta}{n}}\right) \ge 1 - 2\delta$$

Theorem (Generalization 1): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[a_i, b_i]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \le \frac{\sqrt{\log 1/\delta \times \sum_{i=1}^{n} (b_i - a_i)^2}}{n}\right) \ge 1 - 2\delta$$

Generalized Versions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^n r_i}{n} - \frac{\sum_{i=1}^n \mu_i}{n}\right| \leq \sqrt{\frac{\log 1/\delta}{n}}\right) \geq 1 - 2\delta$$

Theorem (Generalization 2): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from σ_i-sub-Gaussian distribution D_i with mean μ_i. Then

$$\Pr\left(\left|\frac{\sum_{i=1}^n r_i}{n} - \frac{\sum_{i=1}^n \mu_i}{n}\right| \leq \frac{\sqrt{\log 1/\delta \times \sum_{i=1}^n (\sigma_i)^2}}{n}\right) \geq 1 - 2\delta$$

- Intuitively, distribution X is σ-sub-Gaussian if its "spreadness" is upper bounded by a variance-σ Gaussian, up to a constant; formally

$$\Pr(|X - \mu_X| \geq t) \leq c \exp(t^2/\sigma^2), \forall t$$

Generalized Versions

Theorem (Hoeffding's inequality): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from a bounded distribution D_i supported on $[0, 1]$, with mean μ_i. Then we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n}\right| \leq \sqrt{\frac{\log 1/\delta}{n}}\right) \geq 1 - 2\delta$$

One-sided version

Theorem (Generalization 3): For $i = 1, \cdots, n$, let r_i be a sample drawn independently from σ_i-sub-Gaussian distribution D_i, with mean μ_i. Then

$$\Pr\left(\frac{\sum_{i=1}^{n} r_i}{n} - \frac{\sum_{i=1}^{n} \mu_i}{n} \geq \frac{\sqrt{\log 1/\delta \times \sum_{i=1}^{n} (\sigma_i)^2}}{n}\right) \leq \delta$$

- Symmetric side also holds
- Together imply the original version

Generalizing from Independence to Martingale

- It turns out that independence among r_i can also be (slightly) relaxed
- A famous/useful generalization is for Martingale

Definition: A sequence of random variables $X_1, X_2, \cdots$ is called a *martingale difference sequence* with respect to another sequence $R_1, R_2 \cdots$ if for any t, random var X_{t+1} is a function of $R_1, \cdots, R_t$, and

$$\mathbb{E}(X_{t+1} | R_1, \cdots, R_t) = 1 \qquad \text{with probability 1.}$$

Theorem (Azuma-Hoeffding inequality): Let $X_1, X_2, \cdots$ be a martingale difference sequence w.r.t. $R_1, R_2 \cdots$. Moreover, for any realized $r_1, \cdots, r_t$ sequence, $X_{t+1}(r_1, \cdots, r_t)$ satisfies (i.e., is σ-subgaussian)

$$\Pr(|X_{t+1}| \geq t) \leq c \exp(t^2/\sigma^2), \forall t$$

Then, we have

$$\Pr\left(\left| \frac{\sum_{i=1}^{n} x_i}{n} \right| \leq \sigma \sqrt{\frac{28c \log 1/\delta}{n}} \right) \leq 1 - 2\delta$$

Generalizing from Independence to Martingale

- Intuitively, even when X_{t+1} depends on the past randomness from $R_1, \cdots, R_{t-1}$, its sum still concentrates so long as its mean is the same under any realized $r_1, \cdots, r_{t-1}$ (and it is subgaussian)
- Unsurprisingly, there is one-sided version as well.
  - For interested audience, refer to a 2-page note "A Variant of Azuma's Inequality for Martingales with Subgaussian Tails" for one-sided version
  - Citing author's note, "the numerical constant can be improved", though not important for the purpose of this course

Theorem (Azuma-Hoeffding inequality): Let $X_1, X_2, \cdots$ be a martingale difference sequence w.r.t. $R_1, R_2 \cdots$. Moreover, for any realized $r_1, \cdots, r_t$ sequence, $X_{t+1}(r_1, \cdots, r_t)$ satisfies (i.e., is σ-subgaussian)

$$\Pr(|X_{t+1}| \geq t) \leq c \exp(t^2/\sigma^2), \forall t$$

Then, we have

$$\Pr\left(\left|\frac{\sum_{i=1}^{n} x_i}{n}\right| \leq \sigma \sqrt{\frac{28c \log 1/\delta}{n}}\right) \leq 1 - 2\delta$$

Remarks

- Previous four versions are most common, but there are many other variants as well
  - If variance/spreadness is nicely small, you can get possibly even sharper bound (e.g., Berstein's inequality)
  - If spreadness (defined in subtle ways) cannot be upper bounded by a Gaussian, you can get weaker upper bounds
- Main takeaways
  - Independent randomness hedges out after being summed up together
  - This generally holds true with roughly $\Theta(\sqrt{n})$ rate, and can be proved under various conditions

Thank You

Haifeng Xu

University of Chicago

haifengxu@uchicago.edu