AI For Games: What Are We Talking About?

Jordan Thayer

Syllabus

1. Introduction To Games
   - Syllabus
   - Types of Games
   - Terminology
   - Brief History of Games and AI
2. Minimax Tree Search
3. α-β pruning
4. Multi-Armed Bandits and Monte Carlo Tree Search
5. Implementing Monte Carlo Tree Search
6. Weak and Strong Solutions to Games, Checkers

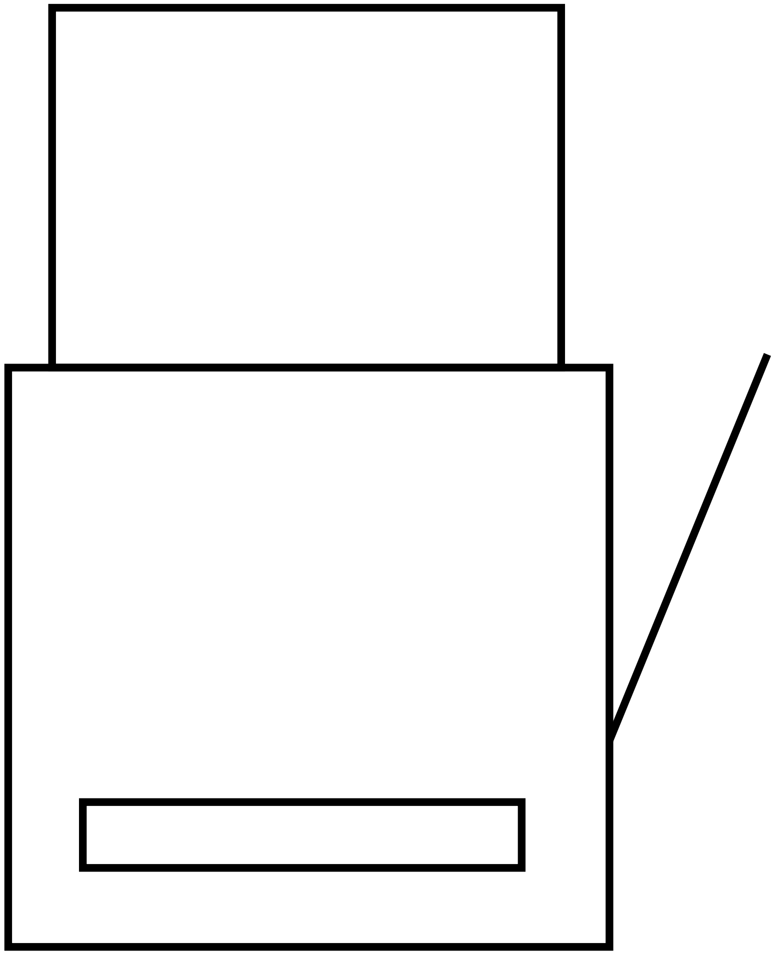
Today Multi-Armed Bandits and Monte Carlo Tree Search

- One Armed Bandits
- Distribution Estimation
- More than One Arm
- Formalizing Regret
- Transformation to Trees

Distribution Estimation

Bandit Games

Distribution Estimation



Distribution Estimation

\$0

\$1

\$2

\$3 |

\$4

\$5

Distribution Estimation

$0

$1 ||

$2 |

$3 ||||

$4 ||

$5

$0

$1 | | | |

$2 | | | | |

$3 | | | | | |

$4 | | |

$5 | |

Distribution Estimation

$0

$1 | | | |

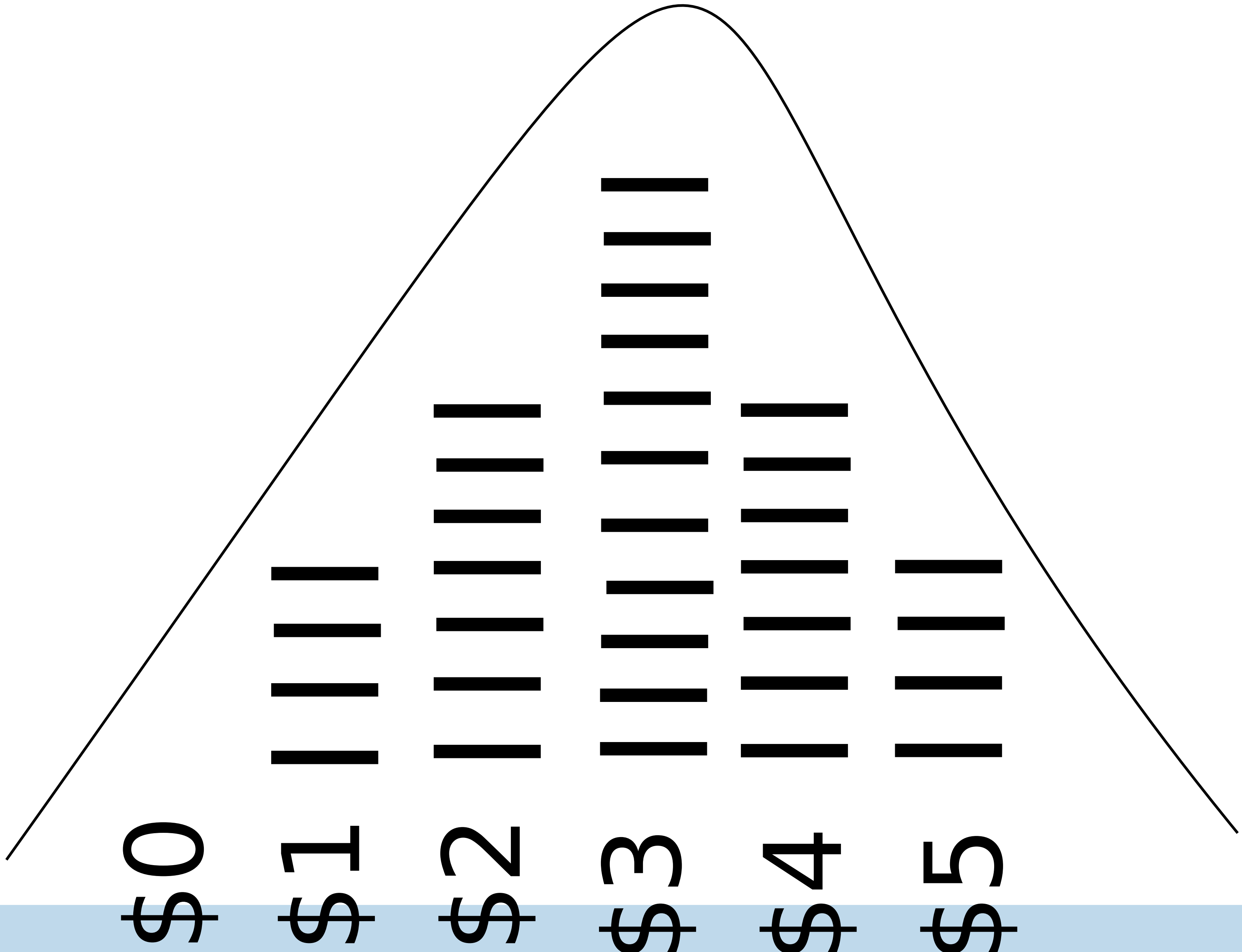
$2 | | | | | | |

$3 | | | | | | | | | | |

$4 | | | | | | |

$5 | | | |

Distribution Estimation



Formally

Let's say we have a model of our distribution:

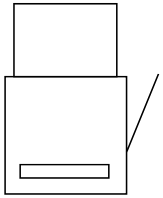
$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Formally

Let's say we have a model of our distribution:

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Further, we have a source of data:

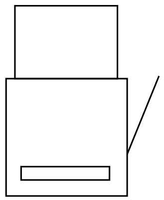


Formally

Let's say we have a model of our distribution:

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Further, we have a source of data:



Now, we just need to divine μ and σ by sampling from the data.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} payout_i$$
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(payout_i - \mu)^2$$

Multiple Arms, Regret

Going from 1 to n

Before, we had a single machine.

Going from 1 to n

But that isn't what's normal.

So, what's the real question?

The Naive Approach

1. Fix a trial budget
2. Sample evenly from the machines
3. Later pick the machine with the best payout with the rest of your cash

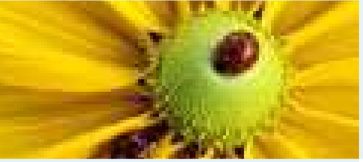
The Informal Model

Given a budget, determine which arm is best to pull and pull it as often as possible.

The Informal Model

Given a budget, determine which arm is best to pull and pull it as often as possible.

Pull levers that aren't the best payout as infrequently as possible



The Informal Model

Given a budget, determine which arm is best to pull and pull it as often as possible.

Pull levers that aren't the best payout as infrequently as possible

But still be really sure that your notion of 'best lever' is correct.

The Formal Model

We have a set of bandits, modled by a set of distributions $B = R_1, ..., R_K$.

The Formal Model

We have a set of bandits, modled by a set of distributions $B = R_1, ..., R_K$.

Each distribution R_i represents the ith bandit, and has mean payout μ_i.

The Formal Model

We have a set of bandits, modled by a set of distributions $B = R_1, ..., R_K$.

Each distribution R_i represents the ith bandit, and has mean payout μ_i.

We have a horizon H which represents the number of pulls available to us.

The Formal Model

We have a set of bandits, modled by a set of distributions $B = R_1, ..., R_K$.

Each distribution R_i represents the ith bandit, and has mean payout μ_i.

We have a horizon H which represents the number of pulls available to us.

The regret after T rounds is $\rho = T\mu^* - \sum_{t=1}^{T} r_t$ where:
μ^* is the mean payout of the best bandit, and r_t is the reward earned from pull t.

Applied to Games

But I Don’t Want To Play Slots