



Bandits

CISC 7404 - Decision Making

Steven Morad

University of Macau

Sets	2
Functions	11
Exercises	16
Bandits	20
Multiarmed Bandits	39
Questions?	52
Coding	53

Sets

Sets

Let us review some notation I will use in the course

Sets

Let us review some notation I will use in the course

If you ever get confused, come back to these slides

Sets

Let us review some notation I will use in the course

If you ever get confused, come back to these slides

Vectors

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Sets

Let us review some notation I will use in the course

If you ever get confused, come back to these slides

Vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrices

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix}$$

Sets

We will represent **tensors** as nested vectors or matrices

Sets

We will represent **tensors** as nested vectors or matrices

Tensor

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$$

Sets

We will represent **tensors** as nested vectors or matrices

Tensor

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$$

Each \mathbf{x}_i is a vector

Sets

Same for matrices

Tensor of matrices

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,n} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{m,1} & \mathbf{x}_{m,2} & \cdots & \mathbf{x}_{m,n} \end{bmatrix}$$

Sets

Question: What is the difference between the following?

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,n} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{m,1} & \mathbf{x}_{m,2} & \cdots & \mathbf{x}_{m,n} \end{bmatrix}$$

Sets

Capital letters will often refer to **sets**

Sets

Capital letters will often refer to **sets**

$$X = \{1, 2, 3, 4\}$$

Sets

Capital letters will often refer to **sets**

$$X = \{1, 2, 3, 4\}$$

We will represent important sets with blackboard font

Sets

Capital letters will often refer to **sets**

$$X = \{1, 2, 3, 4\}$$

We will represent important sets with blackboard font

\mathbb{R}

Set of all real numbers

$\{1, 2.03, \pi, \dots\}$

Sets

Capital letters will often refer to **sets**

$$X = \{1, 2, 3, 4\}$$

We will represent important sets with blackboard font

\mathbb{R}

Set of all real numbers

$$\{1, 2.03, \pi, \dots\}$$

\mathbb{Z}

Set of all integers

$$\{-2, -1, 0, 1, 2, \dots\}$$

Sets

Capital letters will often refer to **sets**

$$X = \{1, 2, 3, 4\}$$

We will represent important sets with blackboard font

\mathbb{R}

Set of all real numbers

$$\{1, 2.03, \pi, \dots\}$$

\mathbb{Z}

Set of all integers

$$\{-2, -1, 0, 1, 2, \dots\}$$

\mathbb{Z}_+

Set of all **positive** integers

$$\{1, 2, \dots\}$$

Sets

$[0, 1]$

Closed interval

0.0, 0.01, 0.00...1, 0.99, 1.0

Sets

$[0, 1]$

Closed interval

0.0, 0.01, 0.00...1, 0.99, 1.0

$(0, 1)$

Open interval 0.01, 0.00...1, 0.99

Sets

$[0, 1]$

Closed interval

0.0, 0.01, 0.00...1, 0.99, 1.0

$(0, 1)$

Open interval 0.01, 0.00...1, 0.99

$\{0, 1\}$

Set of two numbers (boolean)

Sets

$[0, 1]$

Closed interval

0.0, 0.01, 0.00...1, 0.99, 1.0

$(0, 1)$

Open interval 0.01, 0.00...1, 0.99

$\{0, 1\}$

Set of two numbers (boolean)

$[0, 1]^k$

A vector of k numbers between 0 and 1

Sets

$$[0, 1]$$

Closed interval

0.0, 0.01, 0.00...1, 0.99, 1.0

$$(0, 1)$$

Open interval 0.01, 0.00...1, 0.99

$$\{0, 1\}$$

Set of two numbers (boolean)

$$[0, 1]^k$$

A vector of k numbers between 0 and 1

$$\{0, 1\}^{k \times k}$$

A matrix of boolean values of shape k by k

Sets

We will use various set operations

Sets

We will use various set operations

$$A \subseteq B$$

A is a subset of B

Sets

We will use various set operations

$$A \subseteq B$$

A is a subset of B

$$A \subset B$$

A is a strict subset of B

Sets

We will use various set operations

$$A \subseteq B$$

A is a subset of B

$$A \subset B$$

A is a strict subset of B

$$a \in A$$

a is an element of A

Sets

We will use various set operations

$$A \subseteq B$$

A is a subset of B

$$A \subset B$$

A is a strict subset of B

$$a \in A$$

a is an element of A

$$b \notin A$$

b is not an element of A

Sets

We will use various set operations

$$A \subseteq B$$

A is a subset of B

$$A \subset B$$

A is a strict subset of B

$$a \in A$$

a is an element of A

$$b \notin A$$

b is not an element of A

$$A \cup B$$

The union of sets A and B

Sets

We will use various set operations

$$A \subseteq B$$

A is a subset of B

$$A \subset B$$

A is a strict subset of B

$$a \in A$$

a is an element of A

$$b \notin A$$

b is not an element of A

$$A \cup B$$

The union of sets A and B

$$A \cap B$$

The intersection of sets A and B

Sets

We will often use **set builder** notation

Sets

We will often use **set builder** notation

$$\{ x + 1 \mid x \in \mathbb{Z} \}$$

Sets

We will often use **set builder** notation

$$\{ \boxed{x + 1} \mid \boxed{x \in Z} \}$$

↑ ↖
Function Domain

Sets

We will often use **set builder** notation

$$\{ \boxed{x + 1} \mid \boxed{x \in Z} \}$$

↑ ↖
Function Domain

You can think of this as a for loop

```
output = {} # Set
for x in Z:
    output.insert(x + 1)
```

Sets

We will often use **set builder** notation

$$\{ \boxed{x + 1} \mid \boxed{x \in Z} \}$$

 ↑ ↑
Function Domain

You can think of this as a for loop

```
output = {} # Set
for x in Z:
    output.insert(x + 1)
```

```
output = {x + 1 for x in Z}
```

Functions

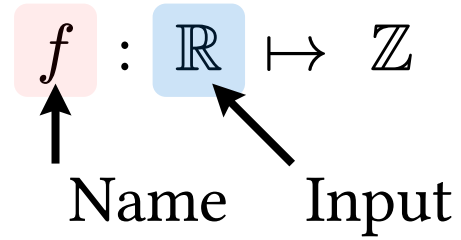
Functions

We define **functions** or **maps** between sets

$$f : \mathbb{R} \mapsto \mathbb{Z}$$

Functions

We define **functions** or **maps** between sets



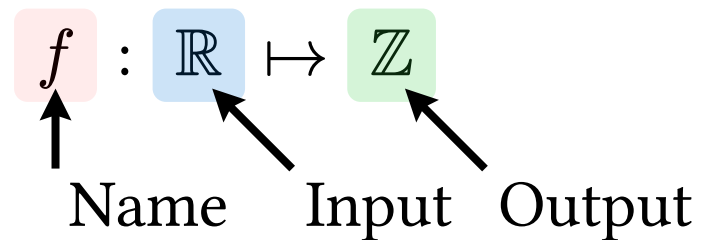
The diagram shows a function definition $f : \mathbb{R} \mapsto \mathbb{Z}$. The symbol f is enclosed in a light red square, and the symbol \mathbb{R} is enclosed in a light blue square. Below the red square, an upward-pointing arrow is labeled "Name". Below the blue square, a diagonal arrow pointing up and to the left is labeled "Input".

$$f : \mathbb{R} \mapsto \mathbb{Z}$$

Name Input

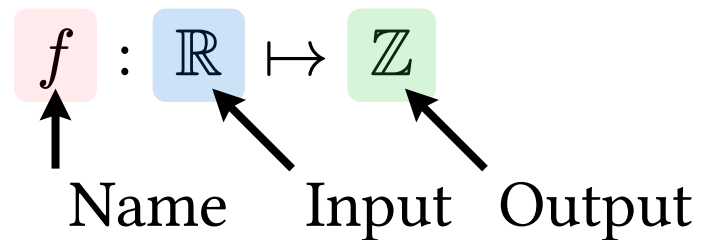
Functions

We define **functions** or **maps** between sets



Functions

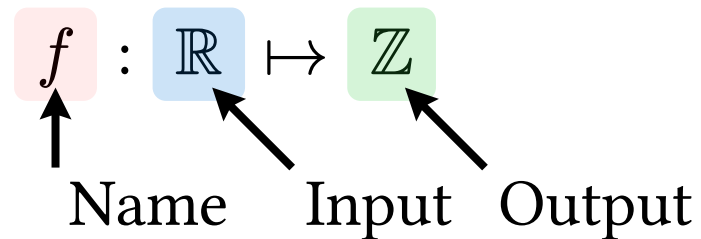
We define **functions** or **maps** between sets



A function f maps a real number to an integer

Functions

We define **functions** or **maps** between sets

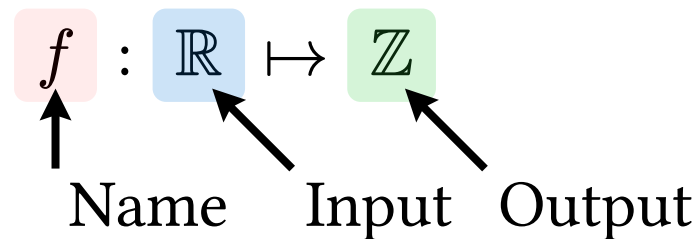


A function f maps a real number to an integer

Question: What functions could f be?

Functions

We define **functions** or **maps** between sets



A function f maps a real number to an integer

Question: What functions could f be?

$$\text{round} : \mathbb{R} \mapsto \mathbb{Z}$$

Functions

Functions can have multiple inputs

$$f : X \times \Theta \mapsto Y$$

Functions

Functions can have multiple inputs

$$f : X \times \Theta \mapsto Y$$

The function f maps elements from sets X and Θ to set Y

Functions

Functions can have multiple inputs

$$f : X \times \Theta \mapsto Y$$

The function f maps elements from sets X and Θ to set Y

I will define variables when possible

$$X \in \mathbb{R}^n; \Theta \in \mathbb{R}^{m \times n}; Y \in [0, 1]^{n \times m}$$

Functions

Functions can have multiple inputs

$$f : X \times \Theta \mapsto Y$$

The function f maps elements from sets X and Θ to set Y

I will define variables when possible

$$X \in \mathbb{R}^n; \Theta \in \mathbb{R}^{m \times n}; Y \in [0, 1]^{n \times m}$$

Functions

The max function returns the maximum of a function over a domain

Functions

The max function returns the maximum of a function over a domain

$$\max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

Functions

The max function returns the maximum of a function over a domain

$$\max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\max_{x \in Z} f(x)$$

Functions

The max function returns the maximum of a function over a domain

$$\max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\max_{x \in Z} f(x)$$

The arg max operator returns the input that maximizes a function

Functions

The max function returns the maximum of a function over a domain

$$\max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\max_{x \in Z} f(x)$$

The arg max operator returns the input that maximizes a function

$$\arg \max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Z$$

Functions

The max function returns the maximum of a function over a domain

$$\max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\max_{x \in Z} f(x)$$

The arg max operator returns the input that maximizes a function

$$\arg \max : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Z$$

$$\arg \max_{x \in Z} f(x)$$

Functions

We also have the min and arg min operators, which minimize f

$$\min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

Functions

We also have the min and arg min operators, which minimize f

$$\min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\min_{x \in Z} f(x)$$

Functions

We also have the min and arg min operators, which minimize f

$$\min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\min_{x \in Z} f(x)$$

$$\arg \min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Z$$

Functions

We also have the min and arg min operators, which minimize f

$$\min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\min_{x \in Z} f(x)$$

$$\arg \min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Z$$

$$\arg \min_{x \in Z} f(x)$$

Functions

We also have the \min and $\arg \min$ operators, which minimize f

$$\min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Y$$

$$\min_{x \in Z} f(x)$$

$$\arg \min : (f : X \mapsto Y) \times (Z \subseteq X) \mapsto Z$$

$$\arg \min_{x \in Z} f(x)$$

We want to make optimal decisions, so we will often take the minimum or maximum of functions

Exercises

Exercises

$$\mathbb{R}^n$$

Exercises

\mathbb{R}^n

Set of all vectors containing n real numbers

Exercises

$$\mathbb{R}^n$$

Set of all vectors containing n real numbers

$$\{3, 4, \dots, 31\}$$

Exercises

$$\mathbb{R}^n$$

Set of all vectors containing n real numbers

$$\{3, 4, \dots, 31\}$$

Set of all integers between 3 and 31

Exercises

$$\mathbb{R}^n$$

Set of all vectors containing n real numbers

$$\{3, 4, \dots, 31\}$$

Set of all integers between 3 and 31

$$[0, 1]^n$$

Exercises

$$\mathbb{R}^n$$

Set of all vectors containing n real numbers

$$\{3, 4, \dots, 31\}$$

Set of all integers between 3 and 31

$$[0, 1]^n$$

Set of all vectors of length n with values between 0 and 1

Exercises

$$\mathbb{R}^n$$

Set of all vectors containing n real numbers

$$\{3, 4, \dots, 31\}$$

Set of all integers between 3 and 31

$$[0, 1]^n$$

Set of all vectors of length n with values between 0 and 1

$$\{0, 1\}^n$$

Exercises

$$\mathbb{R}^n$$

Set of all vectors containing n real numbers

$$\{3, 4, \dots, 31\}$$

Set of all integers between 3 and 31

$$[0, 1]^n$$

Set of all vectors of length n with values between 0 and 1

$$\{0, 1\}^n$$

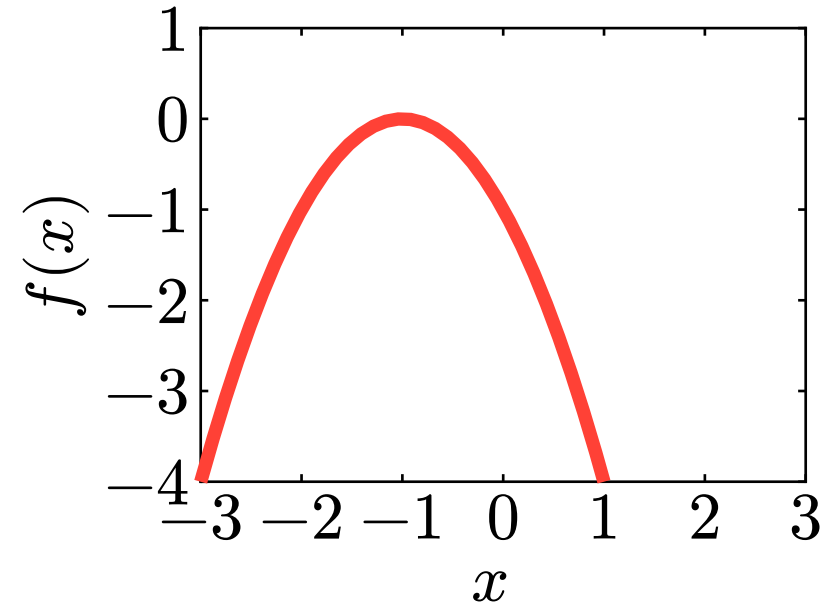
Set of all boolean vectors of length n

Exercises

$$f(x) = -(x + 1)^2$$

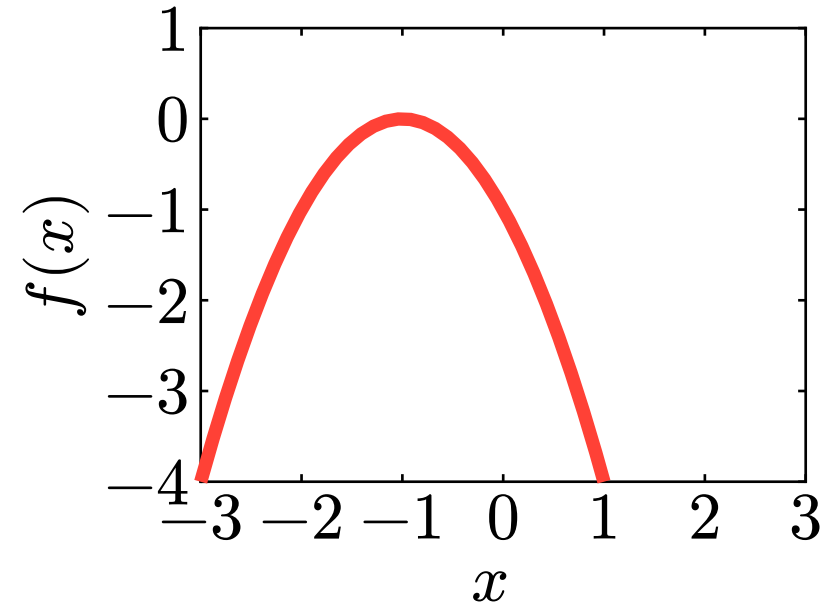
Exercises

$$f(x) = -(x + 1)^2$$



Exercises

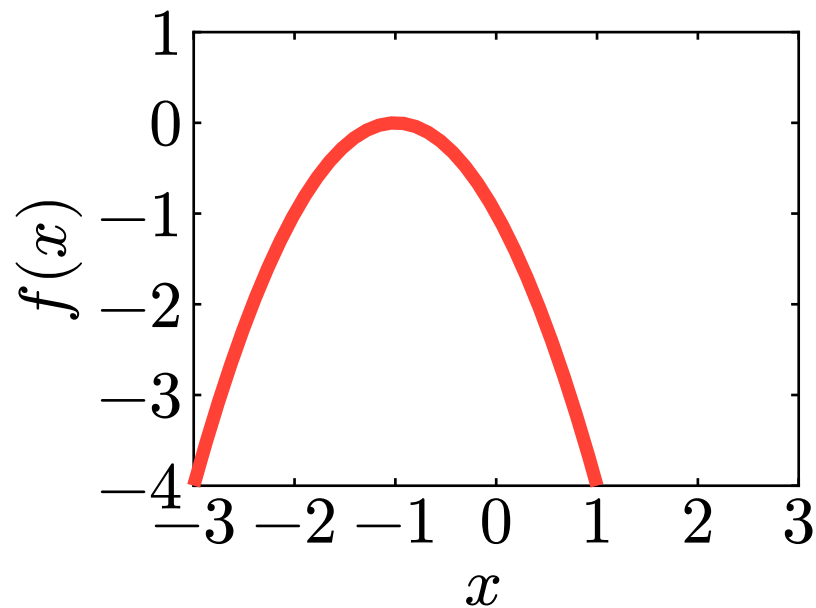
$$f(x) = -(x + 1)^2$$



$$\max_{x \in \mathbb{R}} f(x)?$$

Exercises

$$f(x) = -(x + 1)^2$$

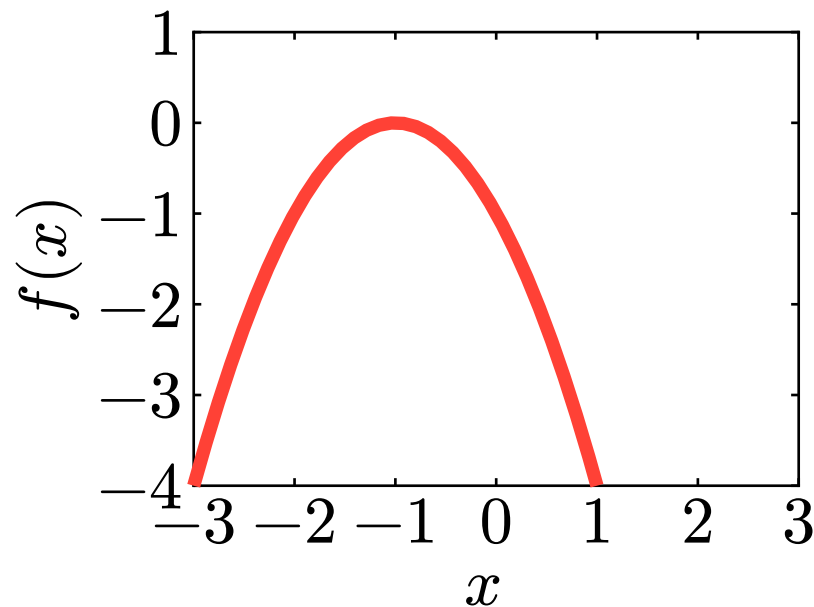


$$\max_{x \in \mathbb{R}} f(x)?$$

$$\arg \max_{x \in \mathbb{R}} f(x)?$$

Exercises

$$f(x) = -(x + 1)^2$$



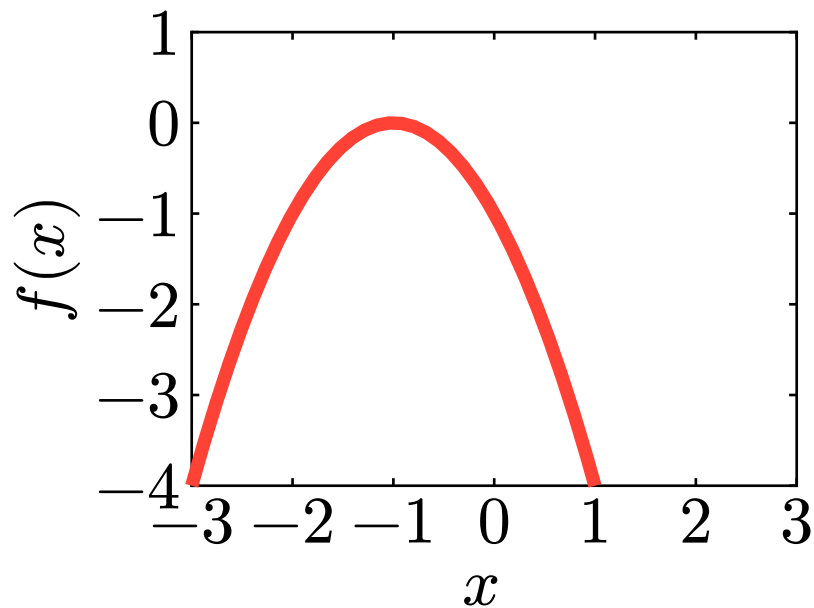
$$\max_{x \in \mathbb{R}} f(x)?$$

$$\arg \max_{x \in \mathbb{R}} f(x)?$$

$$\arg \max_{x \in \mathbb{Z}_+} f(x)?$$

Exercises

$$f(x) = -(x + 1)^2$$



$$\max_{x \in \mathbb{R}} f(x)?$$

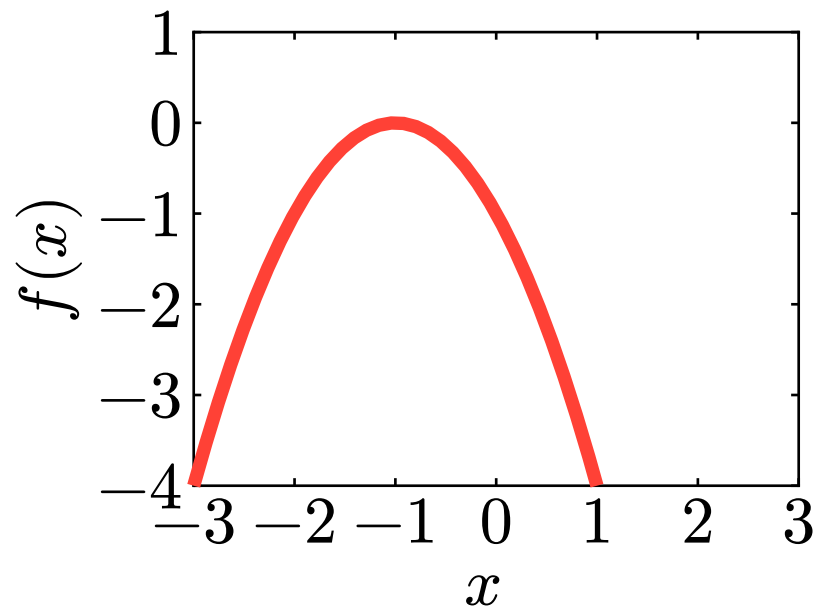
0

$$\arg \max_{x \in \mathbb{R}} f(x)?$$

$$\arg \max_{x \in \mathbb{Z}_+} f(x)?$$

Exercises

$$f(x) = -(x + 1)^2$$



$$\max_{x \in \mathbb{R}} f(x)?$$

0

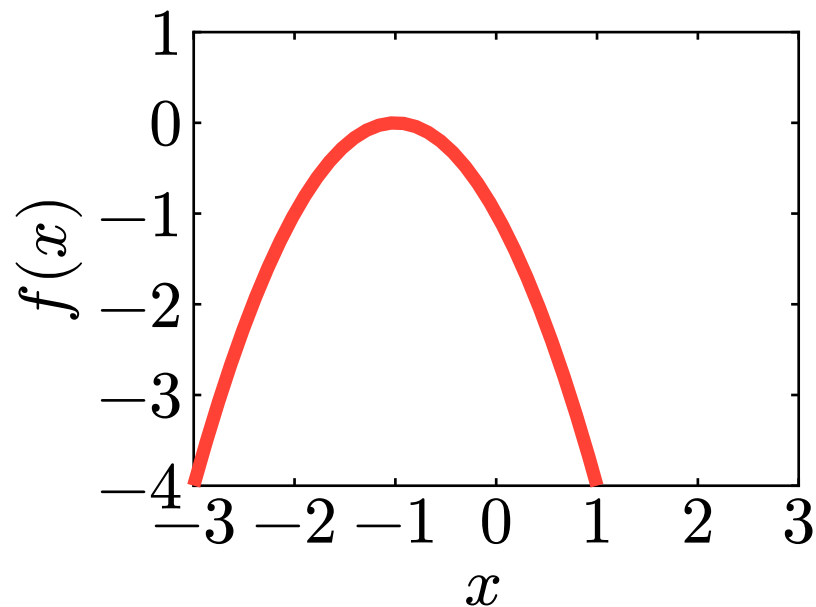
$$\arg \max_{x \in \mathbb{R}} f(x)?$$

-1

$$\arg \max_{x \in \mathbb{Z}_+} f(x)?$$

Exercises

$$f(x) = -(x + 1)^2$$



$$\max_{x \in \mathbb{R}} f(x)?$$

0

$$\arg \max_{x \in \mathbb{R}} f(x)?$$

-1

$$\arg \max_{x \in \mathbb{Z}_+} f(x)?$$

1

Exercises

$$\left\{ x^{\frac{1}{2}} \mid x \in \mathbb{R}_+ \right\}$$

Exercises

$$\left\{ x^{\frac{1}{2}} \mid x \in \mathbb{R}_+ \right\}$$

Question: What is this?

Exercises

$$\left\{ x^{\frac{1}{2}} \mid x \in \mathbb{R}_+ \right\}$$

Question: What is this?

Answer:

Exercises

$$\left\{ x^{\frac{1}{2}} \mid x \in \mathbb{R}_+ \right\}$$

Question: What is this?

Answer:

- An infinitely large set of all real numbers greater than zero

Exercises

$$\left\{ x^{\frac{1}{2}} \mid x \in \mathbb{R}_+ \right\}$$

Question: What is this?

Answer:

- An infinitely large set of all real numbers greater than zero
- The results of evaluating $f(x) = \sqrt{x}$ for all positive real numbers

Bandits

Bandits

The Sutton and Barto textbook reviews bandits before introducing reinforcement learning

Bandits

The Sutton and Barto textbook reviews bandits before introducing reinforcement learning

Bandits are a simplified version of reinforcement learning

Bandits

The Sutton and Barto textbook reviews bandits before introducing reinforcement learning

Bandits are a simplified version of reinforcement learning

It provides a “taste” of reinforcement learning in a single lecture

Bandits

The Sutton and Barto textbook reviews bandits before introducing reinforcement learning

Bandits are a simplified version of reinforcement learning

It provides a “taste” of reinforcement learning in a single lecture

Today's lecture will be difficult

Bandits

The Sutton and Barto textbook reviews bandits before introducing reinforcement learning

Bandits are a simplified version of reinforcement learning

It provides a “taste” of reinforcement learning in a single lecture

Today’s lecture will be difficult

But if you can understand it, then reinforcement learning will be easy for you

Bandits

Bandits are the simplest decision making problem

Bandits

Bandits are the simplest decision making problem

Question: What is a bandit?

Bandits

Bandits are the simplest decision making problem

Question: What is a bandit?



Bandits

Bandits are the simplest decision making problem

Question: What is a bandit?



A bandit steals your money

Bandits

Here is the bandit we will focus on in this course

Bandits

Here is the bandit we will focus on in this course



Bandits

Here is the bandit we will focus on in this course



This is a **one-armed** bandit

Bandits



Bandits

Question: How does a one-armed bandit steal your money?



Bandits



Question: How does a one-armed bandit steal your money?

Answer: You win less money than you put in

Bandits



Question: How does a one-armed bandit steal your money?

Answer: You win less money than you put in

Example: Costs 10 MOP to play,
you can win 1000 MOP each spin

Bandits



Question: How does a one-armed bandit steal your money?

Answer: You win less money than you put in

Example: Costs 10 MOP to play, you can win 1000 MOP each spin

Your chance of winning is $\frac{1}{200}$

Bandits



Question: How does a one-armed bandit steal your money?

Answer: You win less money than you put in

Example: Costs 10 MOP to play, you can win 1000 MOP each spin

Your chance of winning is $\frac{1}{200}$

Let us see if we can make money playing this game

Bandits

We will use **probability** to understand how much money we will make

Bandits

We will use **probability** to understand how much money we will make

First, we should briefly review probability theory

Bandits

We will use **probability** to understand how much money we will make

First, we should briefly review probability theory

The world is based on random **outcomes**

Bandits

We will use **probability** to understand how much money we will make

First, we should briefly review probability theory

The world is based on random **outcomes**

For our bandit, we have two possible outcomes

$$\Omega \in \{\text{win}, \text{lose}\}$$

Bandits

We will use **probability** to understand how much money we will make

First, we should briefly review probability theory

The world is based on random **outcomes**

For our bandit, we have two possible outcomes

$$\Omega \in \{\text{win}, \text{lose}\}$$

An **event** is a set of outcomes

$$E \subseteq \Omega$$

Bandits

We will use **probability** to understand how much money we will make

First, we should briefly review probability theory

The world is based on random **outcomes**

For our bandit, we have two possible outcomes

$$\Omega \in \{\text{win}, \text{lose}\}$$

An **event** is a set of outcomes

$$E \subseteq \Omega$$

$$E_{\text{win}} = \{\text{win}\}; \quad E_{\text{lose}} = \{\text{lose}\}; \quad E_{\text{any}} = \{\text{win}, \text{lose}\}$$

Bandits

We define the probabilities over the outcome and event spaces

Bandits

We define the probabilities over the outcome and event spaces

$$\Pr(\text{win}) = \frac{1}{200}, \quad \Pr(\text{lose}) = \frac{199}{200}$$

Bandits

We define the probabilities over the outcome and event spaces

$$\Pr(\text{win}) = \frac{1}{200}, \quad \Pr(\text{lose}) = \frac{199}{200}$$

Outcome probabilities **must be positive** and **must sum to one**

Bandits

We define the probabilities over the outcome and event spaces

$$\Pr(\text{win}) = \frac{1}{200}, \quad \Pr(\text{lose}) = \frac{199}{200}$$

Outcome probabilities **must be positive** and **must sum to one**

$$\sum_{\omega \in \Omega} \Pr(\omega) = 1$$

Bandits

We define the probabilities over the outcome and event spaces

$$\Pr(\text{win}) = \frac{1}{200}, \quad \Pr(\text{lose}) = \frac{199}{200}$$

Outcome probabilities **must be positive** and **must sum to one**

$$\sum_{\omega \in \Omega} \Pr(\omega) = 1$$

Event probabilities do not always sum to one

Bandits

We define the probabilities over the outcome and event spaces

$$\Pr(\text{win}) = \frac{1}{200}, \quad \Pr(\text{lose}) = \frac{199}{200}$$

Outcome probabilities **must be positive** and **must sum to one**

$$\sum_{\omega \in \Omega} \Pr(\omega) = 1$$

Event probabilities do not always sum to one

$$E_{\text{win}} = \{\text{win}\}$$

$$\sum_{\varepsilon \in E} \Pr(\varepsilon) \leq 1$$

Bandits

A **random variable** \mathcal{X} maps an outcome to a real number

Bandits

A **random variable** \mathcal{X} maps an outcome to a real number

$$\mathcal{X} : \Omega \mapsto \mathbb{R}$$

Bandits

A **random variable** \mathcal{X} maps an outcome to a real number

$$\mathcal{X} : \Omega \mapsto \mathbb{R}$$

Our bandit has two outcomes, lose (-10) or win (1000)

Bandits

A **random variable** \mathcal{X} maps an outcome to a real number

$$\mathcal{X} : \Omega \mapsto \mathbb{R}$$

Our bandit has two outcomes, lose (-10) or win (1000)

Question: What is the random variable for the bandit?

Bandits

A **random variable** \mathcal{X} maps an outcome to a real number

$$\mathcal{X} : \Omega \mapsto \mathbb{R}$$

Our bandit has two outcomes, lose (-10) or win (1000)

Question: What is the random variable for the bandit?

$$\mathcal{X} : \{\text{lose}, \text{win}\} \mapsto \{-10, 1000\}$$

Bandits

A **random variable** \mathcal{X} maps an outcome to a real number

$$\mathcal{X} : \Omega \mapsto \mathbb{R}$$

Our bandit has two outcomes, lose (-10) or win (1000)

Question: What is the random variable for the bandit?

$$\mathcal{X} : \{\text{lose}, \text{win}\} \mapsto \{-10, 1000\} \quad \mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

Bandits

We can also compute the probability over random variables

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real



Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

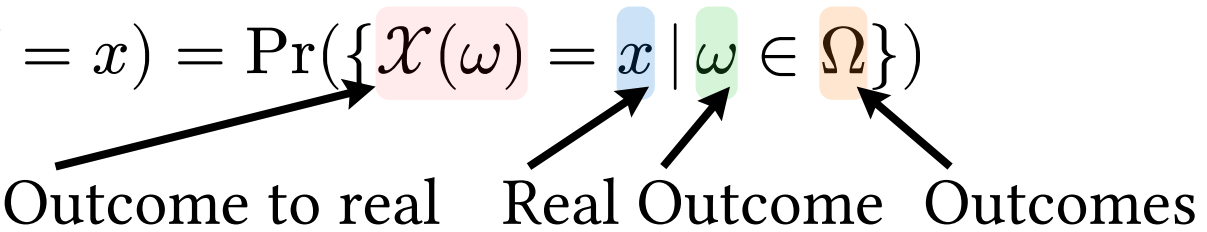
Outcome to real Real Outcome

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real Outcome Outcomes

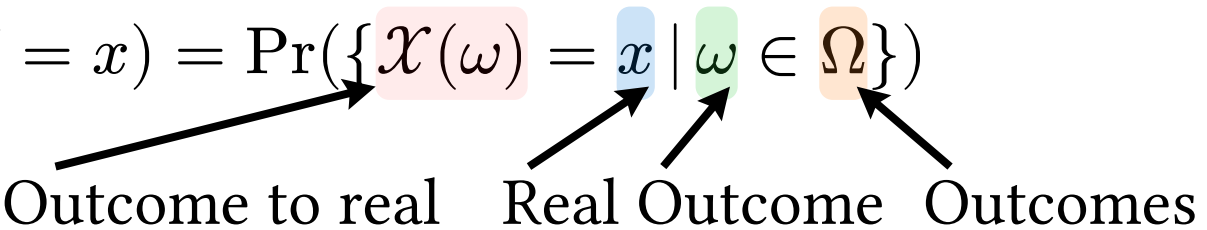


Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real Outcome Outcomes



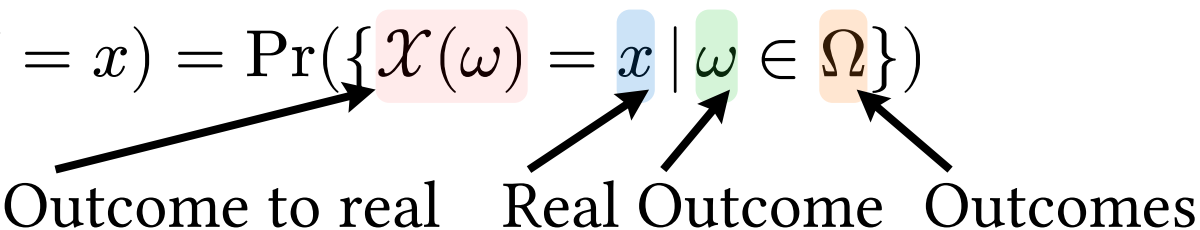
$$\mathcal{X} : \{\text{lose, win}\} \mapsto \{-10, 1000\}$$

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real Outcome Outcomes



$$\mathcal{X} : \{\text{lose}, \text{win}\} \mapsto \{-10, 1000\} \quad \mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\Pr(\mathcal{X}) = \begin{bmatrix} \Pr(\mathcal{X} = -10) \\ \Pr(\mathcal{X} = 1000) \end{bmatrix} =$$

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real Outcome Outcomes

$$\mathcal{X} : \{\text{lose}, \text{win}\} \mapsto \{-10, 1000\} \quad \mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

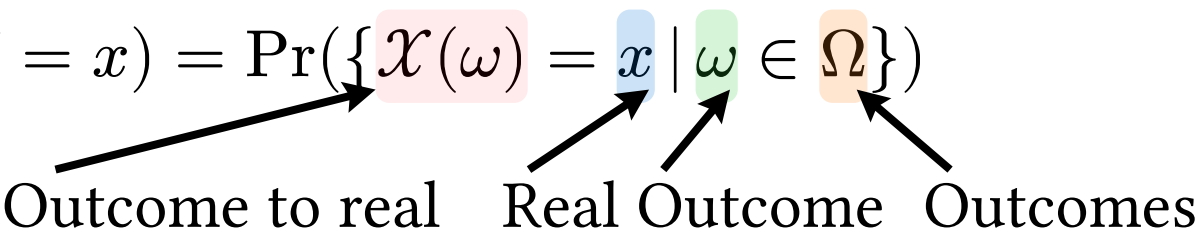
$$\Pr(\mathcal{X}) = \begin{bmatrix} \Pr(\mathcal{X} = -10) \\ \Pr(\mathcal{X} = 1000) \end{bmatrix} = \begin{bmatrix} \frac{199}{200} \\ \frac{1}{200} \end{bmatrix} = \begin{bmatrix} 0.995 \\ 0.005 \end{bmatrix}$$

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real Outcome Outcomes



$$\mathcal{X} : \{\text{lose}, \text{win}\} \mapsto \{-10, 1000\} \quad \mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\Pr(\mathcal{X}) = \begin{bmatrix} \Pr(\mathcal{X} = -10) \\ \Pr(\mathcal{X} = 1000) \end{bmatrix} = \begin{bmatrix} \frac{199}{200} \\ \frac{1}{200} \end{bmatrix} = \begin{bmatrix} 0.995 \\ 0.005 \end{bmatrix}$$

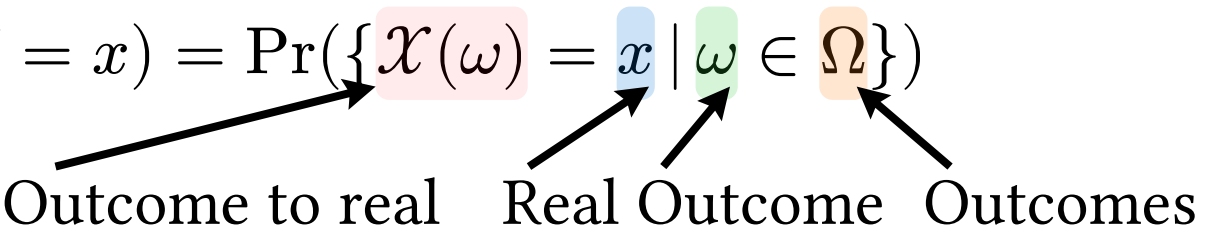
$$\Pr(\mathcal{X} = 1000) = \begin{bmatrix} \Pr(\mathcal{X} = -10) \\ \Pr(\mathcal{X} = 1000) \end{bmatrix} =$$

Bandits

We can also compute the probability over random variables

$$\Pr(\mathcal{X} = x) = \Pr(\{\mathcal{X}(\omega) = x \mid \omega \in \Omega\})$$

Outcome to real Real Outcome Outcomes



$$\mathcal{X} : \{\text{lose}, \text{win}\} \mapsto \{-10, 1000\} \quad \mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\Pr(\mathcal{X}) = \begin{bmatrix} \Pr(\mathcal{X} = -10) \\ \Pr(\mathcal{X} = 1000) \end{bmatrix} = \begin{bmatrix} \frac{199}{200} \\ \frac{1}{200} \end{bmatrix} = \begin{bmatrix} 0.995 \\ 0.005 \end{bmatrix}$$

$$\Pr(\mathcal{X} = 1000) = \begin{bmatrix} \Pr(\mathcal{X} = -10) \\ \Pr(\mathcal{X} = 1000) \end{bmatrix} = \begin{bmatrix} \frac{199}{200} \\ \frac{1}{200} \end{bmatrix} = \begin{bmatrix} 0.995 \\ 0.005 \end{bmatrix}$$

Bandits

Like before, the probability over the random variable **must sum to one**

Bandits

Like before, the probability over the random variable **must sum to one**

$$\sum_{\omega \in \Omega} \Pr(X(\omega)) = 1$$

Bandits

Like before, the probability over the random variable **must sum to one**

$$\sum_{\omega \in \Omega} \Pr(X(\omega)) = 1$$

$$\Pr(\mathcal{X}(\text{lose}) = -10) + \Pr(\mathcal{X}(\text{win}) = 1000) = 1$$

Bandits

Like before, the probability over the random variable **must sum to one**

$$\sum_{\omega \in \Omega} \Pr(X(\omega)) = 1$$

$$\Pr(\mathcal{X}(\text{lose}) = -10) + \Pr(\mathcal{X}(\text{win}) = 1000) = 1$$

$$\frac{199}{200} + \frac{1}{200} = 1$$

Bandits

We defined our bandit's probabilities

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

Bandits

We defined our bandit's probabilities

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

And the random variable

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

Bandits

We defined our bandit's probabilities

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

And the random variable

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

But we still do not know how much money we will make!

Bandits

We defined our bandit's probabilities

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

And the random variable

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

But we still do not know how much money we will make!

We can combine probabilities and random variables to find out

Bandits

The **expectation** or **expected value** \mathbb{E} is the mean of the random variable

Bandits

The **expectation** or **expected value** \mathbb{E} is the mean of the random variable

The expectation tells us how much money we make on average

Bandits

The **expectation** or **expected value** \mathbb{E} is the mean of the random variable

The expectation tells us how much money we make on average

$$\mathbb{E} : \underbrace{(\Omega \mapsto \mathbb{R})}_{\text{random variable}} \mapsto \mathbb{R}$$

Bandits

The **expectation** or **expected value** \mathbb{E} is the mean of the random variable

The expectation tells us how much money we make on average

$$\mathbb{E} : \underbrace{(\Omega \mapsto \mathbb{R})}_{\text{random variable}} \mapsto \mathbb{R}$$

$$\mathbb{E}[\mathcal{X}] = \sum_{\omega \in \Omega} \mathcal{X}(\omega) \cdot \Pr(\omega)$$

Bandits

The **expectation** or **expected value** \mathbb{E} is the mean of the random variable

The expectation tells us how much money we make on average

$$\mathbb{E} : \underbrace{(\Omega \mapsto \mathbb{R})}_{\text{random variable}} \mapsto \mathbb{R}$$

$$\mathbb{E}[\mathcal{X}] = \sum_{\omega \in \Omega} \mathcal{X}(\omega) \cdot \Pr(\omega)$$

Bandits

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

Bandits

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

Bandits

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\mathbb{E}[\mathcal{X}] = \sum_{\omega \in \Omega} \Pr(\omega) \cdot \mathcal{X}(\omega)$$

Bandits

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\mathbb{E}[\mathcal{X}] = \sum_{\omega \in \Omega} \Pr(\omega) \cdot \mathcal{X}(\omega)$$

Question: What is the expected value of the bandit?

Bandits

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\mathbb{E}[\mathcal{X}] = \sum_{\omega \in \Omega} \Pr(\omega) \cdot \mathcal{X}(\omega)$$

Question: What is the expected value of the bandit?

$$\Pr(\text{lose}) \cdot \mathcal{X}(\text{lose}) + \Pr(\text{win}) \cdot \mathcal{X}(\text{win})$$

Bandits

$$\Pr(\text{lose}) = \frac{199}{200}; \quad \Pr(\text{win}) = \frac{1}{200}$$

$$\mathcal{X}(\text{lose}) = -10; \quad \mathcal{X}(\text{win}) = 1000$$

$$\mathbb{E}[\mathcal{X}] = \sum_{\omega \in \Omega} \Pr(\omega) \cdot \mathcal{X}(\omega)$$

Question: What is the expected value of the bandit?

$$\Pr(\text{lose}) \cdot \mathcal{X}(\text{lose}) + \Pr(\text{win}) \cdot \mathcal{X}(\text{win})$$

$$\frac{199}{200} \cdot -10 + \frac{1}{200} \cdot 1000 = -4.95$$

Bandits

Question: What does $\mathbb{E}[\mathcal{X}] = -4.95$ mean?

Bandits

Question: What does $\mathbb{E}[\mathcal{X}] = -4.95$ mean?

Expect to lose 4.95 MOP on average each time you spin the bandit

Bandits

Question: What does $\mathbb{E}[\mathcal{X}] = -4.95$ mean?

Expect to lose 4.95 MOP on average each time you spin the bandit

We call the value after each spin the **reward**

Bandits

Question: What does $\mathbb{E}[\mathcal{X}] = -4.95$ mean?

Expect to lose 4.95 MOP on average each time you spin the bandit

We call the value after each spin the **reward**

$$r_1 = -10$$

$$r_2 = -10$$

$$\vdots$$

$$r_n = -10$$

Bandits

Question: What does $\mathbb{E}[\mathcal{X}] = -4.95$ mean?

Expect to lose 4.95 MOP on average each time you spin the bandit

We call the value after each spin the **reward**

$$r_1 = -10$$

$$r_2 = -10$$

$$\vdots$$

$$r_n = -10$$

Negative reward means we lose money

Bandits

$$r_1 = -10$$

$$r_2 = -10$$

$$\vdots$$

$$r_n = -10$$

Bandits

$$r_1 = -10$$

$$r_2 = -10$$

$$\vdots$$

$$r_n = -10$$

If play the game more, the mean reward converges to the expectation

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n r_t = n \cdot \mathbb{E}[\mathcal{X}] = -4.95n$$

Bandits

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n r_t = -4.95n = n\mathbb{E}[\mathcal{X}]$$

Bandits

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n r_t = -4.95n = n\mathbb{E}[\mathcal{X}]$$

If you play 1,000 times ($n = 1000$), expect to lose -4950 MOP

Bandits

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n r_t = -4.95n = n\mathbb{E}[\mathcal{X}]$$

If you play 1,000 times ($n = 1000$), expect to lose -4950 MOP

Question: What is the best way to make money with the bandit?

Bandits

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n r_t = -4.95n = n\mathbb{E}[\mathcal{X}]$$

If you play 1,000 times ($n = 1000$), expect to lose -4950 MOP

Question: What is the best way to make money with the bandit?

Answer: Do not play! If you must, play as little as possible

Bandits

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n r_t = -4.95n = n\mathbb{E}[\mathcal{X}]$$

If you play 1,000 times ($n = 1000$), expect to lose -4950 MOP

Question: What is the best way to make money with the bandit?

Answer: Do not play! If you must, play as little as possible

The more you play, the closer you get to $n \cdot \mathbb{E}[\mathcal{X}]$

Bandits

If you know $\mathbb{E}[\mathcal{X}]$, you know the result of gambling

Bandits

If you know $\mathbb{E}[\mathcal{X}]$, you know the result of gambling

Question: Do gamblers know $\mathbb{E}[\mathcal{X}]$?

Bandits

If you know $\mathbb{E}[\mathcal{X}]$, you know the result of gambling

Question: Do gamblers know $\mathbb{E}[\mathcal{X}]$?

Answer: No! This is a secret of the casino

Bandits

If you know $\mathbb{E}[\mathcal{X}]$, you know the result of gambling

Question: Do gamblers know $\mathbb{E}[\mathcal{X}]$?

Answer: No! This is a secret of the casino

Question: Could a gambler find out $\mathbb{E}[\mathcal{X}]$?

Bandits

Gambler only has access to the rewards

$$r_1, r_2, \dots, r_n = -10, -10, \dots, 1000$$

Bandits

Gambler only has access to the rewards

$$r_1, r_2, \dots, r_n = -10, -10, \dots, 1000$$

Question: How could a gambler find out $\mathbb{E}[\mathcal{X}]$?

Bandits

Gambler only has access to the rewards

$$r_1, r_2, \dots, r_n = -10, -10, \dots, 1000$$

Question: How could a gambler find out $\mathbb{E}[\mathcal{X}]$?

We can sum the rewards

$$\sum_{t=1}^n r_t \approx n \cdot \mathbb{E}[\mathcal{X}]$$

Bandits

Gambler only has access to the rewards

$$r_1, r_2, \dots, r_n = -10, -10, \dots, 1000$$

Question: How could a gambler find out $\mathbb{E}[\mathcal{X}]$?

We can sum the rewards

$$\sum_{t=1}^n r_t \approx n \cdot \mathbb{E}[\mathcal{X}]$$

Divide by number of plays

$$\frac{1}{n} \sum_{t=1}^n r_t \approx \mathbb{E}[\mathcal{X}]$$

Bandits

Gambler only has access to the rewards

$$r_1, r_2, \dots, r_n = -10, -10, \dots, 1000$$

Question: How could a gambler find out $\mathbb{E}[\mathcal{X}]$?

We can sum the rewards

$$\sum_{t=1}^n r_t \approx n \cdot \mathbb{E}[\mathcal{X}]$$

Divide by number of plays

$$\frac{1}{n} \sum_{t=1}^n r_t \approx \mathbb{E}[\mathcal{X}]$$

After playing enough, the gambler can approximate the expectation!

Bandits

Exercise: You start a new casino in Macau.

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\text{Pr}(\omega) \mid \omega \in \Omega\}$

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\Pr(\omega) \mid \omega \in \Omega\}$
- The random variable \mathcal{X} for each outcome $\{\mathcal{X}(\omega) \mid \omega \in \Omega\}$

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\Pr(\omega) \mid \omega \in \Omega\}$
- The random variable \mathcal{X} for each outcome $\{\mathcal{X}(\omega) \mid \omega \in \Omega\}$
- The expected value of the random variable $\mathbb{E}[\mathcal{X}]$

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\Pr(\omega) \mid \omega \in \Omega\}$
- The random variable \mathcal{X} for each outcome $\{\mathcal{X}(\omega) \mid \omega \in \Omega\}$
- The expected value of the random variable $\mathbb{E}[\mathcal{X}]$
- How much money we expect to make if the gambler plays 1000 times

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\Pr(\omega) \mid \omega \in \Omega\}$
- The random variable \mathcal{X} for each outcome $\{\mathcal{X}(\omega) \mid \omega \in \Omega\}$
- The expected value of the random variable $\mathbb{E}[\mathcal{X}]$
- How much money we expect to make if the gambler plays 1000 times

Make sure the expected value is **negative but near zero**:

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\Pr(\omega) \mid \omega \in \Omega\}$
- The random variable \mathcal{X} for each outcome $\{\mathcal{X}(\omega) \mid \omega \in \Omega\}$
- The expected value of the random variable $\mathbb{E}[\mathcal{X}]$
- How much money we expect to make if the gambler plays 1000 times

Make sure the expected value is **negative but near zero**:

- Negative: The gambler loses money and you make money

Bandits

Exercise: You start a new casino in Macau. Create a bandit with the following outcomes $\Omega \in \{\text{Win Lemon}, \text{Win Cherry}, \text{Win 7}, \text{Lose}\}$

Write down:

- Probability for each outcome $\{\Pr(\omega) \mid \omega \in \Omega\}$
- The random variable \mathcal{X} for each outcome $\{\mathcal{X}(\omega) \mid \omega \in \Omega\}$
- The expected value of the random variable $\mathbb{E}[\mathcal{X}]$
- How much money we expect to make if the gambler plays 1000 times

Make sure the expected value is **negative but near zero**:

- Negative: The gambler loses money and you make money
- Near zero: The gambler wins sometimes and will continue to play

Multiarmed Bandits

Multiarmed Bandits

The bandit problem is useful for casino owners and gamblers

Multiarmed Bandits

The bandit problem is useful for casino owners and gamblers

But it is a trivial decision making problem

Multiarmed Bandits

The bandit problem is useful for casino owners and gamblers

But it is a trivial decision making problem

If $\mathbb{E}[\mathcal{X}] > 0$ you should gamble

Multiarmed Bandits

The bandit problem is useful for casino owners and gamblers

But it is a trivial decision making problem

If $\mathbb{E}[\mathcal{X}] > 0$ you should gamble

If $\mathbb{E}[\mathcal{X}] < 0$ you should not gamble

Multiarmed Bandits

The bandit problem is useful for casino owners and gamblers

But it is a trivial decision making problem

If $\mathbb{E}[\mathcal{X}] > 0$ you should gamble

If $\mathbb{E}[\mathcal{X}] < 0$ you should not gamble

We will make the problem more interesting

Multiarmed Bandits

You arrive at the Londoner with 1000 MOP and want to win money

Multiarmed Bandits

You arrive at the Londoner with 1000 MOP and want to win money



Multiarmed Bandits

You arrive at the Londoner with 1000 MOP and want to win money



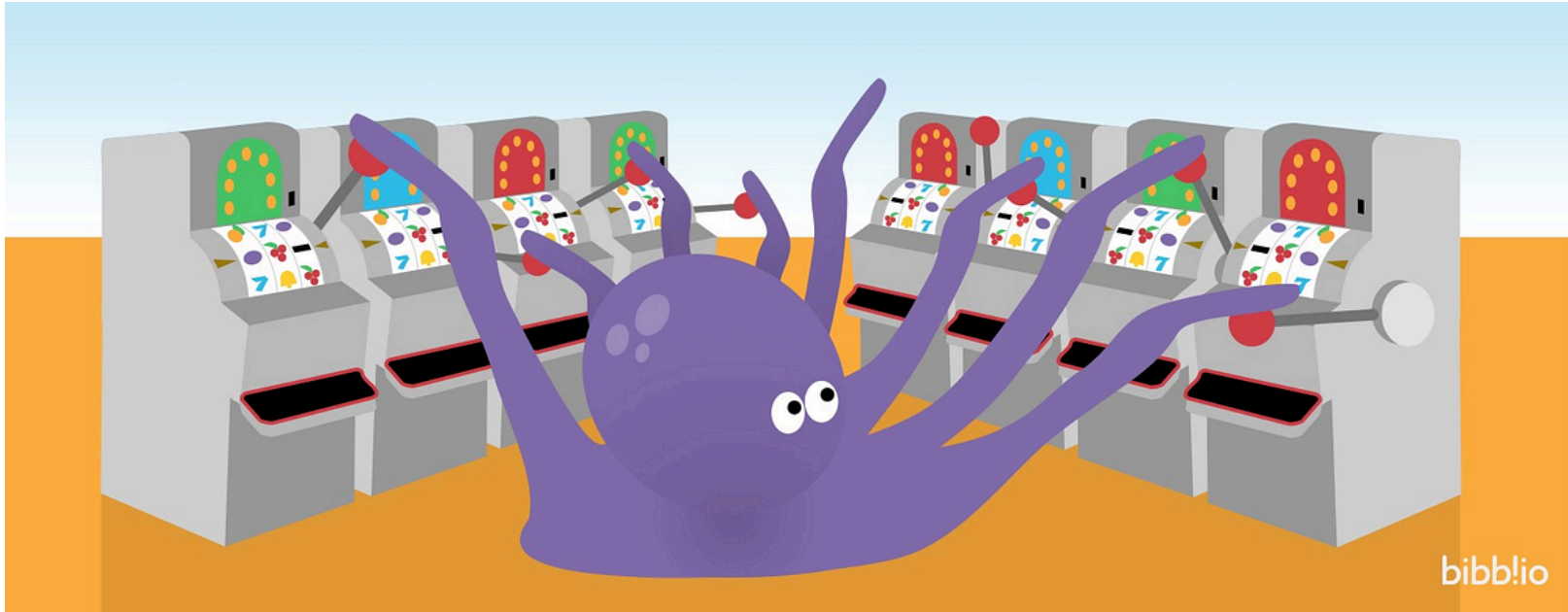
Question: Which machine do you play?

Multiarmed Bandits

We call this the **multi-armed bandit** problem

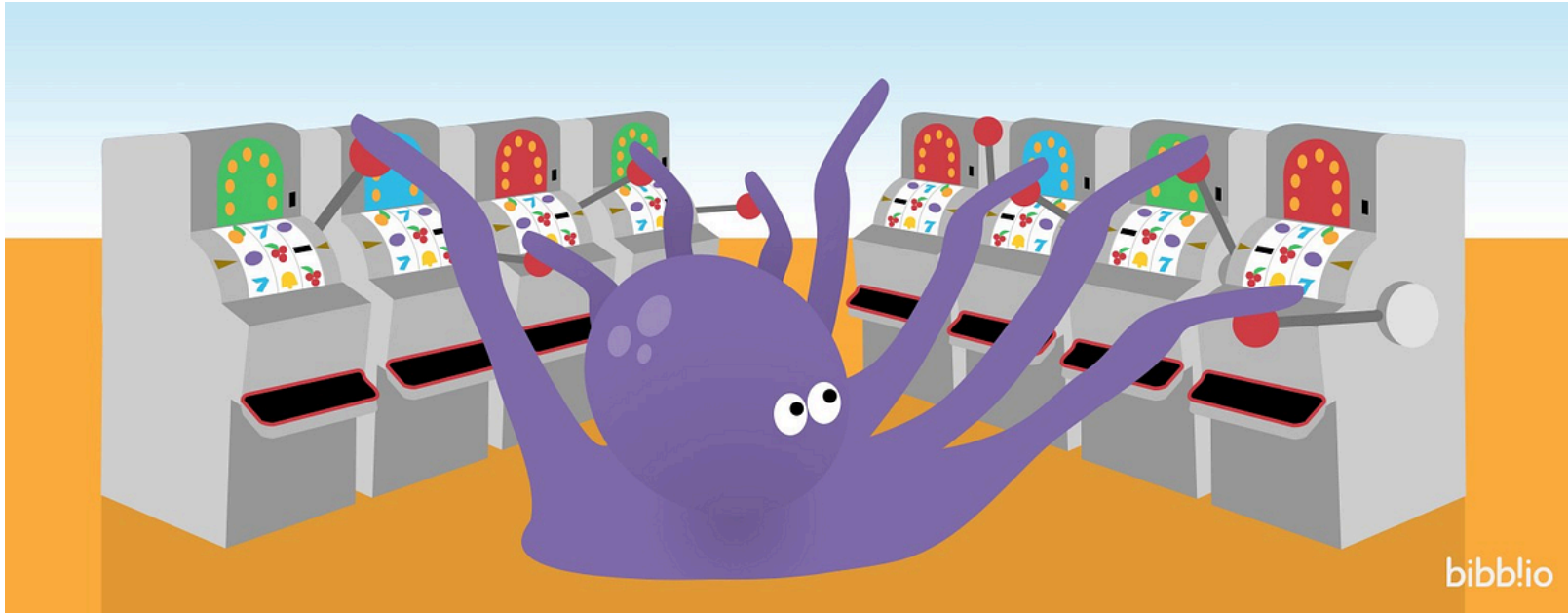
Multiarmed Bandits

We call this the **multi-armed bandit** problem



Multiarmed Bandits

We call this the **multi-armed bandit** problem



You don't know the expected value of each arm. Which should you pull?

Multiarmed Bandits

We can model many real problems as multiarmed bandits

Multiarmed Bandits

We can model many real problems as multiarmed bandits

For example, we can model hospital treatment as multiarmed bandits

Multiarmed Bandits

We can model many real problems as multiarmed bandits

For example, we can model hospital treatment as multiarmed bandits

We have new medicines, but do not know their effectiveness

Multiarmed Bandits

We can model many real problems as multiarmed bandits

For example, we can model hospital treatment as multiarmed bandits

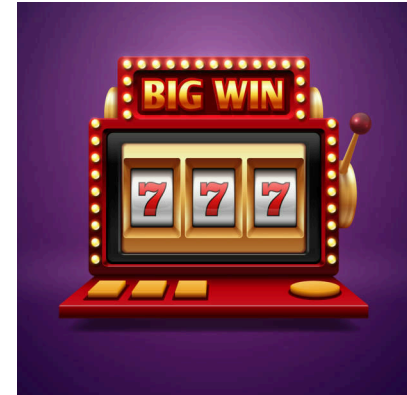
We have new medicines, but do not know their effectiveness



Medicine A



Medicine B



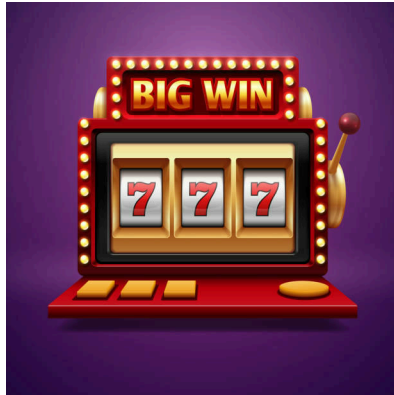
Medicine C

Multiarmed Bandits

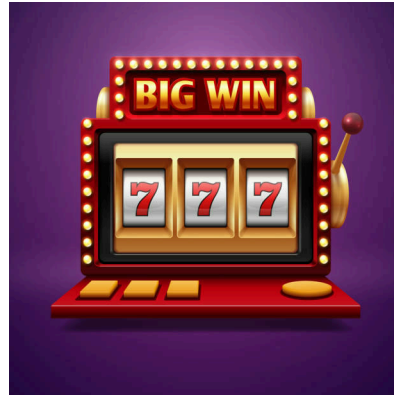
We can model many real problems as multiarmed bandits

For example, we can model hospital treatment as multiarmed bandits

We have new medicines, but do not know their effectiveness



Medicine A



Medicine B



Medicine C

We can find the best medicine while healing the most people

Multiarmed Bandits

YouTube, Youku, BiliBili, TikTok, Netflix use bandits to suggest videos

Multiarmed Bandits

YouTube, Youku, BiliBili, TikTok, Netflix use bandits to suggest videos



Dog videos



Gaming videos



Study videos

Multiarmed Bandits

YouTube, Youku, BiliBili, TikTok, Netflix use bandits to suggest videos



Dog videos



Gaming videos



Study videos

The “money” is your ❤️

Multiarmed Bandits

YouTube, Youku, BiliBili, TikTok, Netflix use bandits to suggest videos



Dog videos



Gaming videos



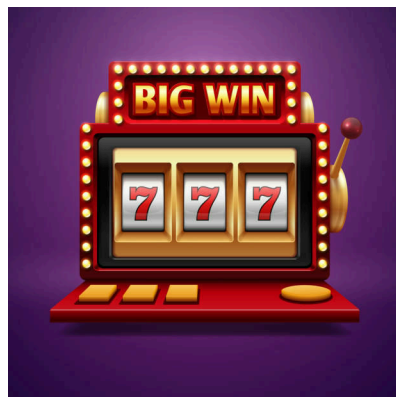
Study videos

The “money” is your ❤️

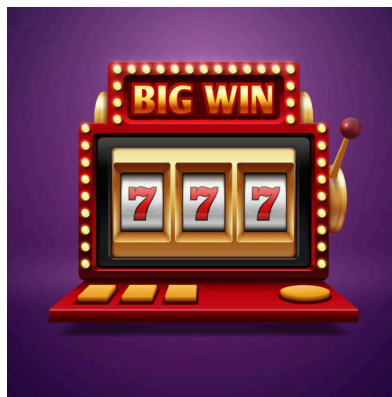
You like a specific type of video, but TikTok does not know what it is

Multiarmed Bandits

YouTube, Youku, BiliBili, TikTok, Netflix use bandits to suggest videos



Dog videos



Gaming videos



Study videos

The “money” is your ❤️

You like a specific type of video, but TikTok does not know what it is

TikTok select videos to maximize your $\mathbb{E}[\text{❤️}]$

Multiarmed Bandits

Problem: We have k bandits, and each bandit is a random variable

$$\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k$$

Multiarmed Bandits

Problem: We have k bandits, and each bandit is a random variable

$$\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k$$

We do not know $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$

Multiarmed Bandits

Problem: We have k bandits, and each bandit is a random variable

$$\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k$$

We do not know $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$

You can take an **action** by pulling the arm of a bandit

$$a \in \{1, 2, \dots, k\}$$

Multiarmed Bandits

Problem: We have k bandits, and each bandit is a random variable

$$\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k$$

We do not know $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$

You can take an **action** by pulling the arm of a bandit

$$a \in \{1, 2, \dots, k\}$$

Which actions should you take to make the most money?

Multiarmed Bandits

This is a hard problem!

Multiarmed Bandits

This is a hard problem!

We need to estimate $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$ to find the best \mathcal{X}

Multiarmed Bandits

This is a hard problem!

We need to estimate $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$ to find the best \mathcal{X}

But it takes ∞ money to find $\mathbb{E}[\mathcal{X}]$!

$$\mathbb{E}[\mathcal{X}] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r_t$$

Multiarmed Bandits

This is a hard problem!

We need to estimate $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$ to find the best \mathcal{X}

But it takes ∞ money to find $\mathbb{E}[\mathcal{X}]!$

$$\mathbb{E}[\mathcal{X}] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r_t$$

Which action $a \in \{1 \dots k\}$ do we choose? Which bandit do we play?

Multiarmed Bandits

This is a hard problem!

We need to estimate $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$ to find the best \mathcal{X}

But it takes ∞ money to find $\mathbb{E}[\mathcal{X}]!$

$$\mathbb{E}[\mathcal{X}] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r_t$$

Which action $a \in \{1 \dots k\}$ do we choose? Which bandit do we play?

We want to:

Multiarmed Bandits

This is a hard problem!

We need to estimate $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$ to find the best \mathcal{X}

But it takes ∞ money to find $\mathbb{E}[\mathcal{X}]!$

$$\mathbb{E}[\mathcal{X}] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r_t$$

Which action $a \in \{1 \dots k\}$ do we choose? Which bandit do we play?

We want to:

- Pick a to estimate bandits

$$\mathbb{E}[\mathcal{X}_a \mid a \in 1 \dots k]$$

Multiarmed Bandits

This is a hard problem!

We need to estimate $\mathbb{E}[\mathcal{X}_1], \mathbb{E}[\mathcal{X}_2], \dots, \mathbb{E}[\mathcal{X}_k]$ to find the best \mathcal{X}

But it takes ∞ money to find $\mathbb{E}[\mathcal{X}]!$

$$\mathbb{E}[\mathcal{X}] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n r_t$$

Which action $a \in \{1 \dots k\}$ do we choose? Which bandit do we play?

We want to:

- Pick a to estimate bandits
- Pick a to make the most money

$$\mathbb{E}[\mathcal{X}_a \mid a \in 1 \dots k]$$

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Multiarmed Bandits

We have names for each goal

Multiarmed Bandits

We have names for each goal

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Multiarmed Bandits

We have names for each goal

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve
our estimate of each random
variable

Multiarmed Bandits

We have names for each goal

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve
our estimate of each random
variable

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Multiarmed Bandits

We have names for each goal

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve
our estimate of each random
variable

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to select the best
bandit and make the most money

Multiarmed Bandits

We have names for each goal

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve
our estimate of each random
variable

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to select the best
bandit and make the most money

Multiarmed Bandits

We have names for each goal

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve
our estimate of each random
variable

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to select the best
bandit and make the most money

It is important to understand the difference between exploration and exploitation! Any questions?

Multiarmed Bandits

Question: How can we choose a to achieve each goal?

Multiarmed Bandits

Question: How can we choose a to achieve each goal?

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve
our estimate of each expectation

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to make money

Multiarmed Bandits

Question: How can we choose a to achieve each goal?

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve our estimate of each expectation

$$a \sim \text{uniform}(\{1 \dots k\})$$

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to make money

Multiarmed Bandits

Question: How can we choose a to achieve each goal?

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve our estimate of each expectation

$$a \sim \text{uniform}(\{1 \dots k\})$$

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to make money

$$a = \arg \max_{a \in \{1 \dots k\}} (\mathbb{E}[\mathcal{X}_a])$$

Multiarmed Bandits

Question: How can we choose a to achieve each goal?

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve our estimate of each expectation

$$a \sim \text{uniform}(\{1 \dots k\})$$

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to make money

$$a = \arg \max_{a \in \{1 \dots k\}} (\mathbb{E}[\mathcal{X}_a])$$

Question: How can we achieve both goals at once?

Multiarmed Bandits

Question: How can we choose a to achieve each goal?

Exploration:

$$\mathbb{E}[\mathcal{X}_a \mid a \in \{1 \dots k\}]$$

Explore our options to improve our estimate of each expectation

$$a \sim \text{uniform}(\{1 \dots k\})$$

Exploitation:

$$\arg \max_{a \in \{1 \dots k\}} \mathbb{E}[\mathcal{X}_a]$$

Use our estimates to make money

$$a = \arg \max_{a \in \{1 \dots k\}} (\mathbb{E}[\mathcal{X}_a])$$

Question: How can we achieve both goals at once?

Answer: Sometimes choose a to explore, sometimes choose a to exploit

Multiarmed Bandits

$$u \sim \text{uniform}([0, 1])$$

if $u < 0.5$ then $a \sim \text{uniform}(\{1 \dots k\})$

if $u \geq 0.5$ then $a = \arg \max(\mathbb{E}[\mathcal{X}_a])$

Multiarmed Bandits

$$u \sim \text{uniform}([0, 1])$$

if $u < 0.5$ then $a \sim \text{uniform}(\{1 \dots k\})$

if $u \geq 0.5$ then $a = \arg \max(\mathbb{E}[\mathcal{X}_a])$

Half the time we explore, half the time we exploit

Multiarmed Bandits

$$u \sim \text{uniform}([0, 1])$$

if $u < 0.5$ then $a \sim \text{uniform}(\{1 \dots k\})$

if $u \geq 0.5$ then $a = \arg \max(\mathbb{E}[\mathcal{X}_a])$

Half the time we explore, half the time we exploit

We can change the explore/exploit ratio using a parameter ε

Multiarmed Bandits

$$u \sim \text{uniform}([0, 1])$$

$$\text{if } u < 0.5 \text{ then } a \sim \text{uniform}(\{1 \dots k\})$$

$$\text{if } u \geq 0.5 \text{ then } a = \arg \max(\mathbb{E}[\mathcal{X}_a])$$

Half the time we explore, half the time we exploit

We can change the explore/exploit ratio using a parameter ε

$$u \sim \text{uniform}([0, 1])$$

$$\text{if } u < \varepsilon \text{ then } a \sim \text{uniform}(\{1 \dots k\})$$

$$\text{if } u \geq \varepsilon \text{ then } a = \arg \max(\mathbb{E}[\mathcal{X}_a])$$

Multiarmed Bandits

$$\varepsilon \in [0, 1]$$

$$u \sim \text{uniform}([0, 1])$$

if $u < \varepsilon$ then $a \sim \text{uniform}(\{1 \dots k\})$

if $u \geq \varepsilon$ then $a = \arg \max(\mathbb{E}[\mathcal{X}_a])$

Multiarmed Bandits

$$\varepsilon \in [0, 1]$$

$$u \sim \text{uniform}([0, 1])$$

$$\text{if } u < \varepsilon \text{ then } a \sim \text{uniform}(\{1 \dots k\})$$

$$\text{if } u \geq \varepsilon \text{ then } a = \arg \max(\mathbb{E}[\mathcal{X}_a])$$

We call this **epsilon greedy**

Multiarmed Bandits

$$\varepsilon \in [0, 1]$$

$$u \sim \text{uniform}([0, 1])$$

$$\text{if } u < \varepsilon \text{ then } a \sim \text{uniform}(\{1 \dots k\})$$

$$\text{if } u \geq \varepsilon \text{ then } a = \arg \max(\mathbb{E}[\mathcal{X}_a])$$

We call this **epsilon greedy**

We take the greedy action (make money) with probability $1 - \varepsilon$

Multiarmed Bandits

$$\varepsilon \in [0, 1]$$

$$u \sim \text{uniform}([0, 1])$$

$$\text{if } u < \varepsilon \text{ then } a \sim \text{uniform}(\{1 \dots k\})$$

$$\text{if } u \geq \varepsilon \text{ then } a = \arg \max(\mathbb{E}[\mathcal{X}_a])$$

We call this **epsilon greedy**

We take the greedy action (make money) with probability $1 - \varepsilon$

Question: When should $\varepsilon \approx 1$? When should $\varepsilon \approx 0$?

Multiarmed Bandits

$$\varepsilon \in [0, 1]$$

$$u \sim \text{uniform}([0, 1])$$

$$\text{if } u < \varepsilon \text{ then } a \sim \text{uniform}(\{1 \dots k\})$$

$$\text{if } u \geq \varepsilon \text{ then } a = \arg \max(\mathbb{E}[\mathcal{X}_a])$$

We call this **epsilon greedy**

We take the greedy action (make money) with probability $1 - \varepsilon$

Question: When should $\varepsilon \approx 1$? When should $\varepsilon \approx 0$?

$\varepsilon \approx 1$ when we trust our estimates
of $\mathbb{E}[\mathcal{X}]$

$\varepsilon \approx 0$ when we do not trust our
estimates of $\mathbb{E}[\mathcal{X}]$

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

- Usually give patients drug A that we know works (exploit)

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

- Usually give patients drug A that we know works (exploit)
- Sometimes test new drug B on patients (explore)

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

- Usually give patients drug A that we know works (exploit)
- Sometimes test new drug B on patients (explore)

Question: Does TikTok or BiliBili use epsilon greedy?

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

- Usually give patients drug A that we know works (exploit)
- Sometimes test new drug B on patients (explore)

Question: Does TikTok or BiliBili use epsilon greedy?

Answer: Yes!

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

- Usually give patients drug A that we know works (exploit)
- Sometimes test new drug B on patients (explore)

Question: Does TikTok or BiliBili use epsilon greedy?

Answer: Yes!

- If you watch dog videos, it usually suggests more dog videos

Multiarmed Bandits

Question: Do we use epsilon greedy in medicine?

Answer: Yes!

- Usually give patients drug A that we know works (exploit)
- Sometimes test new drug B on patients (explore)

Question: Does TikTok or BiliBili use epsilon greedy?

Answer: Yes!

- If you watch dog videos, it usually suggests more dog videos
- Sometimes it suggests study videos, to understand if you like study videos more

Questions?

Coding

Coding

Let us code some multiarmed bandits!

Coding

Let us code some multiarmed bandits!

https://colab.research.google.com/drive/1cyNLRa-J8oe7pgy_gs2mcypZPqqaquoa