DEPARTAMENTO DE INFORMÁTICA
PUC-RIO

Introduction to Probability

Deep Reinforcement Learning

Alberto Sardinha
sardinha@inf.puc-rio.br

What is probability?

- Used to **measure** the likelihood of occurrence of events

Natural tool to model uncertainty

What is probability?

- Classical definition of **probability of event A** (assuming all outcomes are equally likely):

$$P(A) = \frac{N_A}{N}$$

where:

- N_A - number of outcomes of event A
- N – total number of outcomes

What is probability?

- A fair die has 6 possible outcomes:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- What is the probability of rolling an even number (i.e., rolling a two, four or a six)?

  - Event $A = \{2, 4, 6\}$

  - $P(A) = \frac{3}{6} = 0.5$

Formally, …

- **Probability space is a triplet $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \boldsymbol{P})$ where:**
  - Ω is the **sample space** ⬅ Things that can happen
  - $\mathcal{F}$ is the **set of events** ⬅ Things we want to measure
  - P is a **probability measure** ⬅ Way to measure them

Formally, …

- **Sample Space Ω**
  - Space of **possible outcomes**
  - For example:
    - When a single die is thrown, there are six possible outcomes: 1, 2, 3, 4, 5, 6.
    - $\Omega = \{1, 2, 3, 4, 5, 6\}$

Formally, …

- **Set of events $\mathcal{F}$**
  - **Subsets of Ω that we can "measure"** (i.e., assign a probability)
  - For example (throwing a die):
    - Drawing an even number: {2, 4, 6}
    - Drawing a number larger than 3: {4, 5, 6}
    - Drawing no number: Ø (empty set)

Formally, …

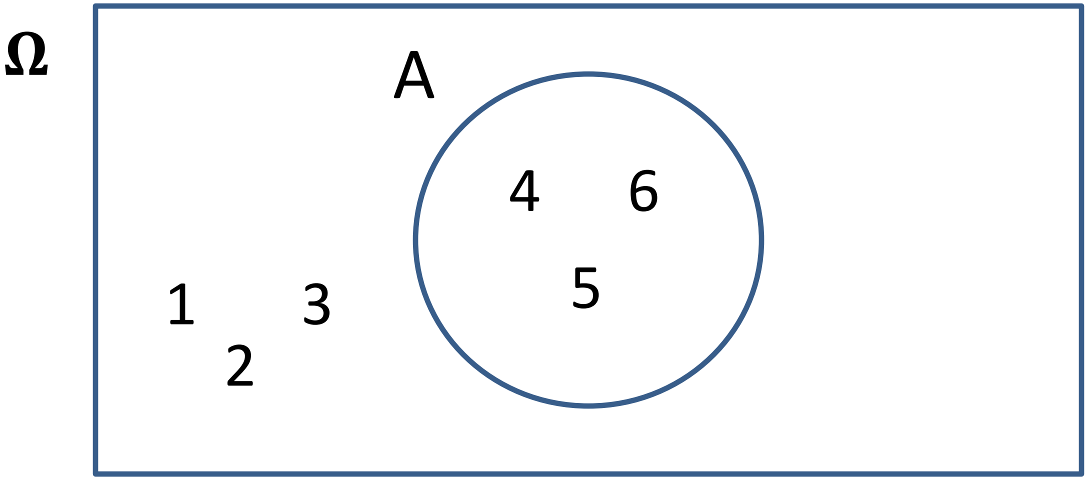
- **Set of events $\mathcal{F}$**
  - Includes the empty set $\emptyset$
  - Includes the full set Ω

Formally, …

- **Probability P**
  - **“Measures” each event in $\mathcal{F}$**
  - For example (throwing a die):
    - What is the probability of drawing a number larger than 3?
      - $A = \{4, 5, 6\}$
      - $P(A) = \frac{3}{6} = 0.5$

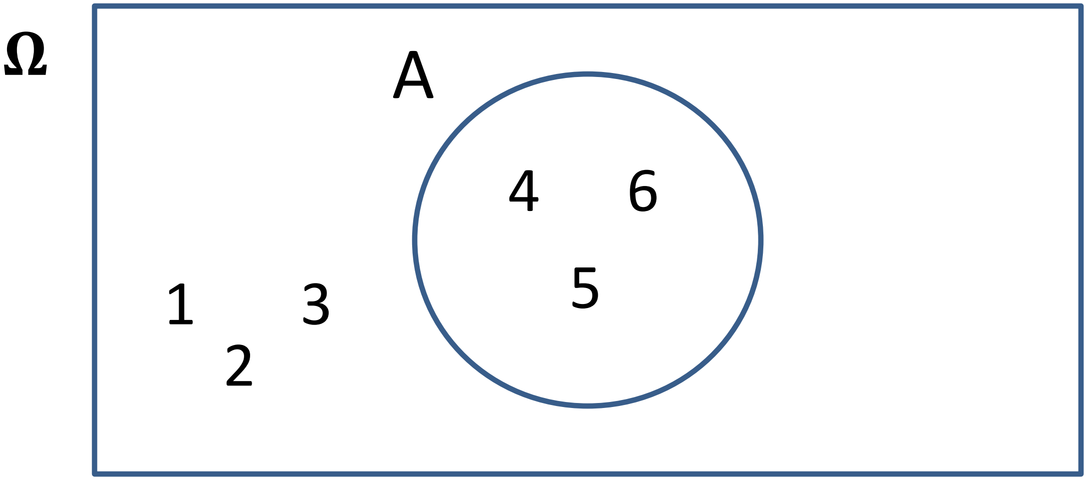
Formally, …

- **Axioms of probability:**
  - $P(A) \geq 0$



Formally, …

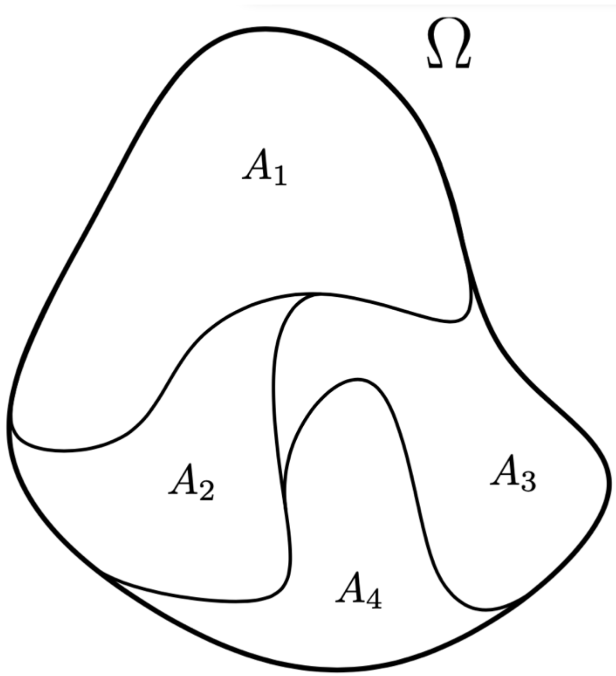
- **Axioms of probability:**
  - $P(A) \geq 0$
  - $P(\Omega) = 1$



Formally, …

- **Axioms of probability:**
  - $P(A) \geq 0$
  - $P(\Omega) = 1$
  - Given disjoint events $A_1, \ldots, A_n$

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$



And all the rest?

$$\boxed{\begin{array}{l} \bullet \quad P(A) \geq 0 \\ \bullet \quad P(\Omega) = 1 \\ \bullet \quad P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) \end{array}}$$

- What is $P(\emptyset)$?

And all the rest?

- $P(A) \geq 0$
- $P(\Omega) = 1$
- $P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$

- What is $P(\emptyset)$?

  - $P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$
  - $1 = P(\Omega) + P(\emptyset)$
  - $1 = 1 + P(\emptyset)$
  - $P(\emptyset) = 0$

Conditional Probability

- Conditional probability of event A given B

- Probability of A occurring, given that B occurred

- Notation:

$$\boldsymbol{P(A|B)}$$

What is the conditional probability?

- A fair die has 6 possible outcomes:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Given an even number is rolled, what is the probability of rolling a 2?

What is the conditional probability?

- Given an even number is rolled, what is the probability of rolling a 2?
  - Reduced sample space: $\Omega_R = \{2, 4, 6\}$
  - $P(2|Even) = \frac{1}{3}$

Conditional Probability

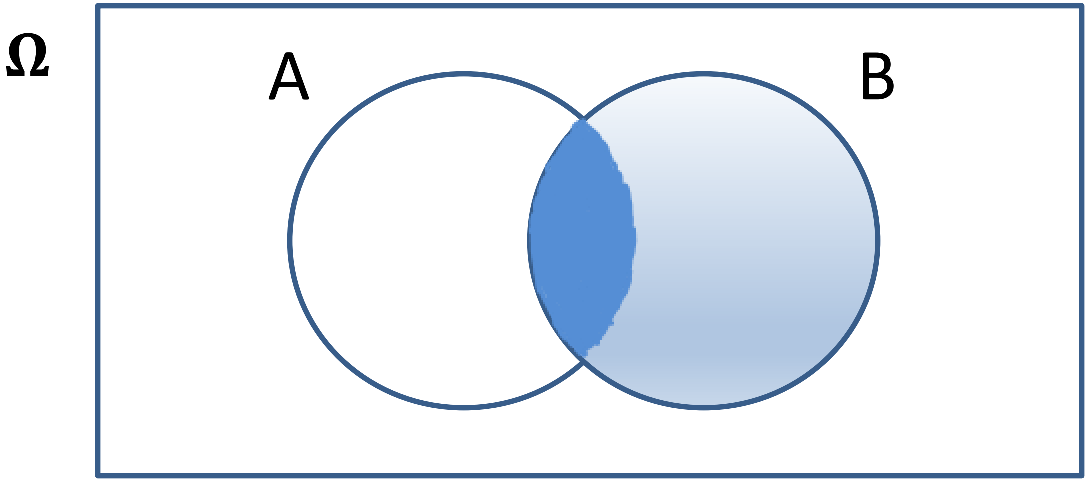
Definition

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

For $P(B) > 0$

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



What is the conditional probability?

- Given an even number is rolled, what is the probability of rolling a 2?
  - $P(2|Even) = ?$
  - $P(2 \cap Even) = \frac{1}{6}$

    $\{2\} \cap \{2, 4, 6\} = \{2\}$
  - $P(Even) = \frac{3}{6}$

    $Even = \{2, 4, 6\}$
  - $P(2|Even) = \frac{P(2 \cap Even)}{P(Even)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$

Conditional Probability

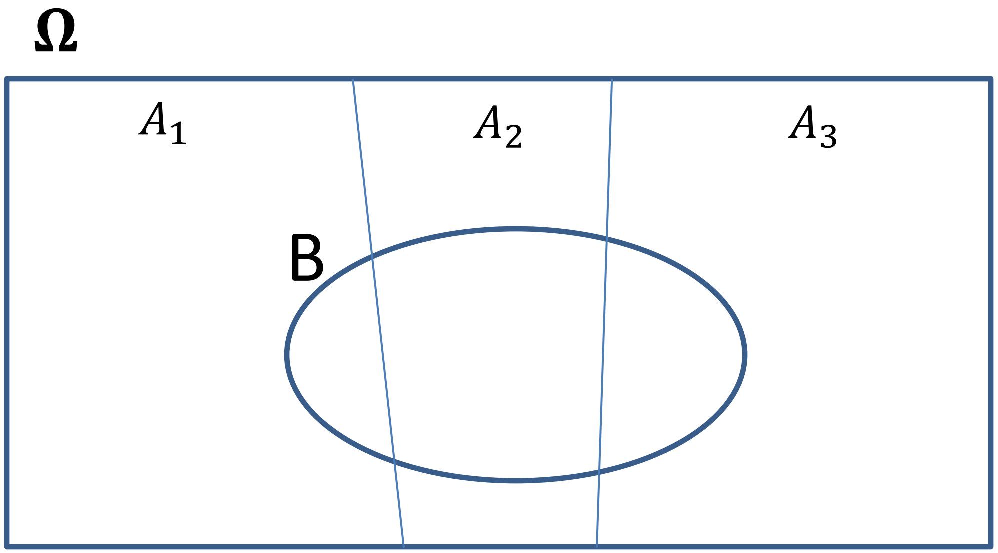
- **Important properties:**
  - **Independent events**:

$$\boxed{\begin{gathered} P(A|B) = P(A) \\ P(B|A) = P(B) \end{gathered}}$$

Conditional Probability

- **Important properties:**
  - **Law of Total Probability**:
    - Given disjoint events $A_1, \ldots, A_n$ with $A_1 \cup \cdots \cup A_n = \Omega$

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$



Conditional Probability

- **Important properties:**
  - **Law of Total Probability**:
    - Given disjoint events $A_1, \dots, A_n$ with $A_1 \cup \cdots \cup A_n = \Omega$

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$

$$\boxed{P(B) = \sum_{i=1}^{n} P(B|A_i).P(A_i)}$$

Conditional Probability

- **Important properties:**
  - **Law of Total Probability**:
    - Given disjoint events $A_1, \ldots, A_n$ with $A_1 \cup \cdots \cup A_n = \Omega$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{\sum_{i=1}^{n} P(B \cap A_i \cap C)}{P(C)}$$

$$P(B|C) = \frac{\sum_{i=1}^{n} P(B|A_i \cap C).P(A_i \cap C)}{P(C)} = \frac{\sum_{i=1}^{n} P(B|A_i \cap C).P(A_i|C).P(C)}{P(C)}$$

$$\boxed{P(B|C) = \sum_{i=1}^{n} P(B|A_i, C).P(A_i|C)}$$

Conditional Probability

- **Important properties:**
  - **Bayes Theorem**:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B).P(B)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \Rightarrow P(A \cap B) = P(B|A).P(A)$$

$$P(A|B).P(B) = P(B|A).P(A)$$

$$\boxed{P(A|B) = \frac{P(B|A).P(A)}{P(B)}}$$

Conditional Probability

- **Important properties:**
  - **Bayes Theorem**:

$$P(\underbrace{A}_{\text{hypothesis}}|\underbrace{B}_{\text{observations}}) = \frac{\overbrace{P(B|A)}^{\text{likelihood}}.\overbrace{P(A)}^{\text{prior}}}{\underbrace{P(B)}_{\text{evidence}}}$$

Random Variable

- Give a probability space $(\Omega, \mathcal{F}, P)$:
  - Sometimes the sample space (Ω) is a cumbersome set to work with
  - It would be more convenient to instead work in some other space, more mathematically convenient
    - For example: $\mathbb{R}$ (real number)

Random Variable

- Give a probability space $(\Omega, \mathcal{F}, P)$:
  - A **random variable** X is a map $X: \Omega \rightarrow E$
    - Where E is a convenient space (e.g., $\mathbb{R}$)
    - We usually work with random variable, ignoring the underlying sample space

Random Variable

- We write $P(X = x)$ to represent

$$P(X = x) = P(\omega \in \Omega : X(\omega) = x)$$

  - If the r.v. takes **values in a discrete set**, we call it a **discrete random variable**

  - If the r.v. takes **values in a continuous set**, we call it a **continuous random variable**

Expectation

- Given a r.v. X and a function $f: X \rightarrow \mathbb{R}$, the **expectation of $\boldsymbol{f}$** is:

$$E[f(X)] = \sum_x f(x).P(X = x)$$

- We can have **conditional expectations**:

$$E[f(X)|Y = y] = \sum_x f(x).P(X = x \mid Y = y)$$

Expectation

- Properties
  - Law of total probability with expectations:

$$E[f(X)] = \sum_y E[f(X)|Y = y].P(Y = y)$$

Thank You

sardinha@inf.puc-rio.br