Probability & Statistics for DS & AI

Introduction to the course

Michele Guindani

Jio Institute

Summer 2022

Probability & Statistics for DS & AI

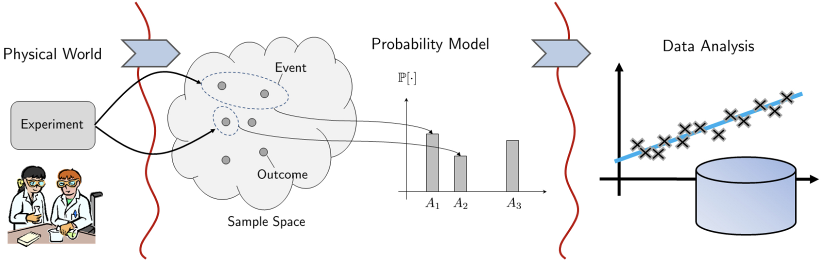
Set Theory and Probability

Michele Guindani

Jio Institute

Summer 2022

- Probability: it is the tool that allows us to describe our world with uncertainty.
- We typically provide an approximate description of phenomena in the real world using the tools of probablity, and then we use probability to allow us to quantify the uncertainty in our description
- It's the base at the foundation of our statistics/data science models and algorithms



Probability and Set theory

- Ch.1 of the textbook provides mathematical backgrounds (skipped, will use what we need)
- We will not got into the phylosophical definition of probability, but we will present the axiomatic framework of Kolmogorov (1933)
- The notion of probability is related to that of an event, which mathematically can be described by the notion of set, a collection of elements

Example (Coin tossing)

- A={obtain head when launching a coin} (event)
- How can we describe such event? First, we can look at the set containing all the results obtained when "launching a coin":

$$\Omega = \{H, T\} \qquad \text{(head, tail)} \Rightarrow \text{universal set}$$

- Then, the result that describes the event

$$A = \{\text{obtain head when launching a coin}\} = \{H\}$$

  can be described by the simple element H.
- Of course, $A \subset \Omega$, A is a subset of Ω
- We will be interested in giving some sense/provide meaning to the following statement

$$P(A) = P(\{H\}) = \frac{1}{2}$$

Example

Die rolling

- B={Launch a regular die and obtain an event number}
- $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\} = \{1, 2, 3, 4, 5, 6\}$ ⇨ universal set (collection of all simple events when launching a die)
- Clearly, $B \subset \Omega$ is a complex event (formed by multiple simple elements) and $B = \{2, 4, 6\}$
- We will be interested in giving meaning to the following statement

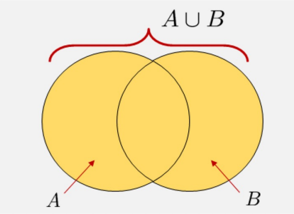
$$P(B) = P(\{\text{even number}\}) = \frac{3}{6} = \frac{1}{2}$$

Set theory

- In order to compute probabilities of complex events, it's often useful to be able to compute and represent an event through operations with sets
- We will recall 4 set operations:
  1. Union of sets
  2. Intersection of sets
  3. Complement of a set
  4. Difference between sets
- The textbook provides a more thorough discussion (invitation to read)

Union of two sets

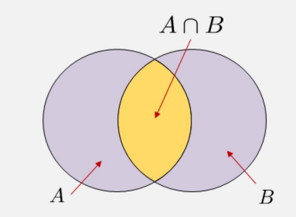
- The union
$$A \cup B = \{\xi \mid \xi \in A \text{ or } \xi \in B\}$$
of two sets contains all elements in A or B.



- Of course, $A \subseteq A \cup B$, $B \subseteq A \cup B$.
- Ex 1: $B = \{\text{even number when launching a die}\} = \{2\} \cup \{4\} \cup \{6\}$
- Ex 2: $A = \{1, 2\}$, $B = \{1, 5\} \Rightarrow A \cup B = \{1, 2, 5\}$
- Ex 3: $A = (3, 4]$, $B = [3.5, \infty)$, $\Rightarrow A \cup B = (3, \infty)$

Intersection of two sets

- The intersection
$$A \cap B = \{\xi \mid \xi \in A \text{ and } \xi \in B\}$$
of two sets contains all elements in A and B.



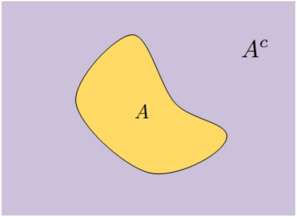
- Of course, $A \cap B \subseteq A$, $A \cap B \subseteq B$.
- Ex 1: $A = \{1, 2, 3, 4\}$, $B = \{1, 5, 6\} \Rightarrow A \cap B = \{1\}$
- Ex 2: $A = (3, 4]$, $B = [3.5, \infty)$, $\Rightarrow A \cap B = [3.5, 4]$
- Ex 3: $A = (3, 4)$, $B = \emptyset$ (empty set) $\Rightarrow A \cap B = \emptyset$

Complement of a set

- The complement of a set A

$$A^c = \{\xi \mid \xi \in \Omega \text{ and } \xi \notin A\}$$

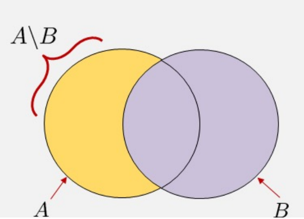
contains all elements that are in Ω but not in A



- Ex 1: Let $A = \{1, 2, 3\}$ and $\Omega = \{1, 2, 3, 4, 5, 6\} \Rightarrow A^c = \{4, 5, 6\}$.
- Ex 2: Let $B = \{\text{even number when launching a die}\} \Rightarrow$ $B^c = \{\text{odd number when launching a die}\}$
- Ex 3: Let $A = [0, 5)$ and $\Omega = \mathbb{R} \Rightarrow A^c = (-\infty, 0) \cup [5, \infty)$.

Difference between two sets

- The difference
$$A \setminus B = \{\xi \mid \xi \in A \text{ and } \xi \notin B\}$$
contains all elements that are in A but not in B.



- Ex 1: Let $A = \{1,3,5,6\}$ and $B = \{2,3,4\} \Rightarrow A \setminus B = \{1,5,6\}$ and $B \setminus A = \{2,4\}$.
- Ex 2: Let $A = [0,1]$, B $= [2,3]$ (non overlapping sets)$\Rightarrow A \setminus B = [0,1]$, and $B \setminus A = [2,3]$.
- It can be shown that $A \setminus B = A \cap B^c$.

Disjoint sets and partitions

- It is important to be able to quantify situations in which two sets are not overlapping (disjoint): $A \cap B = \emptyset$.
- Ex: Let $A = (1, \infty)$ and $B = (-\infty, 0)$ ⇨ A and B are disjoint

Partition

A collection of sets $\{A_1, \ldots, A_n\}$ is a partition of the universal set Ω if

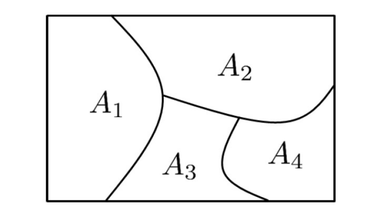
① the set don't overlap, i.e. $\{A_1, \ldots, A_n\}$ is dijoint:

$$A_i \cap A_j = \emptyset$$

② The union of $\{A_1, \ldots, A_n\}$ gives the universal set:

$$\bigcup_{i=1}^{n} A_i = \Omega$$

Example (Examples of Partition)



- Let Ω be the result of rolling a die. Then the event $B = \{2, 4, 6\}$ and $B^c = \{1, 3, 5\}$ form a partition of Ω.
- Let $\Omega = \{1, 2, 3, 4, 5, 6\}$. The following sets form a partition:

$$A_1 = \{1, 2, 3\}, A_2 = \{4, 5\}, A_3 = \{6\}$$

- Sec. 2.1.9: look at other (simple) set operations

Probability & Statistics for DS & AI

Building the concept of Probability from a mathematical perspective

Michele Guindani

Jio Institute

Summer 2022

Sample space

- Sets corresponds to description of "events"⇨ we want to be able to express the probability of events

Sample space Ω

A sample space Ω is the set of all possible outcomes from an experiment. We denote ξ as an element in Ω.

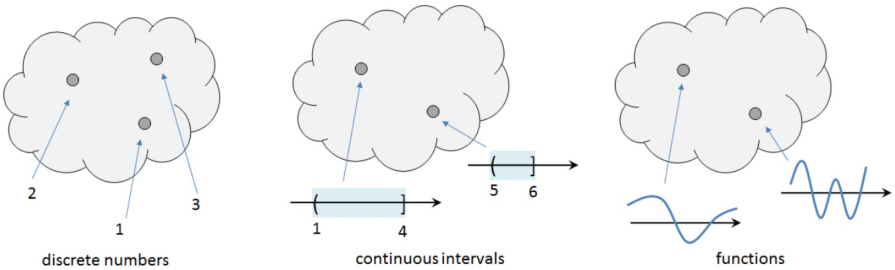
Example (Discrete outcomes)

- Coin flip: $\Omega = \{H, T\}$
- Throw a die: $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$
- Numnber of points by Lebron James in a match: $\Omega : \{0, 1, 2, \ldots\}$
- Recovery from a disease $\Omega = \{\text{No, Yes}\} = \{0, 1\}$.

Example (Continuous Outcomes)

- Waiting time for a bus in Vashi: $\Omega = \{t \mid 0 \leq t \leq 30 \text{ minutes } \}$
- Winnings of golfers on the PCGA (Professionals' golfers association) tour $\Omega : \{x \mid x \geq 0\}$
- Individual height: $\Omega = \{x \mid 0 \leq x \leq 2.5m\}$

Elements in the sample space can be anything.



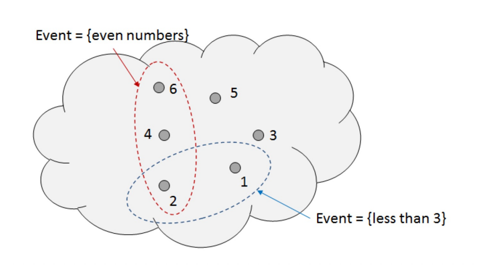
Event space $\mathcal{F}$

- The sample space contains all the possible outcomes. However, in many practical situations, we are not interested in each of the individual outcomes; we are interested in the combinations of the outcomes.

Die rolling

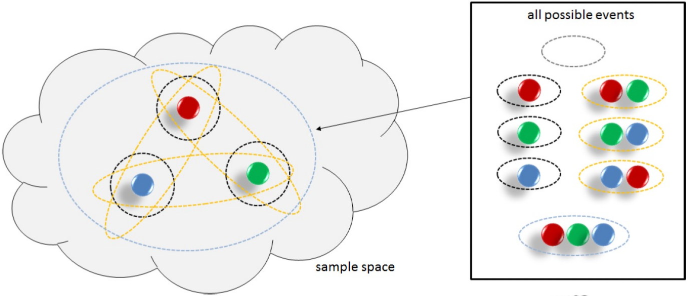
An event E is a subset in the sample space Ω. The set of all possible events is denoted as $\mathcal{F}$.

Example



How many events

- Question: If you have n elements in the sample space, how many events can you construct?
- Solution:



$$2^n$$

(power set)

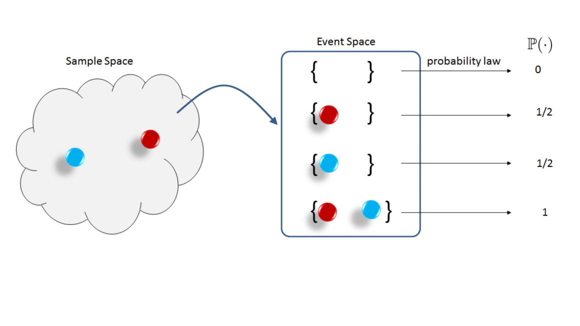
Probability law $\mathbb{P}$

Definition: Part 1

A probability law is a function $\mathbb{P}$ that associates to any event $E \in \mathcal{F}$ a real number in $[0, 1]$.

Example (Coin flip)

The event space $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$.



Probability law $\mathbb{P}$

Definition: Part 2

The function must satisfy three axioms (Kolmogorov, 1933):

① Non-negativity: $\mathbb{P}[A] \geq 0$, for any $A \subseteq \Omega$.

② Normalization: $\mathbb{P}[\Omega] = 1$

③ (Countable) Additivity: For any disjoint sets $\{A_1, A_2, \ldots\}$, it must be true that

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}\left[A_i\right]$$

Why these three axioms?

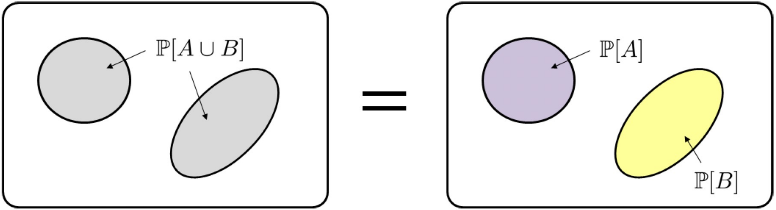
Axiom ① (Non-negativity) ensures that probability is never negative.

Axiom ② (Normalization) ensures that probability is never greater than 1.

Understanding additivity

If A and B are disjoint, then

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B],$$



Why these three axioms?

Axiom ③ (Additivity) allows us to add probabilities when two events do not overlap.

Example (Throwing a fair die)

- $\Omega = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\} = \{1, 2, 3, 4, 5, 6\}$
- Let's consider the probability of getting $\{\text{⚁}, \text{⚃}, \}$:

$$P(\{\text{⚁}, \text{⚃}, \}) = P(\{\text{⚁}\} \cup \{\text{⚃}\}) = P((\{\text{⚁}\}) + P(\{\text{⚃}\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

Example (Baseball)

- In Baseball, a single is the most common type of base hit, accomplished through the act of a batter safely reaching first base by hitting a fair ball.

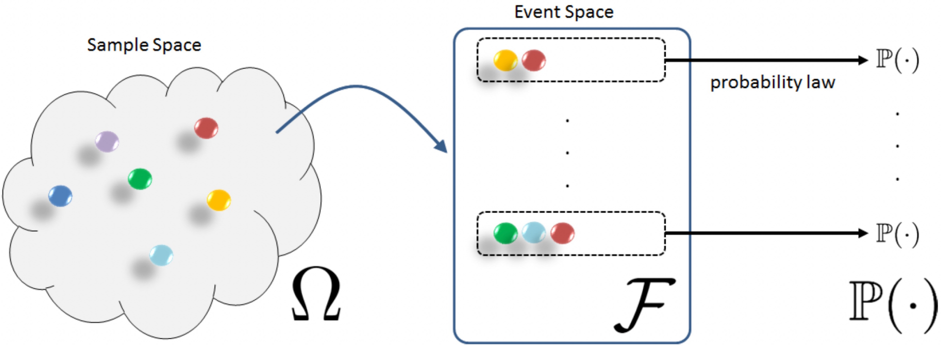
  A double is the act of a batter striking the pitched ball and safely reaching second base.
- Suppose that for a baseball player, $P(S) = 0.2$ and $P(D) = 0.05$. Then

$$P(S \text{ or } D) = P(S \cup D) = P(S) + P(D) = 0.25$$

Probability Space

A probability space consists of a triplet:

$$(\Omega, \mathcal{F}, \mathbb{P})$$



Properties of Probability

1. $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$

2. For any $A \subseteq \Omega, \mathbb{P}[A] \mid\leq 1$

3. $\mathbb{P}[\emptyset] = 0$