# Chapter 1: Basics of Probability

Mariona Segú

Assistant Professor
Thema, CY Cergy Paris Université

Fall 2023

# Contents

# Contents

# Introduction

Definition of Probability

- the mathematical formalism to describe and analyze situations for which we don't have perfect knowledge.
- Probabilistic reasoning is crucial to follow many recent debates in economics, finance and about society.

Probability = a way to describe the genuine risk/chance of events

Inductive reasoning

- how to draw general conclusions from a few observed special cases?
- ex: political polls, surveys, medical trials, etc.

# Introduction

Example - Political Poll

- We observe a subset of individuals in an exit poll (from a population of interest = all voters)
- it is uncertain whether our sample is *representative* of the whole population
- Ensuring this representativity makes heavy use of probabilistic theory

# Introduction

Statistics - definition

- "A branch of mathematics dealing with the collection, analysis, interpretation and presentation of data"
- Statistics "measure properties of a population"

Key Concepts

- *Population*: our object of interest
- *Sample*: a subset of the population that will be used to infer conclusion about it.

"Statistics is concerned with the design of experiments or surveys to obtain a specified quantity of information at minimum cost and the optimum use of this information in making an inference about a population."

# Introduction

Statistics - characterizing a measurement
Numerical descriptive measures

1. Measure of central tendency: the mean.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$\bar{y}$ is the sample mean, $\mu$ is the population mean
Two measurement can have the same mean but a very different distribution.

# Introduction

Statistics - characterizing a measurement

Numerical descriptive measures

2. Measures of dispersion:
    - The variance: measures the deviation from the mean

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- The standard deviation

$$s = \sqrt{s^2}$$

The population parameters are noted $\sigma^2$ and $\sigma$, respectively

# Contents

# 2. Set Theory and Events

Definition 1 A **random experiment** is any procedure which can - at least theoretically

1. be repeated arbitrarily often and under identical conditions
2. has a well-defined set of possible outcomes.

Example:

- a coin flip (two possible outcomes: heads and tails)
- an exit poll: we draw a sample of 2,000 randomly selected voters
  - it is possible redraw arbitrary many alternative samples (condition 1)
  - Outcomes: all possible vote choice times the number of respondants

# 2. Set Theory and Events

Definition 2 The **sample space S** is the collection of all possible outcomes of an experiment.

For many purposes, we are not primarily interested in single outcomes, but instead group collections of outcomes together into events.

Therefore we will in the following describe the experiment in terms of sets.

## 2. Set Theory and Events

Definition 3 An **event** A can be any collection of outcomes (this includes individual outcomes, the empty set, or the entire sample space).

If the realized outcome is a member of the event A, then A is said to occur.

French presidential race

- Sample space S={Macron, Le Pen, Melenchon, Zemmour, Pécresse, Jadot, Others}

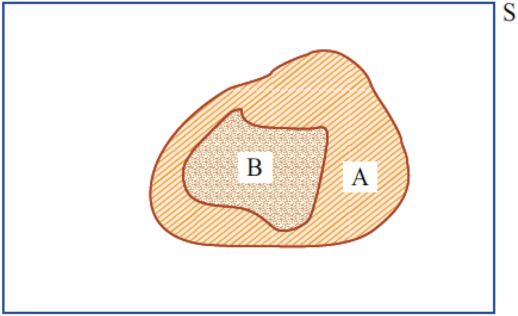
Some events of interest could be:

- "France has a far-right president" = {Le Pen, Zemmour}
- "France has a re-elected president" = {Macron}

## 2.1. Set inclusion $\subset$ (subset)

The event B is contained in A if every outcome in B also belongs to A, or in symbols

$$B \subset A \text{ if } (s \in B \Rightarrow s \in A)$$



# 2.1. Set inclusion $\subset$ (subset)

## Properties

1. Clearly, any event C is contained in the sample space S, i.e.

$$C \subset S \text{ for any event } C$$

2. If A and B contain each other, they are equal,

$$A \subset B \text{ and } B \subset A \Rightarrow A = B$$

3. and set inclusion is transitive, i.e.

$$A \subset B \text{ and } B \subset C \Rightarrow A \subset C$$

# 2.1. Set inclusion $\subset$

## Example

"The French President is 44 years old" $=$ {Macron} $\subset$ "The president is younger than 60" $=$ {Macron, Le Pen, Jadot, Pecresse}

and

{Macron, Le Pen, Jadot, Pecresse} $\subset$ "The President is born in France"

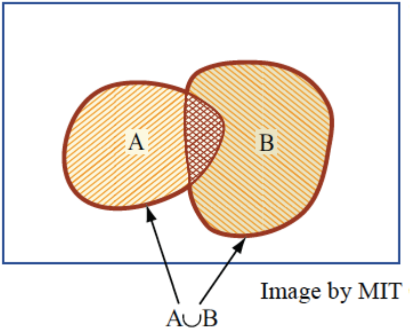
we can conclude that

"The French President is 44 years old" $\subset$ "The President is born in France"

## 2.2. Set union ∪

The union of A and B is the collection of all outcomes that are members of A or B or both. In symbols:

$$A \cup B = \{s \in S \mid s \in A \vee s \in B\}$$



# 2.2. Set union $\cup$

## Properties

1. The set union is symmetric:

$$A \cup B = B \cup A$$

2. Also

$$B \subset A \Rightarrow A \cup B = A \text{ for any } A, B \subset S$$

3. Associative property: the order does not matter

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$$

## 2.3. Set intersection ∩

The intersection of A and B is the (possibly empty) collection of outcomes that are members of both A and B, written as

$$A \cap B = \{s \in S \mid s \in A \wedge s \in B\}$$

where "$\wedge$" denotes the logical "and" . Some texts use the alternative notation $A \cap B =$ AB.

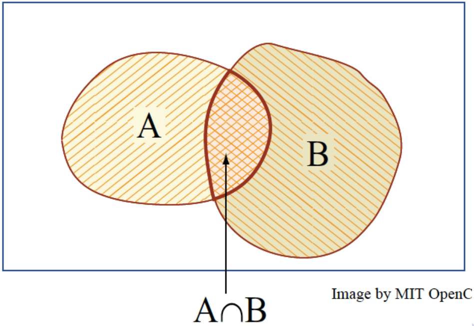



Image by MIT OpenC

# 2.3. Set intersection $\cap$

## Properties

1. The intersection is symmetric:

$$A \cap B = B \cap A$$

2. Also

$$B \subset A \Rightarrow A \cap B = B \text{ for any } A, B \subset S$$

3. Associative property: the order does not matter

$$A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$$

# 2.3. Set intersection

## Distributive properties $\cup$ and $\cap$

In addition, set intersection and union have the distributive properties

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

And

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

## 2.3. Set intersection

### Distributive properties ∪ and ∩ – Example

A="The President is right-wing" $=$ {Macron, Le Pen, Zemmour, Pecresse}

B="The president is a women"={Le Pen, Pecresse}

C="The president is foreign-born"={Melenchon}

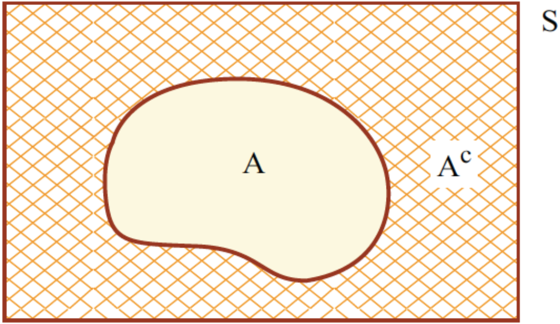
$A \cap (B \cup C)$ ={Macron, Le Pen, Zemmour, Pecresse} $\cap$ {Le Pen, Pecresse, Melenchon}=**{Le Pen, Pecresse}**

$(A \cap B) \cup (A \cap C)$ ={Le Pen, Pecresse} $\cup \varnothing =$ **{Le Pen, Pecresse}**

## 2.4. Set complements $A^C$

The complement $A^C$ of A is the set of outcomes in S which do not belong to A, i.e.

$$A^C = \{s \in S \mid s \notin A\}$$



# 2.4. Set complements $A^C$

## Properties

1. $$(A^C)^C = A$$
2. $$A \cup A^C = S$$
3. $$A \cap A^C = \varnothing \text{ hence } S^C = \varnothing$$

One useful set of relationships between intersections and unions is the following

$$(A \cup B)^C = A^C \cap B^C$$
$$(A \cap B)^C = A^C \cup B^C$$

# 2.5. Partition of Events

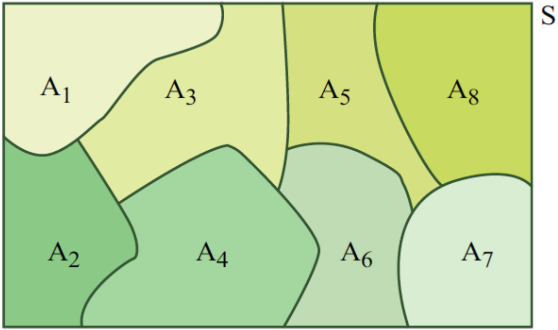
- A and B are **disjoint** (or mutually exclusive) if they have no outcomes in common, i.e.

$$A \cap B = \varnothing$$

- A collection of events $A_1, A_2, \ldots$ is said to be **exhaustive** if their union equals $S$
- A collection of events $A_1, A_2, \ldots$ is called a **partition** of the sample space if
  1. any two distinct events $A_i, A_j$ are disjoint
  2. the collection $A_1, A_2, \ldots$ is exhaustive.

  In a similar fashion we can define partitions of an event $B$ as collections of mutually exclusive subevents whose union equals $B$.

# 2.5. Partition of Events



# Contents

## 3.1 Probability of Events

So far, we have defined events. But how likely are they to happen? To every event $A$ in $S$, we assign a number, $P(A)$, called the *probability* of $A$

$$P : \left\{ A \longmapsto P(A) \right.$$

P satisfies the following axioms:

- (P1) $P(A) \geq 0$
- (P2) $P(S) = 1$ i.e. "for sure, *something* is going to happen"
- (P3) For any sequence of *disjoint* sets $A_1, A_2, ...$

$$P\left(\bigcup A_i\right) = \sum_{i \geq 1} P(A_i)$$

If two events are mutually exclusive, the relative frequency of their union is the sum of their respective relative frequencies

# 3.1 Probability of Events

Definition 4 A probability distribution on a sample space $S$ is a collection of numbers $P(A)$ which satisfies the axioms (P1)-(P3).

- (P1)-(P3) are only the minimal requirements which any probability distribution should satisfy
- In principle any function $P(.)$ satisfying these properties constitutes a valid probability
- The question is how good $P(.)$ describes a random experiment
- In later chapters we will see some popular choices of $P(.)$

## 3.2 Properties of Probabilities

Are (P1)-(P3) enough? Let's verify that

1. The probability that an event happens plus the probability that it doesn't happen should sum to one,
2. The probability that the impossible event, $\varnothing$, happens should equal zero,
3. If an event $B$ is contained in an event $A$, its probability can't be greater than $P(A)$,
4. The probability for any event should be in the interval $[0, 1]$.

# 3.2 Properties of Probabilities

Proof of 1: "The probability that an event happens plus the probability that it doesn't happen should sum to one"

**Proposition 1**

- $1 = P(S)$ as for (P2)
- $P(S)$ is defined as $P(A \cup A^C)$
- $P(A \cup A^C) = P(A) + P(A^C)$ as for (P3)
- We find that $P(A) + P(A^C) = 1$

# 3.2 Properties of Probabilities

Proof of 2: "The probability that the impossible event, $\varnothing$, happens should equal zero"

**Proposition 2**

- $P(\varnothing) = P(S^C)$ as for (P2)
- $P(S^C) = 1 - P(S)$ following proposition 1
- Let's replace $P(S) = 1$ as for (P2)
- $P(\varnothing) = 1 - 1 = 0$
- We have shown that $P(\varnothing) = 0$

## 3.2 Properties of Probabilities

Proof of 3: "If an event B is contained in an event A, its probability can't be greater than P(A)"
Formally, if $B \subset A$ then $P(B) \leq P(A)$

**Proposition 3**

- Let's partition event A
  $A = A \cap S = A \cap (B \cup B^C) = (A \cap B) \cup (A \cap B^C) = B \cup (A \cap B^C)$
- Let's prove that $B$ and $A \cap B^C$ are disjoint (their interaction should be null) which allow us to apply (P3)
- $B \cap (A \cap B^C) = B \cap (B^C \cap A) = (B \cap B^C) \cap A = \varnothing \cap A = \varnothing$
- We conclude that $P(A) = P(B) + P(A \cap B^C) \geq P(B)$

# 3.2 Properties of Probabilities

Proof of 4: "The probability for any event should be in the interval $[0, 1]$"
Formally, for any event $A$, $0 \leq P(A) \leq 1$

**Proposition 4**

- $0 \leq P(A)$ is axiom (P1)
- which also implies that $P(A^C) \geq 0$
- Using proposition 1 $P(A) = 1 - P(A^C) \leq 1$

# 3.2 Properties of Probabilities

**Proposition 5**

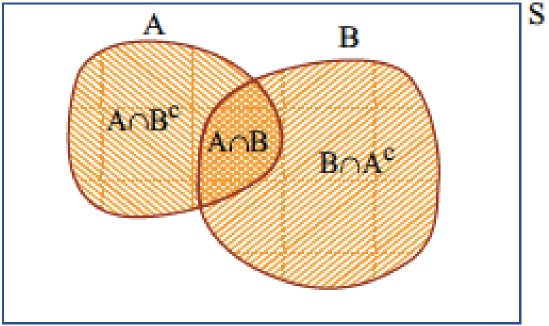
Let's prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- As in proposition 3, we partition A and B
  $A = (A \cap B) \cup (A \cap B^C)$ and $B = (B \cap A) \cup (B \cap A^C)$
- Since the sets are disjoint: $P(A) = P(A \cap B) + P(A \cap B^C)$ and $P(B) = P(B \cap A) + P(B \cap A^C)$
- Then $P(A) + P(B) = P(B \cap A) + [P(A \cap B^C) + P(B \cap A) + P(B \cap A^C)] = P(B \cap A) + P(A \cup B)$

# 3.2 Properties of Probabilities

**Proposition 5**

Since $[A \cap B^C, B \cap A, B \cap A^C)]$ is a partition of $A \cup B$



## 3.3 Example of Simple Probabilities

- Suppose we have a finite sample space
- with symmetric outcomes (all outcomes are equally likely)
- Let $n(A)$ denote the number of outcomes in an event A, then we can define a probability:

$$P(A) = \frac{n(A)}{n(S)}$$

- i.e. the probability equals the fraction of all possible outcomes in S included in A
- This distribution is called the "simple" probability distribution or the "logical" probability

# 3.3 Example of Simple Probabilities

Suppose a fair die is rolled once.

- Sample space equals $S = \{1, 2, ..., 6\}$, so $n(S) = 6$.
- What is the probability of rolling a number strictly greater than 4? $\Rightarrow$ Event A: $A = \{5, 6\}$
- $n(A) = n(\{5, 6\}) = 2$.
- Hence $P(A) = \frac{n(A)}{n(S)} = \frac{2}{6} = \frac{1}{3}$

# 3.3 Example of Simple Probabilities

Suppose a fair die is rolled twice.

- Sample space equals
  $S = \{(1,1),(1,2)...,(2,1),(2,2),...,(6,6)\}$, so
  $n(S) = 6^2 = 36$.
- what is the probability that the sum of the numbers is less than or equal to 4?
- Event B="Sum of dice $\leq 4$"
  $B = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}$
- Hence $P(B) = \frac{n(B)}{n(S)} = \frac{6}{36} = \frac{1}{6}$

# Contents

# 4. Counting Rules

So far... it was easy to count the outcomes in $A$ and $S$.

- If $S$ is very large and $A$ is complex, counting the outcomes might be tedious... hard to obtain $n(A)$ and $n(s)$
- We need to use counting rules to count the number of combinations or permutations

# 4. Counting Rules

Multiplication Rule: If an experiment has 2 parts, where the first part has $m$ possible outcomes, and the second part has $n$ possible outcomes regardless of the outcome in the first part, then the experiment has $m \times n$ outcomes.

**Example**

If a password is required to have 8 characters (letters or numbers), then the corresponding experiment has 8 parts, each of which has $2 \times 26 + 10 = 62$ outcomes (assuming that the password is case-sensitive). Therefore, we get a total of $62^8$ (roughly 218 trillion) distinct passwords. Clearly, counting those up by hand would not be a good idea.

# 4. Counting Rules

Sampling with(out) replacement

**Example**

A card deck has 52 cards, so if we draw one card each from a blue and a red deck, we get $52 \times 52 = 2704$ possible combinations of cards. If, on the other hand, we draw two cards from the same deck without putting the first card back on the stash, regardless of which card we drew first, only 51 cards will remain for the second draw. Therefore, if we draw two cards from the same deck, we'll have $52 \times 51 = 2652$ possible combinations.

## 4.1 Permutation

Permutation: Any **ordered** rearrangement of $N$ objects

1. Permutation **with replacement**, $k$ draws

$$N \times N \times ... \times N = N^k$$

2. Permutation **without replacement**: Sampling k draws from a group of N without replacement, where $N \leq k$

$$P_{N,k} = N(N-1)(N-2)...(N-(k-1)) = \frac{N(N-1)(N-2)...3 \cdot 2 \cdot 1}{(N-k)(N-(k-1)...3 \cdot 2 \cdot 1} = \frac{N!}{(N-k)!}$$

$k! = 1 \cdot 2 \cdot ...(k-1)k$ and it is read as *k-factorial* (with $0! = 1$)

# 4.1 Permutation – Examples

### Choosing Officers

Suppose that a club consists of 25 members and that a president and a secretary are to be chosen from the membership. We shall determine the total possible number of ways in which these two positions can be filled. The first person chosen is the president, the second is the secretary. How many combinations are there?

### Arranging Books

Suppose that six different books are to be arranged on a shelf. How many combinations are there?

# 4.1 Permutation – Examples

## Choosing Officers

Since the positions can be filled by first choosing one of the 25 members to be president and then choosing one of the remaining 24 members to be secretary, the possible number of choices is $P_{25,2} = \frac{25!}{23!} = (25)(24) = 600$.

## Arranging Books

The number of possible permutations of the books is $P_{6,6} = \frac{6!}{(6-6)!} = 6! = 720$.

We can conclude that, when $N = k$, then $P_{N,N} = N!$

## 4.2 Combinations

Combination: Any **unordered** collection of objects

**Poker hand**

If we want to count how many different poker hands we can draw from a single deck of cards (5 cards drawn from a single deck without replacements), we don't care about the order in which cards were drawn, but rather whether each of the card was drawn at all.

A combination is a draw without replacement from a group, but since we now do not care about the order, we don't want to double-count series of draws which consist of the same elements, only in different orders.

# 4.2 Combinations

For n elements, the number of orders in which we can draw them is equal to $n!$
Hence, a combination of k objects from N is:

$$C_{N,k} = \frac{\#\text{outcomes of permutation without replacement}}{\#\text{orders in which we can draw n elements}} = \frac{\frac{N!}{(N-k)!}}{k!}$$

$$C_{N,k} = \frac{N!}{(N-k)!k!} = \binom{N}{k}$$

This number is known as the **binomial coefficient**

# 4.2 Combinations

## Poker hand

$$C_{52,5} = \binom{52}{5} = \frac{52!}{47!5!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47!}{47!5!}$$

$$= \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2} = 2598960$$

# 4.2 Combinations

**Study group**

A functional study group should not have more than, say, 5 members. There are currently 28 students registered for a class. How many possibilities for viable study groups (including students working on their own) would be possible? We'll have to calculate the number of study groups for each group size 1, 2, 3, 4, 5 and add them up...

# 4.2 Combinations

## Study group

$$S = \binom{28}{1} + \binom{28}{2} + \binom{28}{3} + \binom{28}{4} + \binom{28}{5} =$$

$$\frac{28!}{27!1!} + \frac{28!}{26!2!} + \frac{28!}{25!3!} + \frac{28!}{24!4!} + \frac{28!}{23!5!} =$$

$$28 + 378 + 3276 + 20475 + 98280 = 122437$$

# 4. Counting rules and probabilities

Let's go back to the simple probabilities formula

$$P(A) = \frac{n(A)}{n(S)}$$

We can now use the counting rules to determine $n(A)$ and $n(S)$ in complex cases.

**Card Decks**

Draw two cards from a deck of 52 cards with replacement, assuming that each card is picked with equal probability. What is the possibility of drawing two different cards?

# 4. Counting rules and probabilities

## Card Decks

Let's count the outcomes in the denominator

$$S = \{(A\clubsuit, A\clubsuit), (A\clubsuit, A\spadesuit), ...\} \Longrightarrow n(S) = 52^2$$

The numerator consists of the event A="two different cards", equivalent to a permutation without replacement

$$S = \{(A\clubsuit, A\spadesuit), (A\clubsuit, A\heartsuit), ...\} =$$

$$\Rightarrow n(A) = P_{N,k} = P_{52,2} = \frac{52!}{(52-2)!} = 52 \cdot 51$$

So that $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{52!}{(52-2)!52^2} = \dfrac{51}{52} = 0.98$

# 4. Counting rules and probabilities

Sometimes, the best way to compute a probability of an event $A$ is through its complement $A^C$

**The birthday problem**

What is the probability $p$ that at least two people in a group of $k$ people will have the same birthday?
Again, let's use

$$P(A) = \frac{n(A)}{n(S)}$$

and the counting rules.
Any wild guess if $k = 50$?

# 4. Counting rules and probabilities

## The birthday problem

Each of the $k$ people can have 365 possible birthdays. Hence, $n(S) = 365^k$

Now, let's count the number of outcomes in which all k birthdays are different (instead of the same):

- This is a permutation without replacement
  $n(A^C) = P_{365,k} = \frac{365!}{(365-k)!}$

Hence $P(A^C) = \frac{P_{365,k}}{365^k} = \frac{365!}{(365-k)! \cdot 365^k}$

And $P(A) = 1 - P(A^C) = 1 - \frac{365!}{(365-k)! \cdot 365^k}$

How much is $P(A)$ if $k = 3$?

# 4. Counting rules and probabilities

## The birthday problem

| k | P(A) |
|---|---|
| 3 | 0.082 |
| 5 | 0.027 |
| 10 | 0.117 |
| 15 | 0.253 |
| 20 | 0.411 |
| 25 | 0.569 |
| 30 | 0.706 |
| 50 | 0.97 |
| 365 | 1 |

# Contents

# 5. Independence and Conditional probability

## Independent Events

Intuitively, we want to define a notion that for two different events $A$ and $B$ the occurrence of $A$ does not "affect" the likelihood of the occurrence of $B$.

**Coin toss**

If we toss a coin two times, the outcome of the second toss should not be influenced in any way by the outcome of the first.

First, let's simplify notation and denote

$$P(A \cap B) = P(AB)$$

# 5.1. Independent Events

Definition 5 The events A and B are said to be independent if

$$P(A \cap B) = P(AB) = P(A)P(B)$$

Independence is merely a property of the probability distribution, not necessarily of the physical nature of the events.
So while in some examples (e.g. the series of coin tosses) we have a good intuition for independence, in most cases we'll have no choice but need to check the formal condition.

# 5.1. Independent Events

## A fair die

Say we roll a fair die once, what are the probabilities for the events $A = \{2, 4, 6\}$, $B = \{1, 2, 3, 4\}$ and their intersection $P(AB)$?

# 5.1. Independent Events

## A fair die

Say we roll a fair die once, what are the probabilities for the events $A = \{2, 4, 6\}$ $B = \{1, 2, 3, 4\}$ and their intersection $P(AB)$?

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2} \text{ and } P(B) = \frac{n(B)}{n(S)} = \frac{4}{6} = \frac{2}{3}$$

$$P(AB) = P(\{2, 4\}) = \frac{2}{6} = \frac{1}{3} = \frac{1}{2} \times \frac{2}{3} = P(A)P(B)$$

so the events are independent even though they resulted from the same roll.

# 5.1. Independent Events

One interpretation of independence is the following:

- Suppose we know that $B$ has occurred, does that knowledge change our beliefs about the likelihood of $A$ (and vice versa)?
- It will turn out that if $A$ and $B$ are independent, there is nothing about event $A$ to be learned from the knowledge that $B$ occurred

## 5.1. Independent Events

**Proposition 6** If $A$ and $B$ are independent, then $A$ and $B^C$ are also independent.
PROOF: Since $A$ is the union of events $AB$ and $AB^C$, we can apply (P3) and rewrite

$$P(A) = P(AB) + P(AB^C) \Rightarrow P(AB^C) = P(A) - P(AB)$$

Given that A and B are independent...

$$P(AB^C) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^C)$$

## 5.1. Independent Events

Definition 6 A collection of events $A_1, A_2, \ldots$ are independent if for any subset $A_{i1}, A_{i2}, \ldots$ of these events (all indices being different)

$$P(A_{i1} \cap A_{i2} \cap \ldots) = P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots$$

### For 3 events

For three events A, B, C,

$$P(AB) = P(A)P(B), \qquad P(AC) = P(A)P(C),$$
$$P(BC) = P(B)P(C)$$

and

$$P(ABC) = P(A)P(B)P(C)$$

## 5.1. Independent Events

**Example: 3 independent events?**

Let the sample space be $S = \{s_1, s_2, s_3, s_4\}$, and $P(s_i) = \frac{1}{4}$ for all outcomes. Then each of the events

$$A = \{s_1, s_2\}, \qquad B = \{s_1, s_3\}, \qquad C = \{s_1, s_4\}$$

occurs with probability $\frac{1}{2}$. The probability for the event $A \cap B$ is

$$P(AB) = P(\{s_1\}) = \frac{1}{4} = P(A)P(B)$$

# 5.1. Independent Events

## Example: 3 independent events?

Likewise for any other pair of events, so the events are pairwise independent.
However, taken together, the full collection is not independent since

$$P(ABC) = P(\{s_1\}) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

Intuitively, once we know that both $A$ and $B$ occurred, we know for sure that $C$ occurred.

# 5.2. Conditional Probability

Suppose the occurrence of $A$ affects the occurrence (or non-occurrence) of $B$ and vice versa.
How do we describe the probability of $B$ given knowledge about $A$?

### A fair die

If we throw a fair die and the outcome is an even number, i.e the event $B = \{2, 4, 6\}$ occurred. What's the probability of having rolled a 6? Since there are only three equally likely possibilities in $B$, 6 being one of them, we'd intuitively expect the answer to be $\frac{1}{3}$. Here we basically simplified the sample space, to $S = B = \{2, 4, 6\}$, and calculated the simple probability for the redefined problem.

## 5.2. Conditional Probability

Definition 7 Suppose $A$ and $B$ are events defined on $S$ such that $P(B) > 0$. The conditional probability of $A$ given that $B$ occurred is given by

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

Intuitively, the numerator *redefines* which outcomes in $A$ are possible once $B$ has occurred. The possibilities are no longer the whole sample space but only the outcomes comprised in $B$.

# 5.2. Conditional Probability

### A fair die

If $B = \{2, 4, 6\}$ has occurred, What is the probability of having rolled a 6?

$$P(A \cap B) = P\{6\} = \frac{1}{6}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

## 5.2. Conditional Probability

Remark 1 Conditional probability and independence: if $A$ and $B$ are independent,

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

so $B$ occurring tells us nothing about $A$, so the conditional probability is the same as the unconditional probability.

# 5.3 Two Laws of Probability

We define two laws that give the probability of unions and intersections.

Theorem 1 (The Multiplicative Law of Probability) The probability of the *intersection* of two events is

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

If $A$ and $B$ are independent, then

$$P(A \cap B) = P(A)P(B)$$

## 5.3 Two Laws of Probability

Theorem 2 (The Additive Law of Probability) The probability of the *union* of two events is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A$ and $B$ are mutually exclusive then $P(A \cap B) = 0$, and

$$P(A \cup B) = P(A) + P(B)$$

# 5.3 Two Laws of Probability

**Example**

Two events $A$ and $B$ such that $P(A) = 0.2$, $P(B) = 0.3$ and $P(A \cup B) = 0.4$. Find

1. $P(A \cap B)$
2. $P(A^C \cup B)$
3. $P(A^C B^C)$
4. $P(A^C|B)$

# 5.3 Two Laws of Probability

**Example**

Two events $A$ and $B$ such that $P(A) = 0.2$, $P(B) = 0.3$ and $P(A \cup B) = 0.4$. Find

1. $P(A \cap B)$
   - We use the Additive law to state...
   - $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.2 + 0.3 - 0.4 = 0.1$

# 5.3 Two Laws of Probability

**Example**

Two events $A$ and $B$ such that $P(A) = 0.2$, $P(B) = 0.3$ and $P(A \cup B) = 0.4$. Find

2. $P(A^C \cup B)$

- Using the Additive Law, $P(A^C \cup B) = P(A^C) + P(B) - P(A^C \cap B)$
- We know that $P(B) = P(A \cap B) + P(A^C \cap B)$
- which implies $P(A \cap B) = P(B) - P(A^C \cap B)$
- Hence $P(A^C \cup B) = 1 - P(A) + P(A \cap B)$
- Hence $P(A^C \cup B) = 1 - 0.2 + 0.1 = 0.9$

# 5.3 Two Laws of Probability

## Example

Two events $A$ and $B$ such that $P(A) = 0.2$, $P(B) = 0.3$ and $P(A \cup B) = 0.4$. Find

3. $P(A^C B^C)$

- Using the Additive Law, $P(A^C B^C) = P(A^C) + P(B^C) - P(A^C \cup B^C)$
- $P(A^C B^C) = (1 - P(A)) + (1 - P(B)) - P(A^C \cup B^C)$
- Using set complements properties, $A^C \cup B^C = (A \cap B)^C$
- Hence $P(A^C \cup B^C) = 1 - P(A \cap B)$
- $P(A^C B^C) = (1 - P(A)) + (1 - P(B)) - (1 - P(A \cap B))$
- $P(A^C B^C) = (1 - 0.2) + (1 - 0.3) - (1 - 0.1) = 0.6$

# 5.3 Two Laws of Probability

## Example

Two events $A$ and $B$ such that $P(A) = 0.2$, $P(B) = 0.3$ and $P(A \cup B) = 0.4$. Find

4. $P(A^C|B)$

- We know that

$$P(A^C|B) = \frac{P(A^C \cap B)}{P(B)} = \frac{P(A^C) + P(B) - P(A^C \cup B)}{P(B)}$$

- $$P(A^C|B) = \frac{1 - 0.2 + 0.3 - 0.9}{0.3} = \frac{0.2}{0.3} = \frac{2}{3}$$

# Contents

# 6. Bayes Theorem

To distinguish it from the conditional probabilities $P(A|B_i)$, $P(A)$ is also called the marginal probability of A. The relationship between marginal and conditional probabilities is given by the **Law of Total Probability:**

Theorem 3 (Law of Total Probability) Suppose that $S = B_1 \cup B_2 ... \cup B_n$ where $P(B_i) > 0, i = 1, 2, ..., n$ and $B_i \cap B_j = \varnothing$ for $i \neq j$. Then

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

for any event A.

# 6. Bayes Theorem

Proof of **Law of Total Probability**
Any subset $A$ of $S$ can be written as

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_k) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k).$$

Thus,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k) \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i). \end{aligned}$$

# 6. Bayes Theorem

- Law of Total Probability: relates the unconditional probability $P(A)$ to the conditional probabilities $P(A|B_i)$.
- **Bayes Law**: relates the conditional probability $P(A|B)$ to the conditional probability $P(B|A)$, i.e. how we can revert the order of conditioning.
  - Very important result in many areas of statistics and probability,
  - most importantly in situations in which we "learn" about the "state of the world" A from observing the "data" B.

# 6. Bayes Theorem

### Praying to Poseidon

The ancient Greeks noticed that each time a ship sunk, all surviving seamen reported having prayed to Poseidon. They inferred that they were saved from drowning because they had prayed.
Let's define the events A ="survives" and B ="prayed", did praying increases the odds of survival?, i.e. whether $P(A|B) > P(A) \equiv p$, say. If all survivors prayed, then $P(B|A) = 1$. Is that information sufficient to say that praying strictly increases the chances of survival? How do we use the information on $P(B|A)$ to learn about $P(A|B)$?

# 6. Bayes Theorem

From the Multiplicative Law of Probability, we obtain

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

Rearranging the second equality, we get

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

We've also seen that we can partition the event

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

so that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

# 6. Bayes Theorem

## Praying to Poseidon

We know that $P(B|A) = 1$, and the (unconditional) survival rate of seamen, $P(A) = p$. By definition, $P(A^C) = 1 - p$. We still need to know $P(B|A^C)$, did the dead seamen pray before drowning?.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{p}{p + P(B|A^C)(1-p)}$$

It is probably safe to assume that those, fearing for their lives, all of them prayed as well (implying $P(B|A^C) = 1$), so that

$$P(A|B) = \frac{p}{p + (1-p)} = p = P(A)$$

# 6. Bayes Theorem

## Survival Bias

- The reasoning of the ancient Greeks is an instance of "survivor bias" (in a very literal sense)
- Bayes theorem shows us that if we can only observe the survivors, we can't make a judgment about why they survived unless we know more about the sub-population which did not survive.
- In this case, we need to know about $P(B|A^C)$

# 6. Bayes Theorem

Let's generalize this to any partition of $S$:

**Theorem 2 (Bayes Theorem)** If $A_1, A_2, \ldots$ is a partition of $S$, for any event $B$ with $P(B) > 0$ we can write

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\Sigma_{j\geq 1} P(B|A_j)P(A_j)}$$

- $P(A_i)$ is the *prior probability* of an event $A_i$ (i.e. probability before experiment is run)
- $P(A_i|B)$ is the *posterior probability* of $A_i$ (i.e. the probability after we ran the experiment and got information $B$)

# 6. Bayes Theorem

**A medical test**

Suppose a doctor tests a patient for a very nasty disease, (A is the event of having the disease). The test can either give a positive result (event B) or a negative result (event $B^C$). The test is not fully reliable, the probabilities of a positive test result is
$P(B|A) = 99\%$, $P(B|A^C) = 5\%$
The disease affects 0.5% of the population. Let's say the test gives a positive result. What is the (conditional) probability that the patient does in fact have the disease?

# 6. Bayes Theorem

## A medical test

Bayes rule gives that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} =$$

$$\frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.05 \cdot 0.995} = 0.0905$$

- Even a positive test result gives only relatively weak evidence for disease
- This is because the overall prevalence of the disease, $P(A)$ is relatively low

# 6. Bayes Theorem

- In real-life situations, most people aren't very good at this type of judgments and tend to overrate the reliability of a test like the one from the last example
- This phenomenon is known as *base-rate neglect*, where in our example "base-rate" refers to the proportions $P(A)$ and $P(A^C)$ of infected and healthy individuals.
- If these probabilities are very different, biases in intuitive reasoning can be quite severe.

**Time for another example?**

# 6. Bayes Theorem

**Trust in Science**

A population of voters contains $40\%$ Republicans and $60\%$ Democrats. It is reported that $30\%$ of the Republicans and $70\%$ of the Democrats declare to have trust in science. A person chosen at random from this population is found to favor the issue in question. Find the conditional probability that this person is a Democrat.

# 6. Bayes Theorem

## Support for Science

Given: $P(R) = 0.40$ (Republican), $P(D) = 0.60$ (Democrat), $P(T|R) = 0.30$ (Republican trusts in science), $P(T|D) = 0.70$ (a Democrat trusts in science):

We are looking for $P(D|T)$, the probability that a person is a Democrat given that they trust in science. Using Bayes' theorem:

$$P(D|T) = \frac{P(T|D) \times P(D)}{P(T|R) \times P(R) + P(T|D) \times P(D)}$$

Substituting into Bayes' theorem:

$$P(D|T) = \frac{0.70 \times 0.60}{0.30 \times 0.40 + 0.70 \times 0.60} \approx 0.7778$$

So, given that a person trusts in science, there is approximately a $77.78\%$ chance they are a Democrat.