



Introduction to Probability Theory

Statistics I

Disclaimer

These slides are a free translation and adaptation from the slide deck for Estatística I by Prof. Sandra Custódio and Prof. Teresa Ferreira from the Lisbon Accounting and Business School, Polytechnic University of Lisbon.

Probability, background

The concept

Consider the following scenario:



Investor: What's the probability this startup will succeed?



Analyst: Hard to say, every startup is different.



Investor: But if you had to guess, based on similar cases?



Analyst: Maybe 1 in 3 succeed under these conditions.



Investor: So, would you bet on it?



Analyst: Yes, I would.



Investor: Even if the odds aren't great?



Analyst: I believe this one has what it takes.

The concept

Here we can define probability in terms of **frequency of occurrence**, *i.e.* as a percentage of successes in a moderately large number of similar situations.

This is the most natural and traditional way of thinking about probability.

- Regarding a fair coin  we could say: “with probability 50% the coin lands on heads” meaning “roughly half of the time.”

But, what if this company belongs to a completely novel market sector?

The concept

There might be situations where the frequency concept is not adequate, because it might refer to a one-time event. These are **subjective beliefs**.

- A company is recruiting a new CEO, and a board member says:
“I believe there’s a 90% chance that our chosen candidate will be an effective CEO.”

The concept

It might seem easy to disregard the second case as unscientific or useless. However, many times people need to make decisions under uncertainty with not enough data (or no data at all!) about previous realizations of the specific event.

Beliefs allow the decision maker to, well, make *some* decision, at least consistently.

- What's the difference between both situations?
- What do they have in common?

Uncertainty

A refresh on Set Theory

Sets and elements

A **set** is a collection of objects, which are **elements** of the set.

Definition

Let S represent a set, and s an element of that set, we write $s \in S$ to mean s **belongs** to S .

If s does not belong to S , we write $s \notin S$.

Definition

If a set S does not have any element, then it is the **empty** set, denoted by $\emptyset$.

How to write down a Set

There are several ways to specify a set.

By extension, or as a list:

- If a set S has a finite number of elements ($x_i \in S$) we can write it like this:

$$S = \{x_1, x_2, \dots, x_n\}$$

- If a set S has an infinite (but countable) elements ($x_i \in S$) we can write it like:

$$S = \{x_1, x_2, \dots\}$$

How to write down a Set

By describing the property (P) that x must satisfy to be included in S :

$$S = \{x | x \text{ satisfies } P\}$$

in this case $|$ reads as *such that*. For example

$$S = \{x \in \mathbb{R} | x \geq 0\}$$

to describe the non-negative real numbers.

This example is special, as the positive real numbers cannot be written down as a list. In this case the interval $[0, \infty)$ is an **uncountable** set.

More definitions

Definition

If $\forall x \in S$ it is also true that $x \in T$, then we say that S is a **subset** of T , and we write it like $S \subseteq T$.

Definition

If $S \subseteq T$ and at the same time $T \subseteq S$ then we say that S and T are **equal**, and we write it $S = T$.

Definition

The **universal set** Ω is the set that contains all objects that could conceivably be of interest in a particular context.

By definition the, any set S must be a subset of Ω .

More definitions

The universal set is important because it defines the scope of our analysis. Say we are studying the performance of students of Statistics I in 2026.

The cars parked outside our institution do not belong to the universal set, because they are not relevant for our purpose. Only students of Statistics I in 2026 belong to the universal set.

Set Operations

Definition

The **complement** of a set S , with respect to Ω , is the set $\{x \in \Omega | x \notin S\}$, that is, all the relevant elements that do not belong in S . We denote it as S^c .

Corollary: It is easy to see that $\Omega^c = \emptyset$.

Set Operations

Definition

The **union** of two sets S and T is the set of all elements that belong to S *or* T (or both), and is denoted by $S \cup T$.

$$S \cup T = \{x \in \Omega | x \in S \vee x \in T\}$$

Definition

The **intersection** of two sets S and T is the set of all elements that belong to S *and* T , and is denoted by $S \cap T$.

$$S \cap T = \{x \in \Omega | x \in S \wedge x \in T\}$$

Note that $\vee$ stands for *or*, and $\wedge$ stands for *and*.

Set Operations

Sometimes we might need to consider the union or intersection of many sets, and for that we can use a notation similar to the one we used for summations:

- $\bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cup \dots = \{x \in \Omega \mid x \in S_n \text{ for some } n\}$
- $\bigcap_{n=1}^{\infty} S_n = S_1 \cap S_2 \cap \dots = \{x \in \Omega \mid x \in S_n \text{ for every } n\}$

Set Operations

Definition

Two sets (say S and T) are said to be **disjoint** if $S \cap T = \emptyset$.

More generally, a collection of sets S_n is disjoint if S_i and S_j are disjoint when $i \neq j$.

Definition

A collection of sets is said to be a **partition** of a set S if the sets in the collection are:

- Disjoint
- Their union is S

We can use the notation of $\mathcal{P}(\Omega)$

Set Definition

The number of elements of a set S is known as its **cardinality** and it is denoted as $\#S$. $\#S$ satisfies:

1. $\#S \geq 0$
2. $\#\emptyset = 0$

Set definition

If we have two sets, S and T , we define the operation *set minus* as $\setminus$ the set that contains all elements of S that do not belong to T

$$S \setminus T = \{x \in S | x \notin T\}$$

? Set Theory · Question 1

Two sets S and T are **disjoint** when:

A. $S \cup T = \emptyset$

B. $S = T$

C. $S \cap T = \emptyset$

D. $S \subseteq T$

 **C.** Disjoint means they share no element at all, so their **intersection** is empty. Their union is empty only if both sets are.

? Set Theory · Question 2

The set $S \setminus T$ contains exactly the elements that:

A. belong to both S and T

B. belong to T but not to S

C. belong to neither

D. belong to S but not to T

✓ **D.** $S \setminus T = \{x \in S \mid x \notin T\}$. Note the operation is not symmetric: $S \setminus T$ and $T \setminus S$ are different sets.



Set Theory · Question 3

Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{1\}$, $B = \{3, 6\}$ and $C = \{2, 4, 6\}$.

Write down $A \cup B$, $B \cap C$, $(B \cup C)^c$ and $C \setminus B$.



Set Theory · Solution

$$A \cup B = \{1, 3, 6\}$$

$$B \cap C = \{6\}$$

$$B \cup C = \{2, 3, 4, 6\}, \text{ so } (B \cup C)^c = \{1, 5\}$$

$$C \setminus B = \{2, 4\}$$

Back to Probability

In probability, Ω , the universal set, is a non-empty set that contains all possible **outcomes** of an experiment. Each outcome is represented by ω , and obviously $\omega \in \Omega$.

Back to Probability

The sample space (Ω) can be:

- Discrete, when $\#\Omega$ is finite, or countable infinite.
- Continuous, when $\#\Omega$ is uncountable.

Back to Probability

Consider the experiment of throwing a die  and noting the number shown on side facing upwards.

1. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$
2. In this case $\#\Omega = 6$
3. Ω is discrete.

Back to Probability

Consider now the random experiment of measuring the life expectancy of a lamp  , measured in hours.

1. The sample space is $\Omega = \{x \in \mathbb{R} | x \geq 0\}$
2. In this case, Ω is all non-negative real numbers.
3. Ω is continuous.

Remember, Ω must include all possible outcomes from your experiment!
Even then ones that seem ludicrous.

Back to Probability

Definition

A subset A of the sample space Ω is called an **event**.

$$A \subseteq \Omega$$

By definition then Ω is also an event.

Definition

We call the **realization** of an event A if, after an experiment, outcome ω is realized, and $\omega \in A$.

Example

Let's go back to our experiment with the 

The sample space is: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Within this space, we can define the following events:

1. $A = \{1, 3, 5\}$, *i.e.* the number is odd.
2. $B = \{3, 4, 5, 6\}$, *i.e.* the number is at least 3.
3. $C = \{1, 2, 3\}$, *i.e.* the number is lower than 4.
4. $D = \{6\}$, *i.e.* the number is larger than 5.

Example

Now let's revisit the example of our 

The sample space is: $\Omega = \{x \in \mathbb{R} | x \geq 0\}$

In this space, we can define the following events:

1. $A = \{x \in \mathbb{R} | 75 < x < 95\}$, i.e. the  lasts between 75 and 95 hours.
2. $B = \{x \in \mathbb{R} | x \leq 100\}$, i.e. the  lasts no longer than 100 hours.
3. $C = \{x \in \mathbb{R} | x \geq 60\}$, i.e. the  lasts at least 60 hours.

Events

Definitions

- An **elementary event** is any event that contains a single element (*i.e.* $\#A = 1$)
- An **impossible event** is an event with no outcome, (*i.e.* $\#A = 0$). As a consequence, an impossible event coincides with the empty set $\emptyset$.
- A **certain event** is indeed the event Ω , as for any outcome we obtain ω , this outcome belongs to the sample space Ω by definition.

Mixing up Sets and Probability

Consider two events A and B both subsets of Ω

1. A^c contains all the outcomes that are not in A . A^c is the event of *not* A .
2. If $A \subseteq B$, then an outcome that realizes event A ($\omega \in A$), also realizes B , as $A \subseteq B \Rightarrow \omega \in B$ as well. $A \Rightarrow B$
3. For $A \cup B$ to happen, we need $\omega \in A$ **or** $\omega \in B$, which means that A happens, or B happens, or both happen simultaneously.

Mixing up Sets and Probability

4. For $A \cap B$ to happen, we need $\omega \in A$ **and** $\omega \in B$, which means that A and B happen simultaneously.
5. A and B are incompatible if $A \cap B = \emptyset$, *i.e.* if an outcome is in one set, it cannot be in another, for example it cannot be that A and A^c happen simultaneously!

Inherited set properties for events

Consider two events A and B both subsets of Ω

1. Commutativity: $A \cup B = B \cup A$; $A \cap B = B \cap A$
2. Associativity: $(A \cup B) \cup C = A \cup (B \cup C)$;
 $(A \cap B) \cap C = A \cap (B \cap C)$
3. Distributivity: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4. De Morgan's Laws: $(A \cap B)^c = A^c \cup B^c$; $(A \cup B)^c = A^c \cap B^c$

Inherited set properties for events

5. $(A^c)^c = A$

6. Complement law: $A \cup A^c = \Omega$; $A \cap A^c = \emptyset$

7. Identity element: $A \cup \emptyset = A$; $A \cap \Omega = A$

8. Absorbing element: $A \cup \Omega = \Omega$; $A \cap \emptyset = \emptyset$

9. Idempotent law: $A \cup A = A$; $A \cap A = A$

10. $A \subset B \Rightarrow A \cap B = A$; $A \subset B \Rightarrow A \cup B = B$

Example

Consider the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$, from our  case.

Define the events:

$$A = \{1\}, \quad B = \{3, 6\}, \quad C = \{2, 4, 6\}, \quad D = \{4, 5, 6\}$$

Example

Let's define the following events in Ω

- $A \cup B$
- $A \cap B$
- A^c
- $(A \cup B)^c$
- $(B \cap C)^c$
- $B \setminus C$
- $C \setminus D$

Concept of probability

Besides the concepts we already saw of frequency and subjectivity for probability, there was an older, called “*classic*” one. This one was introduced by Pierre-Simon **Laplace** in 1812.

Laplace or Classic interpretation of probability

Let A be an event defined over a finite Ω . The probability of event A is defined as:

$$P(A) = \frac{\#A}{\#\Omega}$$

Example

Consider now an experiment throwing two dice 🎲 🎲

1. How many outcomes are in Ω ? $6^2 = 36$, $\#\Omega = 36$.
2. Let A be the event where both dice show the same number:

$$A = \{(1, 1), (2, 2), \dots, (6, 6)\}$$

here $\#A = 6$

3. The probability that both dice show the same number is

$$P(A) = \frac{\#A}{\#\Omega} = \frac{6}{36} = \frac{1}{6}$$

Laplace or Classic interpretation of probability

The problem with this interpretation, is that we cannot use it, or it becomes meaningless, it when Ω is uncountable or infinite. Also, what if the outcomes are not equally likely? (*i.e.* if the dice are not fair?)

Frequency interpretation

This is today still the dominant interpretation of probability.

In this case, what we want is to observe several independent repetitions of the experiment. After a while, some statistical regularity begins to emerge.

Frequency interpretation

Logically, if you run an experiment, and are interested in the probability of event A , then the events you are registering are A and A^c or *not* A .

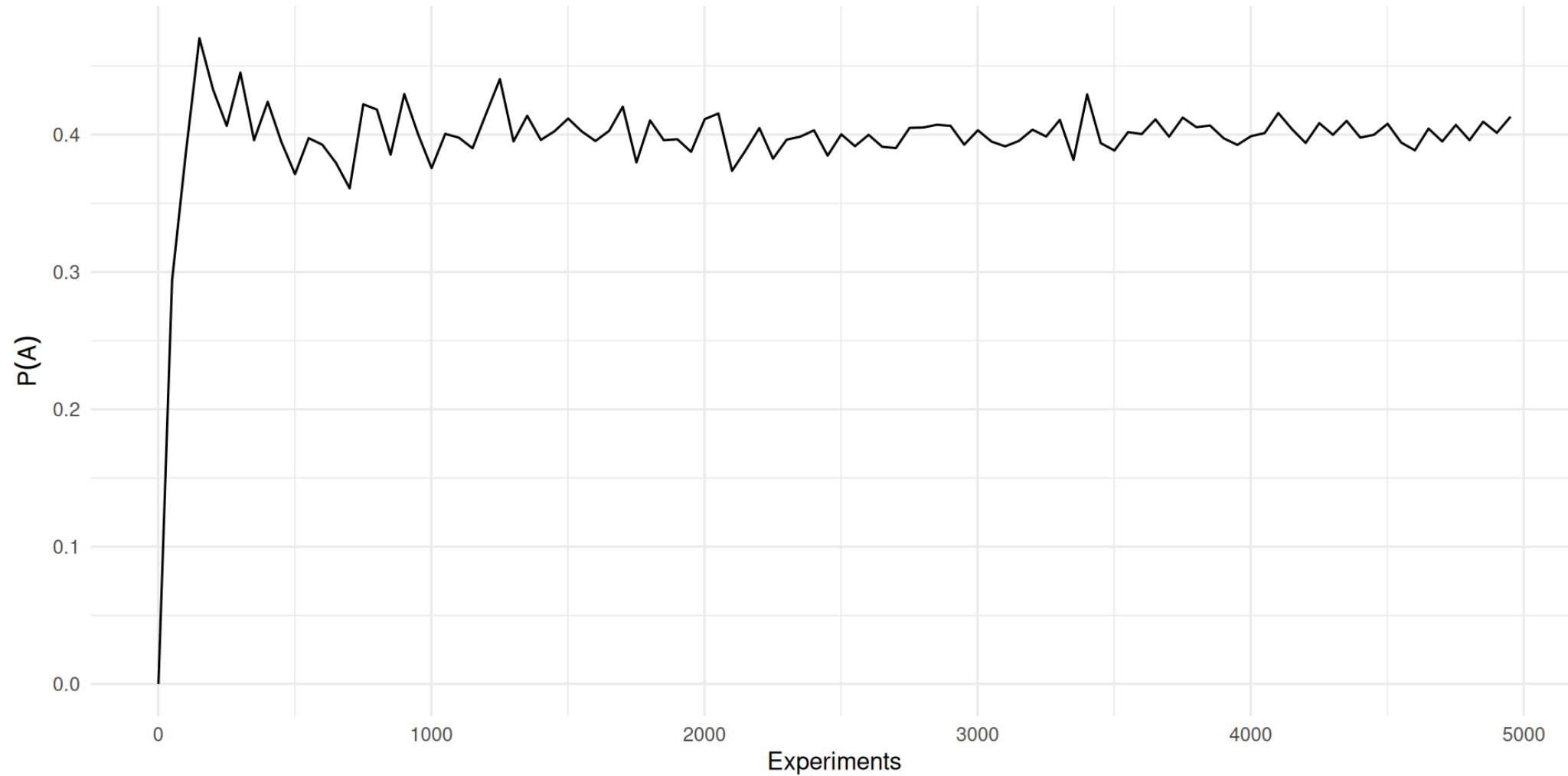
Every time you run your experiment, you count when you get an A and when you observe an A^c event. Obviously, the total number of experiments is how many times you observed A and how many times you observed A^c .

Example:

Experiment: Draw a random number in the interval $[0, 1]$. A denotes $x < 0.4$.

A	A^c	N	$P(A)$
0	1	1	0
2	8	10	0.2
17	33	50	0.34
39	61	100	0.39
217	283	500	0.434
802	1198	2000	0.401

Example:



Frequency interpretation

As you can see, the more experiments we run, the more stabilized the ratio of occurrences for A over the total number of experiments. More generally:

$$P(A) = \lim_{N \rightarrow \infty} \frac{A \text{ occurrences}}{N - \text{Number of Experiments}}$$

This is the relative frequency of A in N experiments: f_A

Not always possible to repeat that many times the experiment in the **same** conditions.

About our previous example

It seems that the probability that the random number between 0 and 1 is below 0.4 is approximately 40%. The more experiments we run, the closer our relative frequency is to that number.

$$P(A) \xrightarrow{N \rightarrow \infty} 0.4$$

By the way, we will see later that theoretically, indeed $P(A) = 0.4$

? The Concept of Probability · Question 1

You roll two fair dice. The probability that the sum equals 7 is:

A. $\frac{1}{12}$

B. $\frac{1}{6}$

C. $\frac{7}{36}$

D. $\frac{1}{9}$



B. There are 6 favourable outcomes out of 36:

$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$, so $P = \frac{6}{36} = \frac{1}{6}$.

? The Concept of Probability · Question 2

Under the **frequency** interpretation, $P(A)$ is:

A. a subjective degree of belief

B. always $\frac{\#A}{\#\Omega}$

C. the limit of the relative frequency of A as the number of experiments grows

D. always 0.5

✓ **C.** The frequency interpretation repeats the experiment many times and takes $P(A) = \lim_{N \rightarrow \infty} \frac{\text{occurrences of } A}{N}$. The $\frac{\#A}{\#\Omega}$ formula is the *classical* definition instead, and it needs equally likely outcomes.



The Concept of Probability · Question 3

Roll two fair dice. Let A be “both dice show the same number” and B be “the sum is at least 10”.

Compute $P(A)$, $P(B)$ and $P(A \cap B)$.



The Concept of Probability · Solution

$$\#\Omega = 36.$$

$$A = \{(1, 1), \dots, (6, 6)\}, \text{ so } P(A) = \frac{6}{36} = \frac{1}{6}.$$

$$B = \{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}, \text{ so } P(B) = \frac{6}{36} = \frac{1}{6}.$$

$$A \cap B = \{(5, 5), (6, 6)\}, \text{ so } P(A \cap B) = \frac{2}{36} = \frac{1}{18}.$$

Probability Axioms

Probability

Andrey Kolmogorov defined a set of characteristics that any probability P measure should have, these are called the Kolmogorov's axioms (1933):

1. $P(A) \in \mathbb{R}$ and $P(A) \geq 0$, for any event $A \subseteq \Omega$.
2. $P(\Omega) = 1$
3. For any A and B disjoint, $P(A) + P(B) = P(A \cup B)$

Corollary

Let A and B be some events in Ω

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. $A \subset B \Rightarrow P(A) \leq P(B)$
4. $0 \leq P(A) \leq 1, \forall A$
5. $P(A \setminus B) = P(A) - P(A \cap B)$

Corollary

$$6. P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) +$$

$$7. \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right)$$

Conditional Probability

Conditional probability, as the wording implies, means the probability of something happening **given** something else has happened. Now, note “something” here makes reference to an *event*.

$$P(A|B)$$

It reads *the probability of A, given B*.

Conditional Probability

Note that if we think on sets, saying *given* B we are immediately excluding everything that could have happened if B did not happen, and therefore our Universal set is no longer Ω , but B .

What we are looking for are, among the events that live in B , how many of those live in A (because those would *trigger* event A). Actually, we are interested on the relative measure of those outcomes, compared to the whole size of B :

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

Example

Consider a factory that makes 10 wrenches . Among those, we know that 2 have imperfections. Suppose you intend to remove, randomly, 2  from the lot (of 10). Consider the following events:

$A = \{\text{The first :wrench: is faulty}\}$ $B = \{\text{The second :wrench: is faulty}\}$

What if we want to compute $P(B)$? For a correct assessment for B , we would better have some information on the realization of A !

Example

If the first  was faulty, then A happened. If the first  was *ok*, then A^c happened, and therefore we can compute $P(B|A)$ and $P(B|A^c)$. We are assuming that we are removing these  without replacing them.

Let's see why this last detail (replacing the ) is so relevant before going on.

Example: With replacement

Initial set:

1	2	3	4	5	6	7	8	9	10
									

Remove one (if randomly you do not know which), but after you remove you can see what happened, let's take out 7.

1	2	3	4	5	6	7	8	9	10
									

We observe, and *voilà* it was a fine wrench . If we replace it though, we would be picking from

1	2	3	4	5	6	7	8	9	10
									

That is in the exact same conditions we made our first choice, and therefore what happens with the first pick is irrelevant: These events are now independent!

Example: with replacement

- With A , we know that when we pick the second wrench (B), in the box there are 2  and 8 .
- With A^c , we know that when we pick the second wrench (B), in the box there are 2  and 8 .

The probability of getting a  is the same in each scenario!

$$P(B|A) = P(B|A^c)!$$

$$P(B|A) = \frac{2}{10} = 0.2 \text{ and } P(B|A^c) = \frac{2}{10} = 0.2$$

Example: no replacement

Initial set:

1	2	3	4	5	6	7	8	9	10
									

Remove one (if randomly you do not know which is broken)

1	2	3	4	5	6	7	8	9	10
									

We observe, and *voilà* it was a broken wrench  A happened!

1	2	3	4	5	6	7	8	9	10
									

We observe, and *voilà* it was a fine wrench  A^c happened!

Example: no replacement

- With A , we know that when we pick the second wrench (B), in the box there are 1  and 9 .
- With A^c , we know that when we pick the second wrench (B), in the box there are 2  and 9 .

The probability of getting a  is different in each scenario!

$$P(B|A) \neq P(B|A^c)!$$

$$P(B|A) = \frac{1}{9} = 0.111 \text{ and } P(B|A^c) = \frac{2}{9} = 0.222$$

Conclusion Conditional Probability

So formally

Definition

Let $A, B \subset \Omega$, then we say the probability of A given B is the conditional probability:

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

Note that from here, we can obtain also $P(A \cap B) = P(A|B)P(B)$. In both situations we need $P(B) \neq 0$.

Corollary

It follows that the identity

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

or

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

with $P(A) \neq 0$ also holds true.

Corollary

Now let's think on $P(A \cap B \cap C)$:

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

Corollary

Note that given the commutativity of the intersection, we could have obtained also:

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

And to make sense of all of this we need $P(X) > 0$, $P(X \cap Y) > 0$ with $X, Y \in \{A, B, C\}$.

Example

Consider a region with 1,000 adults. Their job data is captured by the following table:

	Employed	Unemployed	Total
Women	470	55	525
Men	430	45	475
Total	900	100	1,000

1. Randomly selecting a person in this region, what is the probability this person is:
 1. Woman
 2. Unemployed
 3. Unemployed woman

Example

Let's define the events:

$$W = \{Woman\}, M = \{Man\}, U = \{Unemployed\}$$

1. Woman: $P(W) = \frac{525}{1000} = 0.525$

2. Unemployed: $P(U) = \frac{100}{1000} = 0.1$

3. Unemployed woman: $P(W \cap U) = \frac{55}{1000} = 0.055$

Example

2. A citizen is randomly chosen from the population, and it happens to be a woman. What is the probability she is unemployed?

$$P(U|W) = \frac{P(U \cap W)}{P(W)} = \frac{0.055}{0.525} = 0.105$$

3. A citizen is randomly chosen from the population, and it happens to be unemployed. What is the probability this citizen is a woman?

$$P(W|U) = \frac{P(W \cap U)}{P(U)} = \frac{0.055}{0.1} = 0.55$$

Independent Events

Definition

Two events A and $B \subset \Omega$, are probabilistically independent if and only if:

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

Independent Events

From the definition of independence, we can obtain several properties.

Let A and B independent events with $P(A)P(B) > 0$:

1. $P(A|B) = P(A)$ and $P(B|A) = P(B)$ (remember conditional example with replacement)
2. A^c and B are independent, as well as A and B^c , and even A^c and B^c .
3. If A and B are incompatible, they cannot be independent.
$$P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$$
4. Any event is independent of Ω and $\emptyset$.

Example

Are W and U from the previous example independent?

$$P(W) = 0.525, P(U) = 0.1, P(W|U) = 0.55, P(U|W) = 0.105.$$

Note that $P(W) \neq P(W|U)$ and $P(U) \neq P(U|W)$. Therefore, **they cannot be independent.**

Example

Consider a die  that is thrown twice. Consider the following two events:

$A = \{\text{The die shows an odd number the first time}\}$

$B = \{\text{The die shows a number } > 4 \text{ the second time}\}$

Are A and B independent events?

Example

In this case, $\Omega = \{(x, y) \in \mathbb{N}^2 \mid x, y \leq 6\}$ with $\#\Omega = 6^2 = 36$

- $P(A) = \frac{18}{36} = \frac{1}{2}$
- $P(B) = \frac{12}{36} = \frac{1}{3}$
- $P(A \cap B) = \frac{1}{6} = P(A)P(B)$
- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{2} = P(A)$
- $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1}{3} = P(B)$
- They are independent events!

Remark

Two events being independent is not the same that they being incompatible:

<i>A</i> and <i>B</i> independent	<i>A</i> and <i>B</i> incompatible
$P(A \cap B) = P(A)P(B)$	$P(A \cap B) = 0$
$P(A B) = P(A)$ and $P(B A) = P(B)$	$P(A B) = 0$ and $P(B A) = 0$

Example

Let A and B be two events such that: $P(A) = 0.6$, $P(B) = t$, and $P(A \cup B) = 0.8$

Find t such that A and B are:

1. Mutually exclusive or incompatible.
2. Independent.

Example

1. In this case, what we need is that $P(A \cap B) = 0$.

From probability theory, we have

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

and therefore we get:

$$0.8 = 0.6 + t \Rightarrow t = 0.2$$

Example

2. To make A and B independent, we need that

$$P(A \cap B) = P(A)P(B) = 0.6t:$$

Using the same identity we just used:

$$0.8 = 0.6 + t - 0.6t \Rightarrow t = 0.5$$

Law of Total Probability

Theorem

Let $\{A_i\}_{i=1}^n$ be a partition of Ω , or $\{A_i\} \in \mathcal{P}(\Omega)$. Then, for any $B \subset \Omega$, it holds that:

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

Example

Consider a financial institution that sells two products, α and β , with very high yields. It is known that, among its clients, 10% invest a share of their wealth in α and the rest in β . From those who invest in α , 70% manage to get returns above the market. From among those who do not invest in α , 55% get returns above the market. Randomly choosing a client of this firm, find the probability this customer gets a return above the market.

Example

Let's define the events:

- A_1 the client invest in α
- A_2 the client invest in β
- B has returns above the market.

Matching with the available data we obtain:

$$P(A_1) = 0.1, P(A_2) = 0.9, P(B|A_1) = 0.7, \text{ and } P(B|A_2) = 0.55.$$

From the Law of Total Probability:

Example

$$P(B) = \sum_{i=1}^2 P(A_i \cap B)$$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$$

$$P(B) = 0.7 \times 0.1 + 0.55 \times 0.9 = 0.565$$

Bayes Theorem

Bayes Theorem

Let events $A_1, A_2, \dots, A_n$ with $n \in \mathbb{N}$ a partition of Ω , then, for any event $B \subset \Omega$, with $P(B) > 0$:

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

with $i = 1, 2, \dots, n$

Note that this is a consequence of the Law of Total Probability.

Bayes Theorem

On the other side, $\sum_i P(A_i) = 1$ and $\sum_i P(A_i|B) = 1$

Bayes Theorem has been widely used in economics, in biomedical sciences, and social sciences when looking for causality.

If event B represents *consequences* and event A_i probable *cause*, Bayes Theorem allows to assess the probability of this cause ($P(A_i)$).

Example

Let's go back to the previous example, about our investors.

Let's compute the probability that the client invested his money on product β , but given that the client had returns above the market (event B).

Example

If the customer invested in β , then the event we are trying to is A_2 , but conditional on event B , $P(A_2|B)$:

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{P(A_2) \times P(B|A_2)}{\sum_i P(A_i) \times P(B|A_i)}$$

We knew from the previous exercise that $P(B) = 0.565$, and therefore we obtain:

$$P(A_2|B) = \frac{0.9 \times 0.55}{0.565} = 0.876$$

Example

How do we interpret this?

The probability that the client invested in β , given that he had a return above the market, is 0.876.

Example

All these computations can be very easy with the help of the following table:

A_i	$P(A_i)$	$P(B A_i)$	$P(A_i)P(B A_i)$	$P(A_i B)$
A_1	0.1	0.7	0.07	0.124
A_2	0.9	0.55	0.495	0.876
	1		0.565	1

A test that is right 99% of the time

A screening test for a rare condition. Move the sliders and read off the only number a patient cares about:
given a positive test, what is the chance of actually being ill?



Prevalence **1.0%**



Sensitivity **99.0%** $P(+ | \text{ill})$



Specificity **99.0%** $P(- | \text{healthy})$



$P(\text{ill} | \text{positive}) = 50.0\%$

In 100 000 people:

- truly ill and positive: 990
- healthy but positive: 990
- ill but missed: 10

What the chart showed

At 99% sensitivity **and** 99% specificity, a positive result on a condition affecting 1 in 100 people leaves you only about **50%** likely to be ill. Push prevalence down to 1 in 1000 and the answer collapses to roughly **9%**.

The reason is in the two blocks of the bar: the healthy group is so much larger that its **1% error rate still produces more positives** than the ill group produces in total.

$P(\text{ill} \mid +)$ is not $P(+ \mid \text{ill})$. Swapping them is the single most common probability mistake outside this classroom, and Bayes is what keeps them apart.

Example

Let's verify now if the event A_1 and B^c are independent or not!

According to the definition of independence: $P(A_1 \cap B^c) = P(A_1)P(B^c)$

- $P(A_1 \cap B^c) = P(A_1) \times P(B^c|A_1) = 0.1 \times 0.3 = 0.03$
- $P(A_1) \times P(B^c) = 0.1 \times (1 - 0.565) = 0.0435$
- Then:

$$P(A_1 \cap B^c) = P(A_1) \times P(B^c) \Leftrightarrow 0.03 \neq 0.0435$$

Then A_1 and B^c are not independent.

? Axioms, Independence and Bayes ·

Question 1

A and B are incompatible, with $P(A) > 0$ and $P(B) > 0$. Then they are:

A. independent

B. independent only if disjoint

C. never independent

D. impossible to classify

✓ **C.** Incompatible means $P(A \cap B) = 0$, but independence needs $P(A \cap B) = P(A)P(B) > 0$. The two cannot hold together.

? Axioms, Independence and Bayes ·

Question 2

$P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cap B) = 0.2$. Then A and B are:

- A. incompatible
- B. neither
- C. independent
- D. both

✓ **C.** $P(A)P(B) = 0.5 \times 0.4 = 0.2 = P(A \cap B)$ ✓



Axioms, Independence and Bayes ·

Question 3

A disease affects 2% of the population. A test is positive for 95% of ill people, and also for 5% of healthy people.

A randomly chosen person tests positive. What is the probability that this person is ill?

Axioms, Independence and Bayes · Solution

Let D be “ill” and $+$ be “tests positive”. By the Law of Total Probability:

$$P(+) = 0.95 \times 0.02 + 0.05 \times 0.98 = 0.019 + 0.049 = 0.068$$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.019}{0.068} \approx 0.279$$

Only about **28%**, even though the test is right 95% of the time. The disease is rare, so most positives come from the large healthy group.

Bibliography

- Murteira, B.; Ribeiro C.; Silva, J. and Pimenta, C. (2010) Introdução à Estatística (2ª Edição). McGraw-hill.
- Paulino C.D.; Branco J.A. (2005). Exercícios de Probabilidade e Estatística. Escolar Editora.
- Pedrosa, A.; Gama, S. (2004). Introdução Computacional à Probabilidade e Estatística. Porto Editora.