PROBABILITY THEORY

TOC

# What is the Probability Theory.

It is the mathematical tool we use to measure the **likelihood** of the **different results** of an **experiment**. For example, if we want to measure the probability of getting each number when throwing a die, we would use the **probability theory**.

For that we denote Ω to be the set of possible outcomes. We can represent this set as a box that will contain all the possible outcomes, for that, the possibility of Ω will be 1. (note that the probabilities are between 0 and 1 for then 1 is the same as 100%)

This set can be divided into different subsets, also called events, each one of them with a **probability measure** between **0 and 1**. In the case of having a subset with probability of 1, this will be the same as Ω.

Subsets can also be combined in the following ways:

* A U B (Union): Either A or B or both happen.
* A ∩ B (Intersection): Both A and B happen.
* A \ B (Difference): A happens but B does not
* A’ or Aᶜ (Complement): The opposite of A. That is the possibility of A not happening.

Properties of events in Ω:

1. P(A) belongs to [0,1]
2. P(Ω) equals 1 and P(Ø[[1]](#footnote-1)) equals 0.
3. If A ⃀ B, then P(A) ⃀ P(B)[[2]](#footnote-2)
4. If A ∩ B = Ø 1, the P(A U B) = P(A) + P(B) [[3]](#footnote-3)
5. P(A\B) = P(A) – P(A ∩ B)

Image of the probability Space


Image 1. Ω Is the sample space and Ø anything outside it

A grey and orange circle with black letters

Description automatically generated

Image 2. The circles containing each letter denotes the probability of either A or B happening.

For that, every time A happens, B has to happen since A ⃀ in B.

A grey box with orange circles and black text

Description automatically generatedA diagram of a venn diagram

Description automatically generated

Image 3 (Left) and Image 4 (Right). In image A we can see that they are not intersecting, thus P(A) U P(B) = P(A) + P(B).

In the other hand, for image 4 we cannot do so since the red region would be added twice.

For that we use P(A) U P(B) = P(A) + P(B) – P(A∩B).

# 2. How to compute the probability of am event ‘A’:

There are 3 main interpretations:

1. Classical:

Where we assume that all the results are equally likely. This makes sense in chance games but falls apart on many other contexts. The formula will be:

Examples of this could be the probability that for throwing a die and having a number greater than 3. That would be:

1. Frequentist:

This can be applied to data sets where we have the statistics, but the experiment cannot be repeated. The formula will be:

1. Subjective:

P(A) is a measure of the evidence supporting A.

This is used for example in crime cases where the possibility of happening by accident is really low, and even though it may have happened, as it is so improbable, would be sentenced as guilty.

Despite of the existence of the Frequentist and Subjective methods, we will focus on the Classical one since for our concern is the most important.

# 3. Conditional Probability:

The probability of B **conditional** on A is the probability of **B happening if A has happened.** Denoted as **P(A|B)**, it will be computed such:

And it will be only applied if P(A) ≠ 0.

One example of this would be, when throwing a die and getting an even number, what is the probability of it being greater than 3. That will be in mathematical form:

Note that this differs from the intersection in the way that the die was thrown, and once thrown we were told: “The number is even, what is the chance of getting a value greater than 3”. One thing occurred before the other and so the result will be biased. Is different from if they tell you: “What is the probability of throwing a die and getting a number even and greater than 3”, this would be the intersection. The difference between both of these cases is subtle but changes the whole thing.

It is also different P(A|B) and P(B|A).

# 4. Independence

We say that A and B are **independent** when knowing that A has happened does not change the probability of B happening. Mathematically that is:

This makes total sense since as B does not depend of A, the probability of B occurring when A did occur, will not be different from the probability of it occurring if A did not occur.

To check whether the condition of independence is met, we must check that the following equality is true:

You may be wondering how we can get any kind of intersection if they are proclaimed **independent** events from each other. For this reason and in this case, we will define the intersection as the probability of getting both values to be true at the same time.

Imagine that you have a die. What is the probability of throwing it and getting exactly the number 3? What is the probability of throwing a die and getting exactly the number 4? For both of these cases the probability will be 1/6 and for both cases happening in a row, the chance will be 1/36. It is typical to think that this falls for the gambler’s fallacy, nevertheless that is a wrong assumption since the second probability it is not affected by the first one in the sense that for both probabilities we get 1/6.

If you think about it, if you write the full sequence of values that will result after throwing a die before it has been thrown, even though you have 1/6 of probabilities to guess each one of them independently, guessing all of them will be significantly less likely.

Properties:

* If A and B are independent, so will be Ā and B as well as A and not(B) and Ā and not(B)[[4]](#footnote-4)
* This could be generalized for a list of events in such way that for the events A1, A2, … An, are independent if and only if:

1. Ø is defined to be the empty set. That is the probability of Ø is the same as the probability Ω not happening for that, is 0. [↑](#footnote-ref-1)
2. The symbol ⃀ denotes that A is contained in B. For that reason if A happens, B MUST also happen since A is inside the probability B. Check image 2. [↑](#footnote-ref-2)
3. If there is no intersection between A and B, they are both different events, so since they have not repeated values their probabilities can be added. Check images 3 and 4. [↑](#footnote-ref-3)
4. The factor of form not(B) will be used to denote the negative of B, being that P(not(B)) = 1 - P(B) and will be used since Word does not include the proper symbol out of the equation menu. [↑](#footnote-ref-4)