# Notebook on probability and statistics

Joaquín Gómez

# Chapter 1

# Counting and series

We will start this notebook by introducing basic probability concepts. After that, we will use combinatorics to describe certain experiments. The subject of combinatorics is essential to calculate probabilities, and it's useful in many areas of science. The topics that will be covered are basics and a *must know* in probability theory and statistics.

Combinatorics requires a basic knowledge of *set theory*. This document is not focused on set theory, so we will neglect it. However, it is expected from the reader to know at least some basic notation and to have a notion of set theory.

## 1.1 Introduction

The next example will be useful to introduce the notion of "frequentist probabilities"*. These rely on the idea that, given $n$ results of an experiment $E$, if the desired result $R$ is present $m$ times after the experiment, then the probability of $R$ is given by

$$p_R = \frac{m}{n} \tag{1.1}$$

♦ **Example 1.1.** *If we toss a coin, what is the probability of it getting the first heads in the* $n^{th}$ *trial? This is, the probabilty of the result set to be* $R = \{TTT\ldots TH\}$ *(with* $n-1$ $T's$*)*

Given that the coin has 2 sides and assuming a fair coin, we must set the probability of getting heads as $p_H = \frac{1}{2}$. The probability of tails, $p_T$ is the same.

The probability of the coin to land heads after $n$ throws will be given by

$$\prod_{k=1}^{n-1} p_T \cdot p_H = p_T^{n-1} p_H \tag{1.2}$$

*In classical probability theory, probabilities are assumed to exist as axioms; frequentist definitions aim to approximate them empirically.

Why? You could ask. Well, think of this: you throw the coin once, you had a 1/2 chance to get tails. Now, throwing it again and getting tails *twice* in the same *experiment* has 1/2 chance too, but given it's the *same* experiment, you had 1/4 chance of success, because to get there you had to succeed the first time. This is a chain of probabilities that can be drawn as a tree, with each result as a node and connected by a certain *weight* given by the probability of that result to happen.

This is the first kind of "probability distribution" that we will see in this text. One could infer the value of $p_T$ because, given that we have two possible results

$$p_H + p_T = 1 \tag{1.3}$$

will always be true, because we either have tails or heads after tossing the coin, nothing else. Then

$$p_T = 1 - p_H \tag{1.4}$$

so we can rewrite (1.2) as

$$\textit{Geometric PMF} = (1-p)^{n-1}p \tag{1.5}$$

Alright! Now we have introduced a new concept, a PMF. See, its derivation was very simple. However it is important to know what a PMF is. This will be covered in the **Discrete Random Variables** chapter.

## 1.2 Counting – Combinatorics

We use counting in probabilities to account for the number of possible *orderings* of certain elements, for example.

⧫ **Example 1.2.** *There are* 12 *people in a party. If we want to make groups of* 3*, such that we have* 4 *groups. How many ways of making these groups we have?*

This problem ask for the total subsets of 3 people in a set of 12 people. The solution is given by the Binomial Coefficient

$$\binom{12}{3} = \frac{12!}{3!(12-3)!} = \frac{12!}{4!9!} = 220 \tag{1.6}$$

⧫ **Example 1.3.** *(Deriving the Binomial Coefficient) To derive the formula we used in (1.6) we need to first know how to count the total combinations of a set of $n$ elements.*

*For the first element, we have $n$ options to arrange it, for the second element we have $n-1$ options, as the first one is already sorted. For the element in position $n-1$ we just have 2 options, and for the last element, we have already determined its position, so it's 1 option only. If we want to count the total ways* *to arrange them, we must multiply the total ways of arranging the first element, by the total ways of arranging the second element... and so on.*

$$\textit{total arrangements} = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n! \tag{1.7}$$

*Now, this formula is meaningful when these $n$ objects are distinct from each other. Now, imagine we want to count the number of permutations of $r$ items from these $n$ items.*

*We have $n$ ways to choose the first, $n-1$ ways for the second, and in general we have $n-k+1$ ways to choose the $k^{th}$ item. The total ways of arranging the $k^{th}$ item is given by*

$$\prod_{j=1}^{k} (n-j+1) \tag{1.8}$$

*This simplifies as*

$$\frac{\prod_{j=1}^{n} (n-j+1)}{\prod_{j=k+1}^{n} (n-j+1)} = \frac{n \cdot (n-1) \cdots 2 \cdot 1}{(n-k) \cdot (n-k-1) \cdots 2 \cdots 1} = \frac{n!}{(n-k)!} \tag{1.9}$$

*Now, this permutation takes in account that we can sort these $k$ objects in $k!$ ways. Thus, if we want to know how to choose the objects instead of choosing AND arranging them, we have to divide by $k!$. Thus, we found the Binomial Coefficient*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{1.10}$$

We have covered a bit of combinatorics. Further topics will be disregarded. However, it is important to at least have a practical notion of combinations and counting.

## 1.3 Sets and probability

Sets and probabilities are strongly related. When we measure probabilities, we measure them on sets of particular objects. For simplicity, these can be numbers, vectors, or any object in a linear space (vector space).

In order to understand probability theory, one must know set theory. The basics at least.

### 1.3.1 Fundamental aspects of set theory

We will define the fundamental operations for sets, given two sets $A$ and $B$ we have 1em

- ▷ $A$ equals $B$: $A = B = \forall x (x \in A \text{ and } x \in B)$
- ▷ Union of $A$ and $B$: $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- ▷ Intersection of $A$ and $B$: $A \cap B = \{x : x \in A \text{ and } x \in B\}$

▷ Difference of $A$ and $B$: $A \backslash B = A - B = \{x : x \in A \text{ and } x \notin B\}$[†]

We introduce a new set $S$. This set contains both $A$ and $B$, this is, for every element $x$ in $A$ or $B$, $x \in S$. But, there exists $y \in S$ such that $y \notin A \cup B$. We say that $A \subset S$ and $B \subset S$. The symbol '$\subset$' is called *subset of.* We can also say that $A$ is a subset or is equal to $S$ if $A = S$ or $A \subset S$; we denote this as $A \subseteq S$. We can invert the symbol and say $S \supset A$, or $S \supseteq A$.

♦ **Definition 1.1.** *Let $A$ and $B$ satisfy $A \cap B \neq \emptyset$[‡] and define $S$ to satisfy $A, B \subset S$. Then we have*

- ▷ $A^c = S \backslash A = \{x : x \notin A \text{ and } x \in S\}$ *(Complement of $A$ in $S$ or absolute complement of $A$)*
- ▷ $A \Delta B = (A \backslash B) \cup (B \backslash A) = \{x : x \in A \text{ or } x \in B \text{ and } x \notin A \cap B\}$ *(Symmetric difference between $A$ and $B$)*

We will check for these formulas, the first one is trivial [§].

For the second one, note that if $x \in (A \backslash B)$, then $x \notin B$. Also, if $x \in (B \backslash A)$, $x \notin A$. Now, let $x \in (A \backslash B)$ and $x \in (B \backslash A)$, suppose that $x \in A$, so $x \notin B$; but, if $x \in B$, then $x \notin A$. However, let $x \in A \cap B$ so that $x \in A$ and $x \in B$, consequently, $x \notin A$ and $x \notin B$, which is a contradiction. So $x \notin A \cap B$. Finally, suppose $x \in S \cap (A \cup B)^c$, this also implies $x \notin A$ and $x \notin B$, so $x \notin S \cap (A \cup B)^c$. We finally arrive at the conclusion that either $x \in A$ or $x \in B$, but $x \notin A \cap B$.

### 1.3.2 The triad of probability theory

In order to introduce ourselves to probability theory formally, let's present what is called a *probability space.*

♦ **Definition 1.2.** *(Probability Space)*

---

[†] We will use the notation $A \backslash B$

[‡] The set '$\emptyset$' is called *empty set* and is defined as $\emptyset = \{\}$.

[§] the definition of complement needs a frame of reference, otherwise if $A \not\subseteq S$ we would have $A^c = \emptyset$, which means that every element that exists in our space is in $A$. For example, if we *restrict* numbers to only be real numbers, then our frame of reference will be $\mathbb{R}$, thus, $\mathbb{R}^c = \emptyset$ and $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$