*It is through the calculation of probabilities*

*that the divine order becomes visible.*

— Jacob Bernoulli

**What is probability?**

The study of probability, though having very ancient roots, began its modern development in the seventeenth century through the famous correspondence between Blaise Pascal and Pierre de Fermat. Their discussion on games of chance, and in particular the *"problem of the division of stakes''* laid the groundwork for the systematic analysis of uncertain events. Years later, Jacob Bernoulli’s *"Ars Conjectandi"* established the first classical definition of probability, providing the study of random events with mathematical clarity. Refinements by De Moivre and Laplace transformed it into a powerful analytical theory, while its true axiomatic structure only crystallised in the twentieth century with Kolmogorov’s *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* in 1933 [...].

At its heart, probability is nothing more—and nothing less—than a branch of mathematics developed to describe random events, also referred to as *stochastic*. Indeed, the word stochastic comes from the Greek word στοχαστικός, which literally means “to guess” or “to aim”. The way we describe such events, characterized by the uncertainty of their outcome, is by defining a quantity we will call \(mathbb{P}\), of probability. That quantity \(mathbb{P}\) will denote a number between \(0\) and \(1\), which reflects the degree of uncertainty, or *surprise*, with which the random event produces a specific outcome. For an event \(A\), such as observing a heads when tossing a coin, or a given face when rolling dice, the numerical convention is written as follows,

* If I am sure \(A\) will never occur, \(mathbb{P}(A)=0\).
* If I am sure \(A\) will always occur, \(mathbb{P}(A)=1\).
* For anything in between, if \(A\) is *uncertain*, then \(mathbb{P}(A) \in (0,1)\).

where the \(\in\) symbol just means "belongs to". Thus, probability measures the whole span between impossibility and absolute certainty.

Consider the classic example of tossing a coin. Let $H$ denote heads and $T$ denote tails. If a coin is symmetric and fair, we would name the number of possible outcomes, or \textit{sample space} $\Omega = \{H,T\}$. Those with less mathematical training may find this notation rather odd. Plain and simple, ideas such as sample space or measure space come from the underlying mathematical theory that was used to build modern probability. For the purpose of this course \textit{sample space} will essentially mean \textit{set of possible outcomes}.

If we are certain  we will get heads heads, then $\mathbb{P}(H)=1$ and $\mathbb{P}(T)=0$; On the other hand, if we are certain we will get tails, the roles reverse, and then $\mathbb{P}(H) = 0$ and $\mathbb{P}(T) = 1$. In the general case, where both outcomes can happen with equal probability, we would write

\begin{equation}

\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2} \; . \nonumber

\label{eq:prob\_coin}

\end{equation}

This assignment of $1/2$ probability is not an arbitrary choice. It reflects both a symmetry of the physical system and an idealisation of experimental repetition. Indeed, the way we define probabilities for a given event $A$ is just by computing the ratio of how many times we get that event $n(A)$, and the total number of trials $N$.

\begin{equation}

    \mathbb{P}(A) = \lim\_{N\to\infty} \frac{n(A)}{N} \; .

    \label{eq:prob\_frequentist}

\end{equation}

This is called the \textit{frequentist} definition of probability, since it relies on the frequency with which each results occur. Tossing a fair coin many times, the observed frequencies of heads and tails converge towards the probabilistic assignment $1/2$. The frequentist definition is built upon the idea of repetition and reproducibility. We \textit{expect} that, if we repeat the toss many times, the number of times we get $H$ and $T$ will approach to a perfect half, as $N$ increases.

\begin{equation}

\mathbb{P}(H) = \mathbb{P}(T) \simeq \frac{1}{2} \; . \nonumber

\label{eq:prob\_coin}

\end{equation}

This convergence principle is formalised in Bernoulli’s Law of Large Numbers and later generalised in the Central Limit Theorem, as we will discuss in next chapter. Further approaches to the definition of probability, such as the \textit{bayesian}, will be discussed in Chapter 5.

\medskip

Beyond frequencies, probability must also obey the principle of \textit{unitarity} or \textit{normalisation}. This is the mathematical formalization of quite a natural intuition: at least one of the possible events must take place. By imposing that the sum of the probabilities of all mutually exclusive outcomes must equal 1, we ensure we have assigned the numerical values in a consistent way. For a finite experiment with outcomes $\{x\_1,\ldots, x\_n\}$, the unitarity property is written as

\begin{equation}

    \sum\_{i=1}^{n} \mathbb{P}(x\_i) = 1 \; .

    \label{eq:unitarity}

\end{equation}

For a coin toss, this reduces to

\begin{equation}

    \mathbb{P}(\text{H}) + \mathbb{P}(\text{T}) = \frac{1}{2} + \frac{1}{2} = 1 \; . \nonumber

\end{equation}

For a dice roll, where each face appears with a probability $P = 1/6$, we would write

\begin{equation}

    \mathbb{P}(\text{1}) + \mathbb{P}(\text{2}) + \dots + P(6) = \frac{1}{6} + \frac{1}{6} + \dots + \frac{1}{6} = 1 \; . \nonumber

\end{equation}

Unitarity is one of the most fundamental properties of probability, and it will prove useful to make calculations further on this very chapter. In addition, just as a note on the formal development of all this framework, it is only once this normalisation condition imposed, that the \textit{probability} space $(\Omega,\mathcal{F},\mathbb{P})$ can be defined from the abstract notion of \textit{measure} space $(\Omega,\mathcal{F},\mu)$.

\medskip

These three notions suffice for now. We have seen probability as a number that quantifies uncertainty, the frequentist definition in terms of ratio, and the idea of unitarity. Let us emphasize, though, that this formulation is actually quite recent. Even though the basic intuitions were already introduce by Bernoulli and Laplace, as we discussed, it was not until the nineteenth and twentieth centuries, that probability theory was properly formalized in the language of analysis. The idea of expectation became formally defined via the Lebesgue integral \cite{lebesgue1902}, stochastic processes were studied by Wiener and Doob \cite{doob1953}, and the idea of convergence - that lied the foundations for unitarity - were explored by Borel, Cantelli, and Kolmogorov \cite{billingsley1995}. From gambling practice, probability grew into a highly abstract and powerful theory.

\medskip

The way probability was mathematically defined is based on the idea of \textit{probability space}. This may seem quite abstract at first, so let's illustrate with an example. Imagine that every experiment we perform - tossing a coin, rolling a dice, or measuring the brightness of a star - has a collection of possible outcomes. This collection is what mathematicians call the \textit{sample space}, often written as $\Omega$. Within this space, we may be interested in particular groups of outcomes, such as ``getting an even number'' from a dice roll or ``obtaining heads'' from a coin toss. These groups of outcomes are what we call \textit{events}.

\medskip

Formally, a probability space is defined as a collection of three mathematical objects $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega$ is the \emph{sample space} of possible outcomes, $\mathcal{F}$ is a collection of measurable events (for mathematicians, a $\sigma$-algebra), and $\mathbb{P}$ is a measure assigning real numbers between $0$ and $1$. It is simply a structured way of saying (1) the set of all possible outcomes of an experiment, (2) the events we are interested in within that set, and (3) a systematic assignment of likelihoods to those events. The more abstract terminology - $\sigma$-algebras and measures - becomes indispensable when probability theory is extended to complicated or infinite cases, but in everyday examples such as coins, dice, or cards, it suffices to remember these three key ingredients: outcomes, events, and probabilities.

\medskip The way we define and assign probabilities to random events is done in accordance with Kolmogorov’s axioms, which we summarize as follows:

\begin{itemize}

  \item \textbf{Non-negativity:} The probability of any event is never negative. Probabilities are numbers that represent likelihood, so they must satisfy $\mathbb{P}(A) \geq 0$. In simple terms, it makes no sense to say an event happens with ``negative chance.''

\begin{equation}

  \mathbb{P}(A) \;\geq\; 0 \; , \qquad \forall A \in \mathcal{F} \; .

\end{equation}

  \item \textbf{Normalisation:} The probability of the whole sample space $\Omega$ is exactly $1$. This expresses the fact that ``something will happen.'' If $\Omega$ is the complete list of possible outcomes of an experiment, then we are certain that the final outcome will be one of them.

\begin{equation}

  \mathbb{P}(\Omega) \;=\; 1 \; .

\end{equation}

  \item \textbf{Countable additivity:} If we have several events $A\_1, A\_2, \ldots$ that cannot overlap (e.g.\ they are mutually exclusive or disjoint), then the probability that one or another occurs is the sum of their probabilities. For instance, in rolling a die, the probability of rolling ``1 or 2'' is $P(1)+P(2)$ because the two outcomes cannot happen at the same time.

\begin{equation}

  \mathbb{P}\!\left(\bigcup\_{i=1}^\infty A\_i\right) \;=\; \sum\_{i=1}^\infty \mathbb{P}(A\_i) \; ,

\qquad \text{for disjoint } A\_i \in \mathcal{F}

\end{equation}

\end{itemize}

where the symbol $\bigcup$ just means the \textit{reunion} of all events $A\_i$. Together, these three axioms form the rigorous foundation of probability theory and ensure consistency in reasoning about uncertainty.

\medskip

This framework allows one to treat uncertainty with mathematical precision, applicable not only to games of chance but also to infinite-dimensional spaces, stochastic processes, and mathematical modelling of various and broad scenarios within the natural sciences. It is this generality that transformed probability theory from a tool of gamblers and actuaries into one of the central languages of modern science.

\medskip

Of course, mathematics provides structure, but actual meaning requires interpretation. As we mentioned already, two main schools of interpretation emerged from this axiomatic development of probability theory. The \textit{frequentist} definition, associated with Richard von Mises \cite{vonmises1928}, identifies probability as we did in \eqref{prob\_frequentist}, with long-run frequency in repeated trials: it is then an objective property of the physical world, revealed through repetition. On the other hand, the \textit{Bayesian} tradition, with origins in Thomas Bayes’ posthumous essay \textit{Towards solving a Problem in the Doctrine of Chances} back in 1763 \cite{bayes1763}, conceives probability as a measure of rational belief, updated by evidence through what we call nowadays Bayes’ rule, or Bayes' theorem. Such idea was later refined by Laplace \cite{laplace1812} and further developed by Bruno de Finetti, who further emphasized that probability expresses degrees of personal belief coherent under rational rules of betting \cite{definetti1974}.

\medskip

These two traditions, frequentist and Bayesian, illuminate different facets of the same mathematical object. One interprets probability as an empirical limit of frequencies, the other as a calculus of information and belief. Both, however, are grounded in the modern axiomatic formulation: probability is a measure, and measures assign form and consistency to uncertainty.