These are lecture notes for an introductory course on probability taught as part of the first year's program at École des Ponts ParisTech. This course is currently under the direction of Aurélien Alfonsi (CERMICS) and based on the following textbook:

B. Jourdain, Probabilités et statistique, Ellipses 2009, 2nd edition 2016.

Chapter 1

Probability measures on finite spaces

1.1 Introduction

1.1.1 Definitions

Let $n \in \mathbb{N}^*$. We consider a *random experiment* which consists in the realisation of a unique outcome ω among n possibilities $\omega_1, \cdots, \omega_n$. The set of all possible outcomes of this experiment is called the *sample space* or *universe* and is usually denoted by $\Omega = \{\omega_1, \cdots, \omega_n\}$. Any element $\omega \in \Omega$ is called a *realisation*.

In this chapter, we will only focus on finite sample spaces, that is random experiments that have only finitely many possible outcomes. Most of the definitions and vocabulary introduced here will be extended in the next chapters.

Example 1.1.1. The typical sample space is:

- If you toss a coin, $\Omega = \{\text{head}, \text{tail}\}$;
- If you throw a die, $\Omega = \{1, 2, 3, 4, 5, 6\}$;
- If you throw two dice, $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$.

Let $N \in \mathbb{N}^*$. When the random experiment is repeated N times, for $1 \leq k \leq n$, the *frequency* of ω_k is defined by

$$F_N(k) = \frac{\text{number of realisations of } \omega_k}{N}.$$

When $N \to +\infty$, $F_N(k)$ varies less and less and its limit corresponds to the intuitive notion of probability of ω_k, namely a measure of the likelihood of the occurence of the realisation ω_k, expressed as a real number between 0 and 1.

We call *event* any subset $A \subset \Omega$. We say that A is realised if the outcome ω of the random experiment is in A. The *frequency* of A is defined by

$$F_N(A) = \frac{\text{number of realisations of } A}{N} = \sum_{\substack{1 \le k \le n \\ \omega_k \in A}} F_N(k). \tag{1.1}$$

Similarly, $F_N(A)$ is expected to converge when $N \to +\infty$ to the intuitive notion of probability of A, that is a measure of the chance that A is realised. Notice that $F_N(\Omega) = 1$.

Definition 1.1.2 (Probability measure)**.** *A probability measure on a finite sample space* $\Omega = \{\omega_1, \cdots, \omega_n\}$ *is a vector* $(p_1, \cdots, p_n)$ *which satisfies*

(i) For all $1 \le k \le n$, $p_k \ge 0$ *;*

(ii) $\displaystyle\sum_{k=1}^{n} p_k = 1$.

To any event $A \subset \Omega$ *we associate the real number*

$$\mathbb{P}(A) = \sum_{\substack{1 \le k \le n \\ \omega_k \in A}} p_k, \tag{1.2}$$

called the probability of A*. In particular, for* $1 \le k \le n$, $\mathbb{P}(\{\omega_k\}) = p_k$.

Remark 1.1.3.

- (1.1) is the statistical analogue of (1.2) ;
- The term “probability measure” can either refer to the vector $(p_1, \cdots, p_n)$ or the map $A \subset \Omega \mapsto \mathbb{P}(A)$;
- (i) and (ii) imply $p_k \in [0, 1]$ for all $1 \le k \le n$.

Example 1.1.4. Suppose you toss a coin: $\Omega = \{\text{head}, \text{tail}\}$. Set $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{\text{head}\}) = \mathbb{P}(\{\text{tail}\}) = \frac{1}{2}$ and $\mathbb{P}(\Omega) = 1$. Then $\mathbb{P}$ is a probability measure which models a fair coin.

Terminology Let Ω be a sample state, $A, B \subset \Omega$ be two events and $\mathbb{P}$ be a probability measure on Ω.

- The event realised when A is not is called *complementary event* of A and denoted $A^{\complement} = \Omega \backslash A$;
- The event realised when A and B are realised is called A *and* B and denoted $A \cap B$;
- The event realised when A or B is realised is called A *or* B and denoted $A \cup B$;
- A is *negligible* is $\mathbb{P}(A) = 0$;
- A is *almost sure* if $\mathbb{P}(A) = 1$;

– The *indicator function* of the event A, denoted $\mathbb{1}_A$, is the function equal to 1 iff A is realised, that is

$$\mathbb{1}_A \ : \ \begin{array}{ccl} \Omega & \to & \{0,1\} \\ \omega & \mapsto & \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{else} \end{cases} \end{array} .$$

Exercise 1.1.5. Let Ω be a sample space, $A, B \subset \Omega$ be two events and $\mathbb{P}$ be a probability measure on Ω. Show the following assertions:

1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
2. $\mathbb{1}_{A \cap B} = \mathbb{1}_A \times \mathbb{1}_B$.
3. $\mathbb{1}_{A^\complement} = 1 - \mathbb{1}_A$.
4. $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B}$.

1.1.2 Uniform distribution

In many situations, all the possible outcomes of a random experiment play similar or symmetric roles and are expected to be equal in likelihood. For instance, the outcomes of the toss of a fair coin or the throw of a fair die are equally likely to happen. In order to model this kind of situation, we assign an equal weight to each outcome.

Definition 1.1.6 (Uniform distribution)**.** *The* uniform distribution *on a sample space* $\Omega = \{\omega_1, \cdots, \omega_n\}$ *is the probability measure* $\mathbb{P}$ *defined for all* $A \subset \Omega$ *by*

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

where $|A|$ *denotes the cardinality of* A.

Remark 1.1.7. The uniform distribution is the only probability measure $\mathbb{P}$ which satisfies $\mathbb{P}(\{\omega_k\}) = \frac{1}{n}$ for all $1 \leq k \leq n$.

In the setting of the uniform distribution, the computation of a probability of an event amounts to the computation of the cardinality of that event. It is therefore useful to recall some basics of combinatorics.

Proposition 1.1.8. *Let* $n, k \in \mathbb{N}^*$ *be such that* $k \leq n$. *Then*

(i) The number of permutations of a set containing n *elements is* $n!$ *;*

(ii) The number of one-to-one functions from a set containing k *elements to a set containing* n *elements is* $\frac{n!}{(n-k)!}$ *;*

(iii) The number of subsets containing k *elements of a set containing* n *elements is denoted* $\binom{n}{k}$ *(read as "n choose k") and equal to* $\frac{n!}{(n-k)!k!}$.

Proof. Let $A = \{a_1, \cdots, a_k\}$ be a set of k elements, $B = \{b_1, \cdots, b_n\}$ be a set of n elements and $f : A \to B$ be a one-to-one function. There are n possibilities for the value of $f(a_1)$. Since f is one-to-one, $f(a_2)$ belongs to $B \backslash \{f(a_1)\}$, hence $n-1$ remaining possibilities for the value $f(a_2)$. By induction, for all $1 \leq i \leq k$, given $f(a_1), \cdots, f(a_{i-1})$, there are $n-i+1$ possibilites for the value of $f(a_i)$. This adds up to

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

possibilities for f, which proves (ii). The assertion (ii) implies (i) for $k = n$.

Let us now prove (iii). The choice of a one-to-one function from a subset of k elements of B to B is equivalent to the choice of k distinct elements of B in a certain order. Moreover, there are $k!$ ways to order k distinct elements of B. Therefore, there are $k!$ times more one-to-one functions from a subset of k elements of B to B than subsets containing k elements of B. By (ii) and by definition of $\binom{n}{k}$, this means

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

which proves (iii). $\square$

1.2 Conditional probability and independence

1.2.1 Conditional probability

Let A and B be two events of a sample space Ω. The probability of A, $\mathbb{P}(A)$, models the likelihood that A occurs. In other words, an observer would expect A to occur with chance $\mathbb{P}(A)$. Suppose now that the observer knows that B occurs. Then his expectation of the likelihood of A has no reason to remain the same. For instance, we would expect more that it rains if we know it is cloudy than if we know nothing about the weather. The concept of conditional probability allows the model to take into account the information we might have and revise the likelihood of an event accordingly.

Definition 1.2.1 (Conditional probability)**.** *Let Ω be a sample space, $A, B \subset \Omega$ be two events and $\mathbb{P}$ be a probability measure on Ω. Suppose that $\mathbb{P}(B) > 0$. The* conditional probability of A given B *is denoted $\mathbb{P}(A|B)$ and defined by*

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

The conditional probability $\mathbb{P}(A|B)$ models the likelihood that A occurs, knowing that B occurs. The map $A \subset \Omega \to \mathbb{P}(A|B)$ can be seen as the update of $\mathbb{P}$ given the information that B occurs. In the same spirit, the next statement is a useful result known as Bayes' theorem.

Theorem 1.2.2 (Bayes' theorem)**.** *Let $n \in \mathbb{N}^*$, Ω be a sample state, $\mathbb{P}$ be a probability measure on Ω, $B \subset \Omega$ be an event and $(A_k)_{1\leq k\leq n}$ be a partition of Ω, that is such that Ω is the disjoint union of $(A_k)_{1\leq k\leq n}$. Suppose that $\mathbb{P}(B) > 0$ and for $1 \leq k \leq n$, $\mathbb{P}(A_k) > 0$. Then*

$$\forall 1 \leq k \leq n, \quad \mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}. \tag{1.3}$$

Proof. Let $k \in \{1, \cdots, n\}$. By definition of the conditional probability, $\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k\cap B)}{\mathbb{P}(B)}$. On the one hand, $\mathbb{P}(A_k \cap B) = \mathbb{P}(B|A_k)\mathbb{P}(A_k)$. On the other hand, since $(A_i)_{1\leq i\leq n}$ is a partition of Ω, $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i \cap B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$, which proves (1.3). □

1.2.2 Independence

We saw in the previous section how the knowledge that an event B occurs can affect the likelihood that another event A occurs. However, the knowledge of B might have no effect on the likelihood of A, in which case $\mathbb{P}(A|B) = \mathbb{P}(A)$. For instance, I would not expect the likelihood that it rains to change if I know that my neighbour wears red shoes. This special case corresponds to the concept of *independence*.

Definition 1.2.3 (Independence)**.** *Let Ω be a sample space. Two events $A, B \subset \Omega$ are said to be* independent *if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Remark 1.2.4. If $\mathbb{P}(B) > 0$, then A is independent of B iff $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Definition 1.2.5. *Let $n \in \mathbb{N}^*$ and Ω be a sample space. A family $(A_k)_{1\leq k\leq n}$ of events is said to be* mutually independent *if*

$$\forall k \in \{1, \cdots, n\}, \quad \forall 1 \leq i_1 \leq \cdots \leq i_k \leq n, \quad \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \cdots \times \mathbb{P}(A_{i_n}).$$

Remark 1.2.6.

- If $(A_k)_{1\leq k\leq n}$ is mutually independent, then it is pairwise independent, that is for all $1 \leq i \neq j \leq n$, A_i is independent of A_j, but the converse is not true ;
- If $(A_k)_{1\leq k\leq n}$ is mutually independent, then $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \times \cdots \times \mathbb{P}(A_n)$, but the converse is not true.