

# Proba Stats

Dario Marcone

Lundi 15 septembre 2025

# Table des matières

<b>1</b>	<b>Probabilities : starter guide</b>	<b>2</b>
1.1	Probabilities spaces and measures . . . . .	2
1.1.1	Probability spaces . . . . .	2
1.1.2	Discrete probability measures . . . . .	3
1.1.3	Continuous probability measures . . . . .	3
1.1.4	Genreal probability spaces and measures . . . . .	4
1.1.5	Inclusion-Exclusion . . . . .	5
1.2	Random variables and expectation . . . . .	5
1.2.1	Random variables . . . . .	5
1.2.2	Expectation . . . . .	8
1.2.3	Transfer Theorem . . . . .	9
1.2.4	Random vectors . . . . .	9
1.2.5	Change of variable formula . . . . .	10
1.2.6	Moments, and Moment Generating Function . . . . .	11
1.3	Conditional probability independence . . . . .	12
1.3.1	Conditional probability . . . . .	12
1.3.2	Independence . . . . .	12
1.3.3	Bayes law, formula of total probability . . . . .	13
1.3.4	Almost sure properties . . . . .	14
1.4	Correlation . . . . .	14
1.4.1	Variance, Covariance . . . . .	14
1.4.2	Pearson correlation coefficient . . . . .	15

# Chapitre 1

## Probabilities : starter guide

### 1.1 Probabilities spaces and measures

#### 1.1.1 Probability spaces

##### Définition 1

A **probability space** is a set of realisations denoted  $\Omega$ , together with a probability measure on  $\Omega$ . A **probability measure** on  $\Omega$  is a function  $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$  such that

1.  $P(\emptyset) = 0, P(\Omega) = 1$ .
2. If  $A_i \in \mathcal{F}, i \in \mathbb{N}$  is a sequence of events with  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i).$$

*Remarque*

- (1) The probability that something happens is 1 and that nothing happens is 0.
  - (2) The probability of events that cannot occur simultaneously is the sum of the probabilities of the events.
- From these properties, we can deduce the next theorems.

##### Théorème 1

Let  $P$  be a probability measure on some realisation set  $\Omega$ . Then,

1.  $P(\emptyset) = 0, P(\Omega) = 1$ ;
2. for any event  $A, P(\Omega \setminus A) = 1 - P(A)$ ;
3. if two events  $A, B$  are such that  $A \subset B, P(B) = P(A) + P(B \setminus A)$ . In particular,  $P(A) \leq P(B)$ ;
4. for two events  $A, B$ ,

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

5. finite  $\sigma$ -additivity : if  $n \geq 2$ , and  $A_1, \dots, A_n$  are events such that  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i);$$

6. countable  $\sigma$ -additivity : if  $A_1, A_2, \dots$  are events such that  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i);$$

7. *finite  $\sigma$ -sub-additivity* : if  $n \geq 2$ , and  $A_1, \dots, A_n$  are events, then

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i);$$

8. *countable  $\sigma$ -sub-additivity* : if  $A_1, A_2, \dots$  are events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i);$$

9. *monotone convergence, increasing sequences* : if  $A_1, A_2, \dots$  are events such that  $A_i \subset A_{i+1}$  for all  $i$ 's then

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right);$$

10. *monotone convergence, decreasing sequences* : if  $A_1, A_2, \dots$  are events such that  $A_{i+1} \subset A_i$  for all  $i$ 's, then

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

*Remarque*

It's not necessary to learn this list by heart

We will encounter two main types of probability spaces in these notes :

- **Discrete probability spaces** : in that case  $\Omega$  is a finite or a countable set, and the set of events really is  $\mathcal{F} = \mathcal{P}(\Omega)$ .
- **Continuous probability spaces** : in that case,  $\Omega = \mathbb{R}^d$  with  $d \geq 1$  integer. We won't go into a formal definition of the *set of Borel sets*, and we will do as if we could take  $\mathcal{F} = \mathcal{P}()$

### 1.1.2 Discrete probability measures

#### Définition 2

Let  $\Omega$  be a finite countable set. A **probability mass function** on  $\Omega$  is a function  $p : \Omega \rightarrow [0, 1]$  such that

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

The **probability measure** associated to a probability mass function  $p$  is the function  $P_p : \mathcal{P}(\Omega) \rightarrow [0, 1]$  given by

$$P_p(A) = \sum_{\omega \in A} p(\omega).$$

### 1.1.3 Continuous probability measures

One cannot make sense of the probability that a drop of water falls at *precisely* one point  $x$ , but it is relatively easy to make sense of the probability that the drop falls *in a small disk* around  $x$ . This is the essence of the next definition.

#### Définition 3

Let  $d \geq 1$ . A **probability density function** on  $\mathbb{R}^d$  is a Riemann integrable function  $f : \mathbb{R}^d \rightarrow [0, +\infty)$  such that

$$\int_{-\infty}^{\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_d f(x_1, \dots, x_d) = 1.$$

The **probability measure** associated to a probability density function  $f$  is the  $[0, 1]$ -valued function  $P_f$  given by

$$P_f(A) = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_d f(x_1, \dots, x_d) \mathbb{1}_A(x_1, \dots, x_d).$$

*Remarque*

This is the equivalent of **the second definition** but for a space where you can't delimit the element to sum : a continuous space.

The **density function** is a function that shows where the variable like to be at most. The probability is the area under this curve.

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

### 1.1.4 Genreal probability spaces and measures

#### Définition 4

Let  $\Omega$  be a set. A **sigma-algebra** on  $\Omega$  is a set  $\mathcal{F} \subset \mathcal{P}(\Omega)$  which satisfies

1.  $\mathcal{F}$  contains the empty set ( $\emptyset \in \mathcal{F}$ ).
2.  $\mathcal{F}$  is stable by taking the complement ( $A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}$ ).
3.  $\mathcal{F}$  is stable by countable unions (if for all  $i \in \mathbb{N}$ ,  $A_i \in \mathcal{F}$ , then  $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$

*Remarque*

$\mathcal{P}(\Omega)$  is the set of all subset of  $\Omega$ .

Now we can deduce the following properties.

#### Théorème 2

Let  $\Omega$  be a set and  $\mathcal{F}$  a sigma-algebra on  $\Omega$ . Then all of the following hold.

1.  $\Omega \in \mathcal{F}$ .
2.  $\mathcal{F}$  is stable by finite intersections : if  $A, B \in \mathcal{F}$ , then  $A \cap B \in \mathcal{F}$ .
3. If  $A, B \in \mathcal{F}$ , then  $A \setminus B \in \mathcal{F}$ .
4.  $\mathcal{F}$  stable by countable intersections : if  $A_1, A_2, \dots \in \mathcal{F}$ , then  $\bigcap_{i \geq 1} A_i \in \mathcal{F}$ .
5.  $\mathcal{F}$  is stable by increasing limits : if  $A_i \in \mathcal{F}$ ,  $i \geq 1$  is such that  $A_i \subset A_{i+1}$ , then,  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .
6.  $\mathcal{F}$  is stable by decreasing limits : if  $A_i \in \mathcal{F}$ ,  $i \geq 1$  is such that  $A_{i+1} \subset A_i$ , then,  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ .

#### Définition 5

Let  $\Omega$  be a set and  $\mathcal{F}$  a sigma-algebra on  $\Omega$ . A **probability measure** on  $(\Omega, \mathcal{F})$  is a function  $P : \mathcal{F} \rightarrow [0, 1]$  such that

1.  $P(\Omega) = 1$ .
2. If  $A_i \in \mathcal{F}$ ,  $i \in \mathbb{N}$  is a sequence of events with  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i).$$

*Remarque*

The same as before but more precise.

### Définition 6

A **probability space** is a triplet  $\Omega, \mathcal{F}, P$  where  $\Omega$  is a set (the set of realisations),  $\mathcal{F}$  is a sigma-algebra on  $\Omega$  (the set of events), and  $P$  is a probability measure on  $(\Omega, \mathcal{F})$ .

Remarque

This is the most important definition of this “introduction”

### 1.1.5 Inclusion-Exclusion

It is a generalisation of the following fact that we encounter when counting objects : to count the number of objects with property A or property B, we can count the number of objects with property A add the number of objects with property B, and correct our over-counting by removing from this the number of objects with both property A and property B (which were counted twice).

### Théorème 3

Let  $P$  be a probability measure on some realisation set  $\Omega$ . Let  $n \geq 1$  and  $A_1, \dots, A_n$  be events. Then,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{i_1 < \dots < i_k \leq n} P\left(\bigcap_{j=1}^k A_{i_j}\right).$$

Moreover, for  $1 \leq l \leq \frac{n}{2}$  integer  $a$ ,

$$P\left(\bigcup_{i=1}^n A_i\right) = \begin{cases} \leq \sum_{k=1}^{2l-1} (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P\left(\bigcap_{j=1}^k A_{i_j}\right) \\ \geq \sum_{k=1}^{2l} (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P\left(\bigcap_{j=1}^k A_{i_j}\right) \end{cases}.$$

## 1.2 Random variables and expectation

### 1.2.1 Random variables

Random variables are therefore functions going from the set of realisations to the real numbers ; for example, if the “experiment ” is looking at all people born in 2000, one could make the measurement of the height of the first individual born that year.

### Définition 7

A (real) **random variable** is a function from the realisation space  $\Omega$  to  $\mathbb{R}$ . The probability that a random variable falls in a set  $A$  is

$$P(X \in A) := P(X^{-1}(A))$$

In words : it is the probability that the realisation of the experiment is such that the measurement  $X$  takes a value  $A$ .

Example

We are throwing a dice :

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $X(\omega) = 1$  if it's even,  $X(\omega) = 0$  if it's odd.
- So we are searching  $\omega$  in  $\Omega$  that gives  $X(\omega) = 1$  :

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = \{2, 4, 6\}$$

Then we apply the probability on these results

$$P(X \in A) = P(\{2, 4, 6\}) = \frac{3}{6} = 0.5$$

We will frequently use notations similar to the following :

$$\begin{aligned} P(X = x) &\equiv P(X \in \{x\}), \\ P(X \leq x) &\equiv P(X \in (-\infty, x]), \\ P(X > x) &\equiv P(X \in (x, +\infty)). \end{aligned}$$

### Définition 8

Let  $\Omega$  be a set of realisation, and let  $P$  be a probability measure on  $\Omega$ . Let  $X : \Omega \rightarrow \mathbb{R}$  be a random variable.  $X$  is **discrete** if there is  $\mathcal{D}_X \subset \Omega$  countable or finite such that  $P(X \in \mathcal{D}_X) = 1$ . The **law of  $X$**  is then the probability measure on  $\mathbb{R}$  given by

$$P_X(A) = \sum_{x \in A \cap \mathcal{D}_X} P(X = x)$$

In words, a **discrete random variable** is a variable that can take only finitely or countably many values with non-zero probability.

The second very important family of variables are **continuous random variables**.

### Définition 9

Let  $\Omega$  be a set of realisation, and let  $P$  be a probability measure on  $\Omega$ . Let  $X : \Omega \rightarrow \mathbb{R}$  be a random variable.  $X$  is **continuous random variable** if there is a density function  $f_X : \mathbb{R} \rightarrow [0, +\infty)$  such that

$$P(X \in A) = \int_{\mathbb{R}} \mathbb{1}_A f_X(x) dx.$$

The **law of  $X$**  is then the probability measure on  $\mathbb{R}$  given by  $P_{f_X}$ .

Remarque

Here we can't say : "the probability that  $X = 2$ ", because it will always be 0. Instead we use a density  $f_X(x)$  to compute on an interval. For example, if  $X$  is measuring the size of somebody,  $P(170 \leq X \leq 180)$  is compute with the density  $f_X$  by an integral, because the probability to have 170.000000cm is 0.

- Discrete : whe can say  $P(X = x)$ .
- Continue : values are infinite, we look at interval not precise points.

There is a similat notion of law for general random variables. Random variables allows us sometime to pass from some continuous probability to some discrete random variables.

### Définition 10

Let  $\Omega$  be a set of realisation, and let  $P$  be a probability measure on  $\Omega$ . Let  $X : \Omega \rightarrow \mathbb{R}$ . The **law of  $X$**  is the probability measure  $P_X$  on  $\mathbb{R}$  given by

$$P_X(A) = P(X \in A).$$

### Définition 11

Let  $X : \Omega \rightarrow \mathbb{R}$  be a random variable. The **cumulative distribution function of  $X$**  is defined by

$$F_X(t) = P(X \leq t).$$

Two random variables have the same law if and only if they have the same cumulative distribution functions. Note that if  $X$  is a continuous random variable with density  $f_X$ , one has that  $F_X$  is a primitive of  $f_X$  :

$$F'_X(t) = f_X(t) \quad \forall t \in \mathbb{R}.$$

Remarque

$$F_x : \mathbb{R} \rightarrow [0, 1]$$

Example 1 : If  $P(X = 0) = P(X = 1) = \frac{1}{2}$

$$F_X(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & 0 \leq t < 1 \\ 1 & t \geq 1 \end{cases}$$

### Théorème 4

Let  $x \rightarrow \int_a^x f(t) dt$  is primitive of  $f$

If  $X$  is a continuous random variable : there exist a density function

$$f_X : \mathbb{R} \rightarrow \mathbb{R}, \quad P(X) = \int_{\mathbb{R}} \mathbb{1}_A(x) f_X(x) dx$$

So  $F_X$  is differentiable and

$$F'_X(t) = f_X(t) = \int_{-\infty}^t f_X(x) dx$$

Example

If  $X$  is a continuous random variable with density  $f_X(x) = \mathbb{1}_{[0,4]}(x)$  then

$$P(X \leq t) = \int_{-\infty}^t f_X(x) dx = \frac{1}{4} \int_{-\infty}^t \mathbb{1}_{[0,4]}(x) dx = \begin{cases} 0 & \text{if } t < 0, \\ \frac{t}{4} & \text{if } t \in [0, 4], \\ 1 & \text{if } t > 4. \end{cases}$$

Remarque

to

**Remarque importante**

To resume an important point :

- for discrete random variables we use a **probability masse function**.
- for continue random variables we use a **density function**.



## 1.2.2 Expectation

### Définition 12

**Discrete case :** If  $\Omega$  is a finite or countable set,  $p : \Omega \rightarrow [0, 1]$  is a probability mass function, and  $X : \Omega \rightarrow \mathbb{R}$  is a random variable such that

$$\sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty,$$

the **expectation of X under p** is

$$E_p(X) := \sum_{\omega \in \Omega} X(\omega) p(\omega).$$

In this case, we say that X is  **$P_p$ -integrable**.

**Continuous case :** If  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is a probability density function, and  $X : \mathbb{R}^d \rightarrow \mathbb{R}$  is a random variable such that

$$\int_{\mathbb{R}^d} dx |X(x)| f(x) < \infty,$$

the **expectation of X under  $P_f$**  is

$$E_f(x) := \int_{\mathbb{R}^d} dx X(x) f(x).$$

In this case, we say that X is  **$P_f$ -integrable**.

**General case :** If  $\Omega$  is a set of realisation, and P is a probability measure on  $\Omega$ , we will denote  $E_p(X)$  the **expectation of X under P**.

*Remarque*

Note that the expected value of a random variable depends only on its law

*Example*

Look at a fair 6-face dice roll :  $\Omega = \{F1, F2, \dots, F6\}$ ,  $p(\omega) = \frac{1}{6}$  for every  $\omega \in \Omega$ .

Take the random variable  $X(F1) = X(F3) = X(F5) = -1$ ,  $X(F2) = X(F4) = X(F6) = 2$ , then

$$\begin{aligned} E_p(X) &= p(F1)X(F1) + p(F2)X(F2) + p(F3)X(F3) + p(F4)X(F4) \\ &\quad + p(F5)X(F5) + p(F6)X(F6) = -\frac{1}{6} + \frac{1}{6} \cdot 2 - \frac{1}{6} + \frac{1}{6} \cdot 2 - \frac{1}{6} + \frac{1}{6} \cdot 2 = \frac{1}{2}. \end{aligned}$$

The properties of expectation are summarized in the next Theorem.

### Théorème 5

Let  $\Omega$  be a realisation set, and P a probability measure on  $\Omega$ . Let  $X, Y : \Omega \rightarrow \mathbb{R}$  be two random variables. Then,

1. *linearity* : for any  $a, b \in \mathbb{R}$ ,  $E_P(aX + bY) = aE_P(X) + bE_P(Y)$  ;
2. *ordering* : if  $P(X \geq Y) = 1$ ,  $E_P(X) \geq E_P(Y)$ . In particular,
  - if  $P(X \geq 0) = 1$ , then  $E_P(X) \geq 0$  ;
  - if  $P(a \leq X \leq b) = 1$ , then  $a \leq E_P(X) \leq b$  ;
  - $|E_P(X)| \leq E_P(|X|)$ .

### 1.2.3 Transfer Theorem

The transfer theorem confirm us that the intuition that if we have a random variable that takes values  $x_1, x_2, x_3$ , with probabilities  $p_1, p_2, p_3$ , and we have a function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , then expectation of  $g(X)$  should be

$$E(g(X)) = p_1 g(x_1) + p_2 g(x_2) + p_3 g(x_3).$$

#### Théorème 6

Let  $X$  be a random variable. Then,

1. if  $X$  is a discrete random variable, for any  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(X)$  is a discrete random variable, and

$$E(g(X)) = \sum_{x \in \text{Image}(X)} g(x) P(X = x)$$

as soon as the sum converges absolutely;

2. if  $X$  is a continuous random variable, for any  $g : \mathbb{R} \rightarrow \mathbb{R}$

$$E(g(X)) = \int_{-\infty}^{+\infty} f_X(x) g(x) dx$$

as soon as the integral converges absolutely.

### 1.2.4 Random vectors

#### Définition 13

Let  $\Omega$  be a realisation set, and  $P$  a probability measure on  $\Omega$ . Let  $d \geq 1$ . A **random vector of dimension  $d$**  is function  $X : \Omega \rightarrow \mathbb{R}^d$ . We will denote

$$X(\omega) = (X_1(\omega), \dots, X_d(\omega)).$$

The functions  $X_i : \Omega \rightarrow \mathbb{R}$  are random variables. They are called the **marginals** of  $X$ .

The **cumulative distribution function** (CDF) of a random vector  $X : \Omega \rightarrow \mathbb{R}^d$  is given by  $F_X : \mathbb{R}^d \rightarrow [0, 1]$ ,

$$F_X(t_1, \dots, t_d) = P(X_1 \leq t_1, \dots, X_d \leq t_d)$$

Remarque

Random vectors are just a list of random variables.

Pas sur de l'utilité de celle la

#### Définition 14

Let  $\Omega$  be a realisation set, and  $P$  a probability measure on  $\Omega$ . A **complex random variable** is a function  $X : \Omega \rightarrow \mathbb{C}$ . The real and imaginary parts of  $X$  are then random variables.

There is then also discrete and continuous random vectors

#### Définition 15

Let  $\Omega$  be a realisation set, and  $P$  a probability measure on  $\Omega$ . Let  $X : \Omega \rightarrow \mathbb{R}^d$  be a random vector. We say that

- $X$  is a **discrete random vector** if there is a finite or countable set  $\mathcal{D}_X \subset \mathbb{R}^d$  with  $P(X \in \mathcal{D}_X) = 1$ ;
- $X$  is a **continuous random vector** if there is a density function  $f_X : \mathbb{R}^d \rightarrow [0, +\infty)$  such

that

$$P(X \in A) = \int_{\mathbb{R}^d} \mathbb{1}_A(x) dx$$

*Example*

Consider  $(X, Y)$  a uniform random vector in the unit disc :

$$f_{(X,Y)}(x, y) = \frac{1}{\pi} \mathbb{1}_{[0,1]}(x^2 + y^2).$$

The first marginal,  $X$ , of this random vector is then a continuous random variable with density given by  $f_X(x) = 0$  for  $|x| > 1$ , and, for  $|x| \leq 1$ ,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy = \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \mathbb{1}_{[0,1-x^2]}(y^2) dy = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1-x^2}. \end{aligned}$$

*Hors sujet important*

Let a random vector have two variables, then his density function will be  $f_{XY}(x, y)$ . So the density of the random variable  $X_1$  will be  $f_X(x) = \int_{\mathbb{R}} f_{XY}(x, y) dy$  and the expected value of  $X_1$  will be  $E[X_1] = \int_{\mathbb{R}} x f_X(x) dx = \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{XY}(x, y) dx dy$ .

### 1.2.5 Change of variable formula

Let  $d \geq 1, U \subset \mathbb{R}^d$  an open set, and  $\phi : U \rightarrow \mathbb{R}^d, \phi(y) = (\phi_1(y), \dots, \phi_d(y)), y = (y_1, \dots, y_d)$ . Then,

- we say that  $\phi$  is **continuously differential** on  $U$  if the partial derivative  $\frac{\delta \phi_i}{\delta y_j}$  exists and are continuous on  $U$ ;
- we denote  $D_\phi(y)$  the **Jacobian matrix** of  $\phi$ ;
- we denote  $\det$  the determinant.

#### Théorème 7

Let  $d \geq 1, U \subset \mathbb{R}^d$  an open set,  $V \subset \mathbb{R}^d$ , and  $\phi : U \rightarrow V$  a continuously differentiable bijection with  $\det D_\phi(y) \neq 0$  for all  $y \in U$ . Then, for any function  $f : V \rightarrow \mathbb{R}$ , we have :

$$\int_U dy_1 \dots dy_d f(\phi(y)) |\det D_\phi(y)| = \int_V dx_1 \dots dx_d f(x).$$

From this, we can deduce :

#### Théorème 8

Let  $\Omega$  be a realisation set,  $P$  be a probability measure on  $\Omega$ , and  $X : \Omega \rightarrow \mathbb{R}^d$  a random vector with a density  $f_X$ . Let  $U, V, \phi : U \rightarrow V$ , be as in **the theorem above**. Suppose that  $P(X \in U) = 1$ . We then have that  $Y = \phi \odot X : \Omega \rightarrow \mathbb{R}^d (Y = \phi(X))$  is a continuous random vector with density

$$f_Y(y) = f_X(\phi^{-1}(y)) |\det D_{\phi^{-1}}(y)| = \frac{1}{|\det D_\phi(\phi^{-1}(y))|} f_X(\phi^{-1}(y)).$$

In the case  $d = 1$ , this formula simplifies to

$$f_Y(y) = f_X(\phi^{-1}(y)) |(\phi^{-1})'(y)| = \frac{1}{|\phi'(\phi^{-1}(y))|} f_X(\phi^{-1}(y)).$$

Example

We take  $U$  a uniform random variable on  $[0, 1]$  :  $U$  is a continuous random variable with density  $f_U(x) = \mathbb{1}_{[0,1]}(x)$ . Then for  $a \in \mathbb{R}$  and  $r > 0$ , define  $X = a + rU$ . Using theorem 8 with  $\phi(x) = a + rx$ ,  $\phi^{-1}(x) = \frac{x-a}{r}$ , we get that  $X$  is a continuous random variable with density

$$f_X(x) = f_U(\phi^{-1}(x)) \frac{1}{\phi'(\phi^{-1}(x))} = \mathbb{1}_{[0,1]} \left( \frac{x-a}{r} \right) \frac{1}{r} = \frac{1}{r} \mathbb{1}_{[a, a+r]}(x).$$

So,  $X$  is a uniform random variable on  $[a, a+r]$ .

Remarque

With polar coordinates, we then have

$$\phi^{-1}(x, y) = \left( \text{atan2}(y, x), \sqrt{x^2 + y^2} \right)$$

where

$$\text{atan2}(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{if } x > 0, \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{if } x < 0, y \geq 0, \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{if } x < 0, y < 0, \\ \frac{\pi}{2} & \text{if } x = 0, y > 0, \\ -\frac{\pi}{2} & \text{if } x = 0, y < 0, \\ \text{undefined} & \text{if } x = y = 0. \end{cases}$$

In this case

$$|\det D_\phi(\theta, r)| = r,$$

which leads to the formula

$$dxdy = rd\theta dr$$

### 1.2.6 Moments, and Moment Generating Function

We already saw the cumulative distribution function (CDF), there is an other useful object some-time : the **moment generating function**

#### Définition 16

Let  $X$  be a random variable. Let  $p > 0$ . We say that  $X$  **admits a moment of order  $p$**  if

$$E(|X|^p) < \infty$$

When  $X$  admits a moment of order  $p$ , we define

- the  *$p$ th moment* of  $X$  :  $E(X^p)$ ;
- the  *$p$ th absolute moment* of  $X$  :  $E(|X|^p)$ .

#### Définition 17

Let  $X$  be a random variable. we say that  $X$  **admits exponential moments of order  $\delta > 0$**  if

$$E\left(e^{\delta|X|}\right) < \infty$$

When  $X$  admits exponential moments, we define the **moment generating function of  $X$**  by

$$M_X(t) = E\left(e^{tx}\right), \quad t \in (-\delta, \delta).$$

### Théorème 9

Let  $X, Y$  be two random variables. Suppose that there is  $\delta > 0$  such that

$$E\left(e^{\delta|X|}\right) < \infty, \quad E\left(e^{\delta|Y|}\right) < \infty.$$

Then, we have the following properties.

- $X$  admits moments of any integer order.
- $M_X$  is analytic in a neighbourhood of 0, and for any  $n \in \mathbb{N}$ ,
- $M_X, M_Y$  characterise  $X, Y$  :

$$M_X(t) = M_Y(t) \text{ for all } t \in (-\delta, \delta) \implies X = Y.$$

## 1.3 Conditional probability independence

### 1.3.1 Conditional probability

#### Définition 18

Let  $\Omega$  be a set of realisations, and  $P$  a probability measure on  $\Omega$ . Let  $A \subset \Omega$  be an event such that  $P(A) > 0$ . Define then the **probability measure  $P$  conditioned on  $A$** , denoted  $P(\cdot|A)$ , by

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad \forall \text{ event } B.$$

We can then define the **conditions expectation** of a random variable  $X : \Omega \rightarrow \mathbb{R}$  by

$$E_P(X|A) = E_{P_A}(X),$$

where  $P_A$  stands for  $P_A = P(\cdot|A)$ .

### 1.3.2 Independence

#### Définition 19

- Two events  $A, B$  are said to be **independent** if

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

- A family of events  $(A_i)_{i \in I}$  is said to be **two-by-two independent** if for any  $i \neq j$ ,  $A_i$  and  $A_j$  are independent.
- A family of events  $(A_i)_{i \in I}$  is said to be **an independent family** if for any  $J \subset I$  finite,

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$

#### Définition 20

- Two random variables  $X, Y$  are said to be **independent** if for any events  $A, B \subset \mathbb{R}$ ,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Equivalently,  $X, Y$  are independent if for any  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$E(f(X)g(Y)) = E(f(X))E(g(Y)).$$

- A family of random variables  $(X_i)_{i \in I}$  is said to be **two-by-two independent** if for any  $i, X_i$  and  $X_j$  are independent.

- A family of random variables  $(X_i)_{i \in I}$  is said to be **an independent family** if for any  $J \subset I$  finite, and any events  $A_i \subset \mathbb{R}, i \in J$ ,

$$P(\cap_{i \in J} \{X_i \in A_i\}) = \prod_{i \in J} P(X_i \in A_i).$$

Equivalently,  $(X_i)_{i \in I}$  is an independent family if for any  $J \subset I$  finite, and any functions  $f_i : \mathbb{R} \rightarrow \mathbb{R}, i \in J$ ,

$$E\left(\prod_{i \in J} f_i(X_i)\right) = \prod_{i \in J} E(f_i(X_i)).$$

The same definition holds with “random vectors” replacing “random variables”.

### Définition 21

A family  $(X_i)_{i \in I}$  of random variables is called an **independent identically distributed** family, abbreviated **i.i.d. family**, if the family  $(X_i)_{i \in I}$  is an independent family, and for any  $i, j \in I, X_i = X_j$ .

### Théorème 10

Let  $d, d' \geq 1$ . Let  $X : \Omega \rightarrow \mathbb{R}^d, Y : \Omega \rightarrow \mathbb{R}^{d'}$  be a random vector.

- If  $X, Y$  are **continuous random vector** :  $X$  and  $Y$  are independent if and only if the random vector  $(X, Y) : \Omega \rightarrow \mathbb{R}^{d+d'}$  has density

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y),$$

where  $f_X : \mathbb{R}^d \rightarrow \mathbb{R}$  is a density for  $X$ , and  $f_Y : \mathbb{R}^{d'} \rightarrow \mathbb{R}$  is a density for  $Y$ .

- If  $X, Y$  are **discrete random vectors** :  $X$  and  $Y$  are independent if and only if for any  $x \in \mathbb{R}^d, y \in \mathbb{R}^{d'}$ ,

$$P(X = x, Y = y) = P(X = x) P(Y = y).$$

- If  $X$  is **discrete** and  $Y$  is **continuous** :  $X$  and  $Y$  are independent if and only if for any  $x \in \mathbb{R}^d, A \subset \mathbb{R}^{d'}$ ,

$$P(X = x, Y \in A) = P(X = x) \int_A f_Y(y) dy$$

where  $f_Y : \mathbb{R}^{d'} \rightarrow \mathbb{R}$  is a density for  $Y$ .

### 1.3.3 Bayes law, formula of total probability

Bayes Law :

### Théorème 11

Let  $\Omega$  be a set of realisations, and let  $P$  be a probability measure on  $\Omega$ . Let  $A, B \subset \Omega$  be two events such that  $P(A), P(B) > 0$ . Then

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

### Théorème 12

Let  $\Omega$  be a set of realisations, and let  $P$  be a probability measure on  $\Omega$ . Let  $I$  be a finite or countable set. Let  $A_i, i \in I$  be a collection of events such that

- if  $i \neq j$ , then  $A_i \cap A_j = \emptyset$ ;
- $\cup_{i \in I} A_i = \Omega$ .

Suppose moreover that  $P(A_i) > 0$  for all  $i \in I$ . Then for any event  $B$ ,

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(B|A_i) P(A_i).$$

In the same fashion, for every random variable  $X$

$$E(X) = \sum_{i \in I} E(X|A_i) P(A_i).$$

*Example*

We throw some dices :

- $A_1$  : we throw an even dice
- $A_2$  : we throw an odd dice
- $B$  : the result is  $\leq 4$

Then

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

### 1.3.4 Almost sure properties

#### Définition 22

An event  $A$  is said to occur **almost-surely** if

$$P(A) = 1.$$

## 1.4 Correlation

### 1.4.1 Variance, Covariance

Variance is a way to quantify “how far from its mean is typically my variable”. If every value that  $X$  can take is not far from the mean of every value of  $X$ , then the variance will be small.

#### Définition 23

Let  $\Omega$  be a realisation set and  $P$  a probability measure on  $\Omega$ . Let  $X : \Omega \rightarrow \mathbb{R}$  be a random variable. The **variance of  $X$**  is given by

$$\text{Var}_P(X) := E_P\left((X - E_P(X))^2\right).$$

Alternatively,  $\text{Var}_P(X) = E_P(X^2) - E_P(X)^2$ .

The inside of the expected value in the definition on  $\text{Var}(X - E_P(X))$  is called the standard deviation

#### Définition 24

The **standard deviation** of a random variable  $X$ , often denoted as  $\sigma_X$ , is the square root of its variance :

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

### Définition 25

Let  $\Omega$  be a realisation set and  $P$  a probability measure on  $\Omega$ . Let  $X, Y : \Omega \rightarrow \mathbb{R}$  be two random variables. The **covariance between  $X$  and  $Y$**  is given by

$$\text{Cov}_P(X, Y) := E_P(XY) - E_P(X)E_P(Y).$$

When  $\text{Cov}_P(X, Y) = 0$ , we say that  $X$  and  $Y$  are **uncorrelated**.

*Remarque*

The covariance between  $X$  and  $Y$  is a measure of how much “typical large values of  $X$ ” and “typical large values of  $Y$ ” are influencing each other. Two independent event have a covariance of 0 (the opposite isn’t true!). But there is one case where uncorrelated implies independent : it’s with Bernoulli random variables

### Théorème 13

Let  $x, Y$  be two random variables such that

$$P(X \in \{0, 1\}) \equiv P(Y \in \{0, 1\}) = 1.$$

Such variable are called **Bernoulli random variables**. Then,  $X$  and  $Y$  are independent if and only if  $\text{Cov}(X, Y) = 0$ .

### Théorème 14

Let  $X, Y, Y_1, Y_2$  be random variables and  $a, b \in \mathbb{R}$ . Then,

$$\text{Cov}(X, Y) = \text{Cov}(Y, X), \quad \text{Cov}(aX, bY) = ab\text{Cov}(x, Y),$$

$$\text{Cov}(X_1, Y_1 + Y_2) = \text{Cov}(X, Y_1) + \text{Cov}(X, Y_2).$$

In words : Cov is symmetric, and linear in each of its arguments.

## 1.4.2 Pearson correlation coefficient

In statistics, a coefficient obtained from the covariance and standard deviation is frequently used : the Pearson correlation coefficient.

### Définition 26

For two random variables  $X, Y$  define their **Pearson correlation coefficient** :

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_Y \sigma_x} \in [-1, 1].$$

$|\rho_{X,Y}| = 1$  if and only if  $X$  and  $Y$  are related by an affine transformation (i.e : there are  $a, b \in \mathbb{R}$  such that  $Y = aX + b$ ).

*Remarque*

We normalize by the product of variance because “ the height of Bob influences the height of Alice” should not depend on unit we chose to measure height, but the covariance does, so it’s a way to correct this