I. Foundations of Probability Theory.

<u>informations</u> : – $2 \times 2h$ / week Tuesday 8h15 → 10h15
Thursday
+ exercise sessions just after.

– $2 \times 6 = 12$ weeks.
partial exam 19–23 oct.
exam 14–18 dec
holidays 26–30 oct.

– source : book by Jean-François Le Gall
"Measure Theory, Probability, and Stochastic Processes".
+ hand written notes.

<u>objective</u> : learn all the basics of probability theory and discrete time stochastic processes
↓ random variables and objects
X ω
↓ sequences of random variables
$(X_n)_{n\in\mathbb{N}}$.

<u>contents</u> :
1. Foundations of probability theory.
2. Conditioning.
3. Convergence of sequences of random variables.
4. Martingales.
5. Markov chains.

Chapter I : Foundations of probability theory.

1. Probability spaces.

In order to model an experiment with random outcome, one can use Lebesgue theory of measure :

<u>Definition</u> (Probability space).

A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where:

- Ω is a set (all the possible outcomes of the experiment).
- $\mathcal{F}$ is a σ-field or σ-algebra on Ω:
  - $\mathcal{F} \subset \mathcal{P}(\Omega)$.
  - $\phi, \Omega \in \mathcal{F}$.
  - if $A \in \mathcal{F}$, $A^c = \Omega \setminus A \in \mathcal{F}$ (stable by complementation)
  - if $(A_n)_{n\in\mathbb{N}}$ is a sequence of measurable subsets with $A_n \in \mathcal{F}$, then $\bigcup_{n\in\mathbb{N}} A_n$ and $\bigcap_{n\in\mathbb{N}} A_n$ are also measurable (i.e. in $\mathcal{F}$). (σ-stability).

The elements of $\mathcal{F}$ are also called events. One should think of them as the subsets $A \subset \Omega$ for which one is able to answer after the experiment: "did A occur or not?".

- $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is a probability measure, which means:
  - $\mathbb{P}[\phi] = 0$, $\mathbb{P}[\Omega] = 1$.
  - if $(A_i)_{i\in I}$ is a countable set of disjoint measurable subsets, then
$$\mathbb{P}\Big[\bigsqcup_{i\in I} A_i\Big] = \sum_{i\in I} \mathbb{P}[A_i].$$
(countable additivity).

This implies that $\mathbb{P}$ is nondecreasing w.r.t. inclusion, and that $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.

Often, one has only access to certain observables of the random experiment.

<u>Definition</u> (Random variable, law).
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable defined on this space

is a measurable map $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ towards another measurable space: $\forall B \in \mathcal{E},\ X^{-1}(B) \in \mathcal{F}$.

Then X induces a probability measure $\mathbb{P}_X$ on $(E, \mathcal{E})$ called the law or distribution of X:

$$\mathbb{P}_X[B] = \underbrace{(X_* \mathbb{P})}_{\text{image measure}}(B) = \mathbb{P}[X^{-1}(B)] \quad \forall B \in \mathcal{E}$$

<u>Fundamental example</u>: The dice throw.

Consider a balanced 6-faces dice.

$\Omega = \{$all the parameters of a throw: speed, orientation, height of the table, etc.$\}$

$\mathcal{F} =$ some σ-algebra which makes the parameter maps

$$\left.\begin{aligned} \omega &\longmapsto \text{speed}(\text{throw } \omega) \in \mathbb{R}^3 \\ &\longmapsto \text{orientation}(\text{throw } \omega) \in \mathbb{S}^2 \\ &\text{etc.} \end{aligned}\right) \text{ measurable.}$$

$\mathbb{P} =$ some probability measure on $(\Omega, \mathcal{F})$, difficult to describe.

One has only access to the random variable

$$\begin{aligned} X : \Omega &\longrightarrow [\![1, 6]\!]. \\ \omega &\longmapsto \text{result of the throw } \omega. \end{aligned}$$

and one expects that $\mathbb{P}_X =$ uniform law on $[\![1,6]\!]$.

$$\mathbb{P}_X[k] = \frac{1}{6} \quad \forall k \in [\![1, 6]\!].$$

$\rightsquigarrow$ one can mostly forget about the initial probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and focus on the space induced by X:

$$([\![1,6]\!], \mathcal{P}([\![1,6]\!]), \text{uniform law}).$$

2. Tools and constructions

1) the σ-algebra

Definition Let Ω be a set and $\mathcal{C} \subset \mathcal{P}(\Omega)$ be a class of subsets. There is a smallest σ-algebra $\mathcal{F}$ on Ω (for the inclusion) which contains the class $\mathcal{C}$. It is called the σ-algebra spanned by $\mathcal{C}$ and is denoted $\mathcal{F} = \sigma(\mathcal{C})$.

Proof: σ-algebras are stable by intersection. Just take the (non-empty) intersection of all σ-algebras containing $\mathcal{C}$. $\square$.

Example: Ω = topological space endowed with $\mathcal{T}$ = family of open subsets.
$\mathcal{B}(\Omega)$ = Borel σ-algebra of $\Omega = \sigma(\mathcal{T})$.

If the set of outcomes of the random experiment is endowed with a topological or metric structure, it would be crazy not to take $\mathcal{F} = \mathcal{B}(\Omega)$, or at least $\mathcal{F} \subset \mathcal{B}(\Omega)$.

Particular case: Ω endowed with an observable $X : \Omega \to (E, \mathcal{E})$.

$$\mathcal{F} = \sigma(X) = \sigma(\{X^{-1}(B), B \in \mathcal{E}\})$$
$$= \sigma\text{-algebra spanned by the map } X$$
$$= \{X^{-1}(B), B \in \mathcal{E}\}.$$

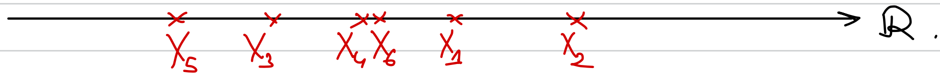
More generally, given a family of observables $(X_i : \Omega \to (E_i, \mathcal{E}_i))_{i \in I}$ one can consider

$$\mathcal{F} = \sigma(X_i, i \in I) = \sigma(\{X_i^{-1}(B_i), B_i \in \mathcal{E}_i, i \in I\})$$
$$= \text{smallest } \sigma\text{-algebra which makes all the maps } X_i \text{ measurable.}$$

Example: Consider a sequence of real random variables $(X_n)_{n \geq 1}$:
$X_n : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The n-th empirical measure of the sequence is the ω-dependent probability measure

$$\mu_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}.$$



Denote $\mathcal{M}^1(\mathbb{R}) = \{$ probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$.

One would like to see μ_n as a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and with values in $\mathcal{M}^1(\mathbb{R})$ (random measure).

Which σ-algebra on $\mathcal{M}^1(\mathbb{R})$?

$$\mathcal{E} = \sigma(\pi_A,\ A \in \mathcal{B}(\mathbb{R})) \quad \text{with } \pi_A(\mu) = \mu(A).$$

Then, μ_n is indeed a measurable map $(\Omega, \mathcal{F}) \to (\mathcal{M}^1(\mathbb{R}), \mathcal{E})$:

μ_n measurable $\iff \forall A \in \mathcal{B}(\mathbb{R})$, $\pi_A \circ \mu_n : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable.

But $\pi_A \circ \mu_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(X_i(\omega) \in A)}$ is a linear combination of measurable maps.

<u>note</u>: actually, there is a nice metric structure on $\mathcal{M}^1(\mathbb{R})$ s.t. $\mathcal{E} = \mathcal{B}(\mathcal{M}^1(\mathbb{R}))$.

<u>Other important case</u>: product of spaces

If $(E_i, \mathcal{E}_i)_{i \in I}$ is a family of measurable spaces, then $E = \prod_{i \in I} E_i$ is naturally endowed with $\bigotimes_{i \in I} \mathcal{E}_i$

$$= \sigma(\{\pi_i : E \to E_i \text{ projection map}\}).$$

$X = (X_i)_{i \in I}$ with values in E is measurable $\Longleftrightarrow$ all the X_i are measurable.

② The probability measure.

Recall the monotone class lemma:

> Ω = some set; $\mathcal{C}$ = some class of subsets, containing ϕ and Ω.
> $\mathcal{M}$ = monotone class spanned by $\mathcal{C}$: stable by complementation and increasing countable union.
> If $\mathcal{C}$ is stable by simple intersection: $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$
> then $\mathcal{M} = \sigma(\mathcal{C})$.

This is often used to prove the unicity of a probability measure with certain properties:

Lemma:

If $\mathcal{C}$ is an algebra of subsets and $\mathbb{P}, \mathbb{Q}$ are probability measures on $(\Omega, \sigma(\mathcal{C}))$ such that $\mathbb{P}_{|\mathcal{C}} = \mathbb{Q}_{|\mathcal{C}}$, then in fact $\mathbb{P} = \mathbb{Q}$.

Proof: Let $\mathcal{M} \subset \sigma(\mathcal{C})$ be the set of subsets on which $\mathbb{P}$ and $\mathbb{Q}$ agree.

- by assumption, $\mathcal{C} \subset \mathcal{M}$.
- if $\mathbb{P}[A] = \mathbb{Q}[A]$ then $\mathbb{P}[A^c] = 1 - \mathbb{P}[A] = 1 - \mathbb{Q}[A] = \mathbb{Q}[A^c]$
  $\Rightarrow$ stability by complementation.
- if $\bigcup_{n \in \mathbb{N}} \uparrow A_n$ is an increasing union of subsets s.t. $\mathbb{P}[A_n] = \mathbb{Q}[A_n]$ $\forall n \in \mathbb{N}$,

then

$$\mathbb{P}\Big[\bigcup_{n \in \mathbb{N}} \uparrow A_n\Big] = \int_\Omega \lim \uparrow \mathbb{1}_{A_n}(\omega)\, \mathbb{P}[d\omega]$$

$$\underset{\text{(monotone convergence)}}{=} \lim \uparrow \int_\Omega \mathbb{1}_{A_n}(\omega)\, \mathbb{P}[d\omega]$$

$$= \lim_{n \to \infty} \mathbb{P}[A_n] = \dots = \mathbb{Q}\Big[\bigcup_{n \in \mathbb{N}} \uparrow A_n\Big].$$

So $\mathcal{M}$ is a monotone class $\Longrightarrow \mathcal{M} = \sigma(\mathcal{C})$ $\square$.

The only non-trivial construction of measure that we shall use: product of probability measures.

$(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i \in I}$ family of probability spaces with arbitrary cardinality.

$\Omega = \prod_{i \in I} \Omega_i$.

$\mathcal{F} = \bigotimes_{i \in I} \mathcal{F}_i = \sigma\left(\text{"cylinders"} \; A_{j_1} \times A_{j_2} \times \dots \times A_{j_n} \times \prod_{i \notin J} \Omega_i\right)$

with $J = \{j_1 \dots j_n\} \subset I$.

the $A_{j_k} \in \mathcal{F}_{j_k}$.

There is a unique probability measure $\mathbb{P} = \bigotimes_{i \in I} \mathbb{P}_i$ on the product space such that for any cylinder,

$$\mathbb{P}\left[A_{j_1} \times \dots \times A_{j_n} \times \prod_{i \notin J} \Omega_i\right] = \mathbb{P}_{j_1}[A_{j_1}] \dots \mathbb{P}_{j_n}[A_{j_n}].$$

Sketch of proof:
- the unicity is immediate from the monotone class lemma.
- the existence is a particular case of the Kolmogorov extension theorem, which is itself an application of Caratheodory-Hahn extension (also used in the construction of the Lebesgue measure).
  (going from a σ-additive measure on an algebra to a true measure on a σ-algebra.

3. Classical distributions of random variables.

Many observables of random experiments are real valued random variables:
$X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Let us review the interesting laws $\mathbb{P}_X$ which occur.

① discrete random variables with values in $\mathbb{Z}$.

If $\mathbb{P}[X(\omega)\in\mathbb{Z}]=1$, then $\mathbb{P}_X$ is entirely determined by the values
$\mathbb{P}_X[k]=\mathbb{P}[X(\omega)=k]$.

Indeed, $\forall B\in\mathcal{B}(\mathbb{R})$, $\mathbb{P}_X[B]=\mathbb{P}[X(\omega)\in B]$

$$= \mathbb{P}[X(\omega)\in B\cap\mathbb{Z}]$$

$$(\text{countable additivity}) = \sum_{k\in B\cap\mathbb{Z}} \mathbb{P}_X[k].$$

We shall also be interested in the following quantities (if well defined):

<u>Definition</u> The expectation or mean of a real random variable $X:(\Omega,\mathcal{F},\mathbb{P})\to\mathbb{R}$ is:

$$\mathbb{E}[X] = \int_\Omega X(\omega)\,\mathbb{P}[d\omega] \qquad (\text{well defined if } X\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbb{P}))$$

$$= \int_{\mathbb{R}} x\ \mathbb{P}_X[dx].$$

The variance of X, well defined if $X\in\mathcal{L}^2(\Omega,\mathcal{F},\mathbb{P})$, is

$$\mathrm{var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2]$$

$$= \mathbb{E}[X^2-2X\mathbb{E}[X]+\mathbb{E}[X]^2]$$

$$= \mathbb{E}[X^2]-\mathbb{E}[X]^2.$$

These quantities allow one to easily control the dispersion of X, through the most important inequality of probability theory:

<u>Proposition</u> (Bienaymé-Chebyshev)

If $X\in\mathcal{L}^2(\Omega,\mathcal{F},\mathbb{P})$, then

$$\mathbb{P}[|X(\omega)-\mathbb{E}[X]|\geq a] \leq \frac{\mathrm{var}(X)}{a^2} \qquad \forall a>0$$

Indeed, $\text{LHS} = \int_\Omega \mathbb{1}_{|X(\omega)-\mathbb{E}[X]|\geq a} \, \mathbb{P}[d\omega]$

$$\leq \int_\Omega \mathbb{1}_{|X(\omega)-\mathbb{E}[X]|\geq a} \frac{(X(\omega)-\mathbb{E}[X])^2}{a^2} \, \mathbb{P}[d\omega]$$

$$\leq \frac{1}{a^2}\int_\Omega (X(\omega)-\mathbb{E}[X])^2 \, \mathbb{P}[d\omega] = \frac{\mathrm{var}(X)}{a^2}. \quad \square$$

<u>notations and remarks</u>:

- If $A \in \mathcal{F}$, then $\mathbb{1}_A$ is a real random variable with expectation $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$.
- $\mathbb{E}[\cdot] = \int_\Omega \cdot(\omega) \, \mathbb{P}[d\omega]$, so all that is true for integrals is in general true for expectations.
  - $\longrightarrow$ additivity, linearity, monotonicity.
  - $\longrightarrow$ monotone convergence theorem, dominated convergence theorem, Fatou.
- One often consider real random variables in a L^p space, whence, defined up to modification on a event with $\mathbb{P}$-probability 0.
  One says that something holds almost surely if $\mathbb{P}[\text{something}] = 1$.
  Thus, if $X = Y$ almost surely, then the random variables are considered equal.
- One often forgets the $\cdot(\omega)$ in the notations, so for instance,
  $$\mathbb{P}[X \in B] = \mathbb{P}[\{X(\omega) \in B\}].$$
  This agrees with the general philosophy: $(\Omega, \mathcal{F}, \mathbb{P})$ is a large universe of possibilities, and one is only interested in observables X.

For discrete random variables, $\mathbb{E}[X] = \mathbb{E}\left[\sum_{k\in\mathbb{Z}} k \, \mathbb{1}_{k\in\mathbb{Z}}\right]$

$$\underset{(\text{TCD})}{=} \sum_{k\in\mathbb{Z}} k \, \mathbb{P}_X[k].$$

If $X \in \mathbb{N}$ a.s., the distribution of X is entirely encoded in the generating series

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k \in \mathbb{Z}} \mathbb{P}_X[k]\, s^k, \quad s \in D(0,1) \text{ (at least)}.$$

$$\mathbb{P}_X[k] = \frac{1}{2i\pi} \oint_\gamma \frac{G(z)}{z^{k+1}}\, dz \quad \text{for any contour } \gamma \text{ cc around } 0.$$

* uniform distribution on $[\![a+1, b]\!]$: $\mathbb{P}_X[k] = \dfrac{1_{a+1 \leq k \leq b}}{b-a}$.

$$\mathbb{E}[X] = \frac{a+b+1}{2}\ ; \quad \mathrm{var}(X) = \frac{(b-a)^2 - 1}{12}$$

$$\mathbb{E}[s^X] = \frac{1}{b-a}\left(\frac{s^{b+1} - s^{a+1}}{s-1}\right). \qquad X \sim \mathcal{U}([\![a+1, b]\!]).$$

* Bernoulli distribution with parameter p: $X \sim \mathcal{B}(p)$. $p \in (0,1)$.

$\mathbb{P}_X[1] = p\ ;\ \mathbb{P}_X[0] = 1-p$.

$\mathbb{E}[X] = p\ ;\ \mathrm{var}(X) = p(1-p)$. $\mathbb{E}[s^X] = 1 + p(s-1)$.

* geometric distribution with parameter p. $(X_n)_{n \geq 1}$ sequence of independent Bernoulli random variables $\mathcal{B}(p)$.

$$G = \inf(\{n \geq 1 \mid X_n = 1\}).$$

$$\mathbb{P}[G \geq k] = \mathbb{P}[X_1 = X_2 = \cdots = X_{k-1} = 0]$$
$$= \prod_{i=1}^{k-1} \mathbb{P}[X_i = 0] = (1-p)^{k-1}.$$

independance, to be introduced later

So $\forall k \geq 1$, $\mathbb{P}_G[k] = \mathbb{P}[\{G \geq k\} / \{G \geq k+1\}] = p(1-p)^{k-1}$.

$$\mathbb{E}[G] = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p\left(\frac{1}{1-z}\right)'\Big|_{z=1-p} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

$$\mathrm{var}(G) = \frac{1}{p}\left(\frac{1}{p} - 1\right)$$

$$\mathbb{E}[s^G] = \sum_{k=1}^{\infty} p(1-p)^{k-1} s^k = \frac{ps}{1-(1-p)s}.$$

* Poisson distribution with parameter λ. $X \sim \text{Poi}(\lambda)$.

$$\mathbb{P}_X[k \geq 0] = e^{-\lambda} \frac{\lambda^k}{k!}.$$ (to be explained later...).

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{j=1}^{\infty} e^{-\lambda} \frac{\lambda^{j-1}}{j-1!} = \lambda.$$

$$\mathbb{E}[X(X-1)] = \lambda^2; \quad \mathbb{E}[X^2] = \lambda^2 + \lambda; \quad \text{var}(X) = \lambda.$$

$$\mathbb{E}[s^X] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k s^k}{k!} = e^{\lambda(s-1)}.$$

<u>exercise</u>: how to recover $\mathbb{E}[X]$ and $\text{var}(X)$ from $\mathbb{E}[s^X]$?

---

② continuous random variables with values in $\mathbb{R}$

To describe a more general law of a real random variable, one can use:

<u>Definition</u> The cumulative distribution function of a random variable X is

$$F_X : \mathbb{R} \to [0,1]$$
$$t \mapsto F_X(t) = \mathbb{P}[X \leq t].$$

It determines entirely the law $\mathbb{P}_X$, because $\mathcal{C} = \{(-\infty, t]\}$ is an algebra of subsets which "generates the" σ-field $\mathcal{B}(\mathbb{R})$.

<u>Proposition</u>: The following are equivalent:

① $F : \mathbb{R} \to [0,1]$ is the CMF of a random variable.

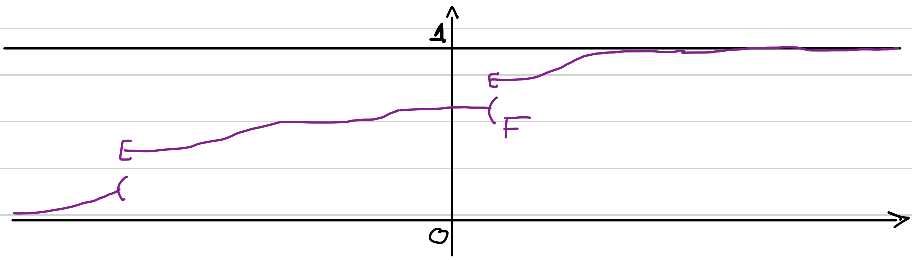
② F is nondecreasing, right continuous, with $\lim_{t \to +\infty} F(t) = 1$; $\lim_{t \to -\infty} F(t) = 0$.

Proof: ① $\Rightarrow$ ②. If $s \le t$, then $\{X \le s\} \subset \{X \le t\}$, so
$$F_X(s) = \mathbb{P}[X \le s] \le \mathbb{P}[X \le t] = F_X(t).$$

If $t_n \downarrow t$ as $n \to +\infty$, then $\{X \le t\} = \bigcap_{n \in \mathbb{N}} \{X \le t_n\}$, so by monotone convergen[ce]
$$\mathbb{P}[X \le t] = \lim_{n \to +\infty} \mathbb{P}[X \le t_n] \Rightarrow \text{right continuity}.$$

Finally,
$$\lim_{t \to +\infty} F(t) \underset{(\text{non decreasing})}{=} \lim_{n \to +\infty} F(n) = \lim_{n \to +\infty} \mathbb{P}[X \le n]$$
$$= \mathbb{P}\Big[\bigcup_{n \in \mathbb{N}}\uparrow \{X \le n\}\Big] = \mathbb{P}[\Omega] = 1.$$

$$\lim_{t \to -\infty} F(t) = \lim_{n \to -\infty} F(n) = \mathbb{P}\Big[\bigcap_{n \in -\mathbb{N}}\downarrow \{X \le n\}\Big] = \mathbb{P}[\emptyset] = 0.$$



② $\Rightarrow$ ① We define the pseudo-inverse:
$$F^{-1}(u) = \inf(\underbrace{\{t \in \mathbb{R} \mid F(t) \ge u\}}_{A_u}) \quad \text{for } u \in (0,1).$$

Since F is non decreasing, $\lim_{t \to \pm\infty} F(t) = 0/1$, A_u is a right interval of the form $(F^{-1}(u), +\infty)$ or $[F^{-1}(u), +\infty)$.

Since F is right continuous, $F^{-1}(u) \in A_u$, so $A_u = [F^{-1}(u), +\infty)$.

Consider the probability space $((0,1), \mathcal{B}(0,1))$, Lebesgue measure).
The law of $U \in (\Omega, \mathcal{F}, \mathbb{P})$ is uniform on $(0,1)$.
Set $X = F^{-1}(U)$. F is measurable $(0,1) \to \mathbb{R}$.

$$\mathbb{P}[X\le t] = \mathbb{P}[t\in A_U] = \mathbb{P}[F(t)\ge U] = \int_0^{F(t)} dx = F(t). \quad \square$$

In many interesting cases, F_X is actually continuous and locally Lipschitz. Its derivative $f_X = F_X'$ is defined almost everywhere on $\mathbb{R}$, and we have:

$$F_X(t) = \int_{-\infty}^{t} f_X(s)\, ds.$$

The law $\mathbb{P}_X$ is then absolutely continuous with respect to the Lebesgue measure, with Radon-Nikodym derivative $\frac{d\mathbb{P}_X}{d\mathrm{Leb}}(s) = f_X(s) \ge 0$.

We have $\int_{\mathbb{R}} f_X(s)\, ds = 1$. f_X is the density of the variable X.

* uniform (continuous) distribution on $[a,b]$: $X \sim U([a,b])$

$$f_X(s) = \frac{\mathbb{1}_{s\in[a,b]}}{b-a}.$$

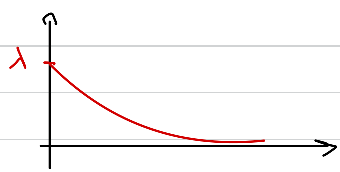
$$F_X(t) = \begin{cases} 0 & \text{if } t\le a, \\ \frac{t-a}{b-a} & \text{if } a<t\le b, \\ 1 & \text{if } b<t. \end{cases}$$

$$\mathbb{E}[X] = \frac{a+b}{2}$$

$$\mathrm{var}(X) = \frac{(b-a)^2}{12}.$$

* exponential distribution with parameter $\lambda>0$: $X\sim \mathrm{Exp}(\lambda)$.

$$f_X(s) = \mathbb{1}_{s>0}\, \lambda e^{-\lambda s}.$$



$$F_X(t) = \begin{cases} 0 & \text{if } t\le 0, \\ 1-e^{-\lambda t} & \text{if } 0<t. \end{cases}$$

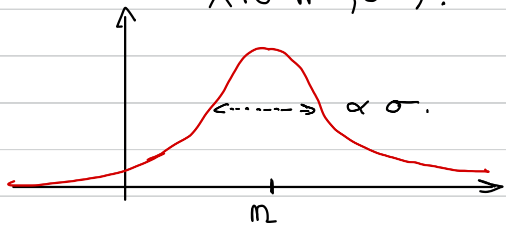
$$\mathbb{E}[X] = \frac{1}{\lambda}$$

$$\mathrm{var}(X) = \frac{1}{\lambda^2}.$$

* Gaussian (or normal) law with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$.

$X \sim N(m, \sigma^2)$.

$$f_X(s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-m)^2}{2\sigma^2}}$$



$\mathbb{E}[X] = m$; $\mathrm{var}(X) = \sigma^2$

(by using change of variables ; one only needs to know that $\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$).

These variables are deeply connected with another description of probability measures on $\mathbb{R}$:

<u>Definition</u> Let μ be a probability measure on $\mathbb{R}$. Its characteristic function or Fourier transform is the continuous function

$$\hat{\mu}(\xi) = \int_{\mathbb{R}} e^{i\xi x} \mu(dx) = \mathbb{E}[e^{i\xi X}] \text{ if } X \sim \mu.$$

<u>example</u>: If $X \sim N(m, \sigma^2)$ then

$$\mathbb{E}[e^{i\xi X}] = \int_{\mathbb{R}} e^{i\xi x - \frac{(x-m)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} \underset{y = \frac{x-m}{\sigma},\ x = m + \sigma y}{=} e^{i\xi m} \int_{\mathbb{R}} e^{i\xi\sigma y - \frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}$$

The map $g(\xi) = \int e^{i\xi y - \frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}$ is smooth, with $g(0) = 1$ and

$$g'(\xi) = \int iy\, e^{i\xi y - \frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} = \underbrace{\left(i \int (y - i\xi) e^{-(\frac{y^2}{2} - i\xi y)} \frac{dy}{\sqrt{2\pi}} \right)}_{i\left[e^{-\frac{y^2}{2} - i\xi y}\right]_{-\infty}^{\infty} = 0} - \xi g(\xi) = -\xi g(\xi).$$

$$\Rightarrow g(\xi) = e^{-\frac{\xi^2}{2}}.$$

So, $\mathbb{E}[e^{i\xi X}] = e^{i\xi m - \frac{\sigma^2 \xi^2}{2}}$.

<u>Proposition</u> The characteristic function $\Phi_X(\xi) = \mathbb{E}[e^{i\xi X}]$ entirely determines the law $\mathbb{P}_X$.

We shall prove this during Chapter III.

4. Independent σ-algebras and random variables.

A common assumption on events A, B in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is their independence:

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\, \mathbb{P}[B].$$

(the simplest formula that one can think of to simplify things).

It is convenient to extend this definition to families of events:

<u>Definition</u> Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
$(\mathcal{F}_i)_{i \in I}$ be a family of σ-subalgebras of $\mathcal{F}$:
$\mathcal{F}_i \subset \mathcal{F}$; $\mathcal{F}_i$ stable by complementation and countable union and intersection.

The σ-algebras $\mathcal{F}_i$ are independent if, for any $i_1 \neq i_2 \neq \cdots \neq i_n \in I$ and for any events $A_{i_j} \in \mathcal{F}_{i_j}$,

$$\mathbb{P}\left[\bigcap_{j=1}^{n} A_{i_j}\right] = \prod_{j=1}^{n} \mathbb{P}[A_{i_j}].$$

<u>example</u>: $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{F}_i, \bigotimes_{i \in I} \mathbb{P}_i\right)$ product of probability spaces.

Each $\overline{\mathcal{F}_i} = \pi_i^{-1}(\mathcal{F}_i) = \left\{ \underbrace{A_i \times \prod_{j \neq i} \Omega_j}_{= \overline{A_i}} \right\}$ is a σ-subalgebra of $\mathcal{F}$

The $\overline{\mathcal{F}_i}$ are independent:

$$\mathbb{P}\left[\bigcap_{j=1}^{n} \overline{A_{i_j}}\right] = \mathbb{P}\left[A_{i_1} \times A_{i_2} \times \cdots \times A_{i_n} \times \prod_{k \notin \{i_1, \ldots, i_n\}} \Omega_k\right]$$

$$\underset{(\text{definition of } \mathbb{P})}{=} \prod_{j=1}^{n} \mathbb{P}_j[A_{i_j}] = \prod_{j=1}^{n} \mathbb{P}[\overline{A_{i_j}}]$$

example: If two events A and B are independent, then the corresponding σ-algebras

$$\mathcal{A} = \sigma(A) = (\phi, A, A^c, \Omega)$$
$$\mathcal{B} = \sigma(B) = (\phi, B, B^c, \Omega)$$ are independent.

Definition A family of random variables $(X_i)_{i \in I}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists in independent random variables if the corresponding σ-algebras $\mathcal{F}_i = \sigma(X_i) \subset \mathcal{F}$ are independent.

Note that by definition, an infinite family of σ-algebras or of random variables is independent if and only if any finite subfamily is independent. Therefore, in the sequel we focus on finite families.

Theorem Given random variables $X_1, \dots, X_n$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with values in $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$, the following are equivalent:

1) $X_1, \dots, X_n$ are independent.

2) The law $\mathbb{P}_X$ of $X = (X_1, \dots, X_n) \in (E_1 \times \dots \times E_n, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n)$ is
$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$.

3) For any bounded measurable functions $f_1 : E_1 \to \mathbb{R}, \dots, f_n : E_n \to \mathbb{R}$
$$\mathbb{E}[f_1(X_1) \dots f_n(X_n)] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)]$$

4) (if $E_1 = \dots = E_n = \mathbb{R}$) The joint characteristic function factorises:
$$\mathbb{E}[e^{i(\xi_1 X_1 + \dots + \xi_n X_n)}] = \prod_{j=1}^{n} \mathbb{E}[e^{i\xi_j X_j}] \quad \forall (\xi_1, \dots \xi_n) \in \mathbb{R}^n.$$

Proof: 1) $\Leftrightarrow$ 2). For any measurable subsets $A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n$,

$$\mathbb{P}_X\Big[\prod_{i=1}^n A_i\Big] = \mathbb{P}\Big[\bigcap_{i=1}^n \{X_i \in A_i\}\Big]$$

$$\underset{\text{(independence assumption)}}{=} \prod_{i=1}^n \mathbb{P}[X_i \in A_i] = \prod_{i=1}^n \mathbb{P}_{X_i}[A_i].$$

$$= \Big(\bigotimes_{i=1}^n \mathbb{P}_{X_i}\Big)\Big[\prod_{i=1}^n A_i\Big].$$

2) $\Rightarrow$ 3). This is the Fubini theorem applied to the product of measures $\bigotimes_{i=1}^n \mathbb{P}_{X_i}$.

3) $\Rightarrow$ 4) is obvious (split each $e^{i\xi_j X_j}$ in $\cos \xi_j X_j + i \sin \xi_j X_j$ to get measurable bounded functions).

4) $\Rightarrow$ 2). We admit the following extension of the already admitted proposition:

A law on $\mathbb{R}^n$ μ is entirely determined by its multidimensional Fourier transform:

$$\hat{\mu}(\vec{\xi} \in \mathbb{R}^n) = \int_{\mathbb{R}^n} e^{i\langle \vec{\xi} | \vec{x} \rangle} \mu(d\vec{x}).$$

see the proof later.

Then, 4) says that $\mathbb{P}_X$ and $\bigotimes_{i=1}^n \mathbb{P}_{X_i}$ have the same Fourier transform, so they are equal.

$\square$

Several basic and fundamental results from probability theory involve sequences of independent events and random variables, and sums of independent random variables. Let us state three preliminary results.

① Borel-Cantelli

$(\Omega, \mathcal{F}, \mathbb{P})$ probability space.
$(A_n)_{n\in\mathbb{N}}$ sequence of events.

$$A = \limsup_{n\to+\infty} A_n = \bigcap_{n=0}^{+\infty} \left(\bigcup_{k\geq n} A_k \right) = \{\text{an infinity of } A_k \text{ occur}\}.$$

Can one relate $\mathbb{P}[A]$ to the $\mathbb{P}[A_n]$?

- if $\sum_{n\in\mathbb{N}} \mathbb{P}[A_n] < +\infty$, then $\mathbb{P}[A] = 0$.
- if the A_n are independent and $\sum_{n\in\mathbb{N}} \mathbb{P}[A_n] = +\infty$, then $\mathbb{P}[A] = 1$.

__Proof__: For the first case, since the series converge, $\forall\varepsilon>0$, $\sum_{k\geq n} \mathbb{P}[A_k] \leq \varepsilon$ for n large enough. Then,

$$\mathbb{P}[A] \leq \mathbb{P}\left[\bigcup_{k\geq n} A_k\right] \leq \sum_{k\geq n} \mathbb{P}[A_k] \leq \varepsilon. \Big) \text{ true for any } \varepsilon, \text{ so } \mathbb{P}[A] = 0.$$

In the second case, by complementation,

$$\mathbb{P}[A^c] = \mathbb{P}\left[\bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k^c\right]$$

$$= \lim_{n\to+\infty} \mathbb{P}\left[\bigcap_{k\geq n} A_k^c\right] \underset{\text{independence}}{=} \lim_{n\to+\infty} \left(\prod_{k\geq n} 1 - \mathbb{P}[A_k] \right)$$

$$\leq \liminf_{n\to+\infty} \exp\left(-\sum_{k\geq n} \mathbb{P}[A_k]\right) = 0. \qquad \square.$$

__example__: (there is an infinity of successes in a sequence of independent Bernoulli experiments with parameter $p_{n\geq1}$) has probability 0 or 1 according to the value of $\sum_{n\geq1} p_n$.

② Kolmogorov 0-1 law.

$(\Omega, \mathcal{F}, \mathbb{P})$ probability space, $(X_n)_{n\in\mathbb{N}}$ sequence of random variables defined on Ω

$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$; we have the inclusions of σ-algebras
$= \{$"events measurable at time n"$\}$.
$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n \subset \dots \subset \mathcal{F}$
(a filtration of the probability space, more on this later...)

$\mathcal{G}_n = \sigma(X_{n+1}, X_{n+2}, \dots, X_{n+m}, \dots)$; $\mathcal{G}_0 \supset \mathcal{G}_1 \supset \dots \supset \mathcal{G}_n \supset \dots$

Lemma: If the X_n are independent, then $\mathcal{F}_n$ and $\mathcal{G}_n$ are independent σ-algebras.

Proof: Fix an event $G = \{X_{n+j_1} \in B_{j_1}, \dots X_{n+j_m} \in B_{j_m}\} \in \mathcal{G}_n$.

By assumption of independence, $\forall F = \{X_{i_1} \in A_{i_1}, \dots X_{i_\ell} \in A_{i_\ell}\} \in \mathcal{F}_n$ with the $i_1, \dots i_\ell \leq n$,

$$\mathbb{P}[F \cap G] = \mathbb{P}[X_{i_1} \in A_{i_1}] \dots \mathbb{P}[X_{i_\ell} \in A_{i_\ell}]\, \mathbb{P}[X_{n+j_1} \in B_{j_1}] \dots \mathbb{P}[X_{n+j_m} \in B_{j_m}]$$
$$= \mathbb{P}[F]\, \mathbb{P}[G]$$

The class of subsets $F \in \mathcal{F}_n$ such that $\mathbb{P}[F \cap G] = \mathbb{P}[F]\,\mathbb{P}[G]$ is a monotone class containing an algebra $\mathcal{C}$ with $\sigma(\mathcal{C}) = \mathcal{F}_n$, so:

$$\forall F \in \mathcal{F}_n, \ \mathbb{P}[F \cap G] = \mathbb{P}[F]\,\mathbb{P}[G] \qquad \text{prop}(G).$$

Then, the class of subsets $G \in \mathcal{G}_n$ with property prop(G) is a monotone class ...

$$\Rightarrow \forall F \in \mathcal{F}_n, \ \forall G \in \mathcal{G}_n, \ \mathbb{P}[F \cap G] = \mathbb{P}[F]\,\mathbb{P}[G] \quad \square$$

We consider in this setting the tail or terminal σ-algebra:

$$\mathcal{G} = \bigcap_{n\in\mathbb{N}} \mathcal{G}_n.$$

Theorem: If the $(X_n)_{n\in\mathbb{N}}$ are independent, then $\mathcal{G}$ is independent of itself.

$$\Longrightarrow \forall G \in \mathcal{G}, \quad \mathbb{P}[G] \in \{0,1\}.$$

<u>Proof</u>: if $G \in \mathcal{G}$, then G is in all $\mathcal{F}_n$, hence independent of any event in a $\mathcal{F}_n$.

Set $\mathcal{F}_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}}^{\uparrow} \mathcal{F}_n\right) = \sigma(X_n, n\in\mathbb{N})$.

By the same monotone class argument as before, G is independent of any event in $\mathcal{F}_\infty$.

But $\mathcal{F}_\infty \supset \mathcal{G}$, so G is independent of any event in $\mathcal{G}$, including itself.

Then, $\mathbb{P}[G] = \mathbb{P}[G\cap G] = (\mathbb{P}[G])^2$ $\square$.

③ <u>convolution of probability measures on $\mathbb{R}$</u>

Consider two independent random real random variables X, Y with laws $\mathbb{P}_X, \mathbb{P}_Y$.
By assumption, the law $\mathbb{P}_{(X,Y)}$ is $\mathbb{P}_X \otimes \mathbb{P}_Y$ on $\mathbb{R}^2$, so the sum $S = X+Y$ has law $\mathbb{P}_S = \Sigma_*(\mathbb{P}_X \otimes \mathbb{P}_Y)$ entirely determined by $\mathbb{P}_X$ and $\mathbb{P}_Y$.

$$\Sigma : \mathbb{R}^2 \to \mathbb{R}$$
$$(x,y) \mapsto x+y.$$

Formula for $\mathbb{P}_S$?

- if X and Y are discrete with values in $\mathbb{Z}$, then $S \in \mathbb{Z}$ a.s., and:

$$\mathbb{P}[S=k] = \sum_{\ell\in\mathbb{Z}} \mathbb{P}[X=\ell \text{ et } S=k]$$
$$= \sum_{\ell\in\mathbb{Z}} \mathbb{P}[X=\ell \text{ et } Y=k-\ell]$$
$$= \sum_{\ell\in\mathbb{Z}} \mathbb{P}_X[\ell]\,\mathbb{P}_Y[k-\ell].$$

In particular, if $X, Y \in \mathbb{N}$, then $\mathbb{P}_S[k] = \sum_{\ell=0}^{k} \mathbb{P}_X[\ell]\,\mathbb{P}_Y[k-\ell]$.

For instance, if $X \perp Y$ and $X \sim Poi(\lambda)$, $Y \sim Poi(\mu)$, then

$$\mathbb{P}_S[k] = e^{-(\lambda+\mu)} \sum_{\ell=0}^{k} \frac{\lambda^\ell}{\ell!} \frac{\mu^{k-\ell}}{(k-\ell)!} = \frac{e^{-(\lambda+\mu)}}{k!} (\lambda+\mu)^k$$

$$\Rightarrow S \sim Poi(\lambda+\mu).$$

- if X and Y have densities f_X and f_Y, then for any positive measurable function $g : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} g(s)\, \mathbb{P}_S[ds] = \mathbb{E}[g(X+Y)] = \int_{\mathbb{R}^2} g(x+y)\, f_X(x) f_Y(y)\, dx\, dy$$

$$\underset{\text{(change of variables } s=x+y)}{=} \int_{(s,x)\in\mathbb{R}^2} g(s)\, f_X(x) f_Y(s-x)\, dx\, ds$$

$$\underset{\text{(Fubini-Tonelli)}}{=} \int_{s\in\mathbb{R}} g(s) \underbrace{\left(\int f_X(x) f_Y(s-x)\, dx\right)}_{\text{almost everywhere convergent, measurable w.r.t } s} ds$$

Taking $g = 1_{\text{Borel subset}}$, we see that $\mathbb{P}_S$ is the probability measure on $\mathbb{R}$ with density

$$f_S(s) = \int_{\mathbb{R}} f_X(x) f_Y(s-x)\, dx.$$

For instance, if $X \perp Y$ and $X \sim \mathcal{N}(m_1, \sigma_1^2)$, $Y \sim \mathcal{N}(m_2, \sigma_2^2)$, then $X+Y \sim \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$.