Discrete Processes

Nicolas Forien

September 8, 2026

These notes are based on notes by Pierre Cardaliaguet, Julien Claisse, Béatrice De Tillière, Paul Gassiat, Massimiliano Gubinelli, Bénédicte Haas, François Huveneers, Jean-François Le Gall and François Simenhaus.

Please feel free to point out any errors!

Evaluation

There are a mid-term exam (P) and a terminal exam (E). The final mark is given by the formula $\max(0.4P + 0.6E,\, E)$.

Notation

In these notes we use the notation $A \subset B$ to mean that A is included or equal to B.

Beware that in English, x positive means $x > 0$, while $x \geq 0$ is called x non-negative.

Similarly, a function f is called increasing if $x < y$ implies $f(x) < f(y)$ (what in French is called "strictement croissante"), while f non-decreasing means that $x < y$ implies $f(x) \leq f(y)$.

Bonus

The parts presented in this kind of "bonus" boxes are not part of the material that you are supposed to master for the exam, but are just here in case you are curious about complements.

These parts can be skipped, and you may read them only if you have time and curiosity, and/or if you feel the need to go a bit deeper into a specific notion.

Contents

Chapter I

Reminders of probability

I.1 Measure theory and integration

I.1.a Measurable sets

Definition 1. *A set $\mathcal{E} \subset \mathcal{P}(E)$ is a **σ-algebra** (also called a **σ-field**) on a set E if it satisfies the three following properties:*

- *The whole set belongs to it: $E \in \mathcal{E}$;*
- *For every $A \in \mathcal{E}$, we have $E \setminus A \in \mathcal{E}$;*
- *If $(A_n)_{n\in\mathbb{N}}$ is a countable family of elements of $\mathcal{E}$, then $\cup_{n\in\mathbb{N}} A_n \in \mathcal{E}$.*

*The sets in $\mathcal{E}$ are called the **$\mathcal{E}$-measurable sets** (or simply the measurable sets, when there is no ambiguity on the σ-field considered), and the pair $(E, \mathcal{E})$ is called a **measurable space**.*

Example 1. *If E is any set, then $\{\varnothing, E\}$ and $\mathcal{P}(E)$ are σ-algebras on E.*

Proposition 1. *An intersection (not necessarily countable) of σ-algebras is also a σ-algebra.*

This property allows to define:

Definition 2. *If $\mathcal{F} \subset \mathcal{P}(E)$, then the **$\sigma$-algebra generated by** $\mathcal{F}$ is defined as*

$$\sigma(\mathcal{F}) \;=\; \bigcap_{\mathcal{E}\ \sigma\text{-algebra}\ :\ \mathcal{E}\supset\mathcal{F}} \mathcal{E}\,.$$

It is the smallest σ-algebra which contains $\mathcal{F}$.

Definition 3. *If E is a topological space, with family of open sets $\mathcal{O}$, the **Borel σ-algebra**, denoted $\mathcal{B}(E)$, is $\sigma(\mathcal{O})$. In the sequel, when we work on $\mathbb{R}$ or more generally $\mathbb{R}^d$, they will always be equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$.*

Definition 4. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A map $f : E \to F$ is called **measurable** if for every set $A \in \mathcal{F}$ we have $f^{-1}(A) \in \mathcal{E}$.*

I.1.b Measures

Definition 5. *A **measure** on a measurable space $(E, \mathcal{E})$ is a map $\mu : \mathcal{E} \to [0, +\infty]$ such that $\mu(\varnothing) = 0$ and if $(A_n)_{n\in\mathbb{N}}$ is a countable family of pairwise disjoint elements of $\mathcal{E}$, then*

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n)\,.$$

*Note that the above sum may be equal to infinity. The triplet $(E, \mathcal{E}, \mu)$ is then called a **measured space**.*

Example 2. *On any measurable space $(E, \mathcal{E})$, we can define the counting measure $\mu : A \in \mathcal{E} \mapsto \mathrm{card}(A)$.*

If $x \in E$, the Dirac mass at x is the measure $\delta_x : A \in \mathcal{E} \mapsto 1_{x\in A}$.

*A more interesting example is the **Lebesgue measure** on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which is the unique measure λ on $\mathcal{B}(\mathbb{R})$ such that $\lambda([a, b]) = b - a$ for any $a \le b$ (its existence is not a trivial fact !).*

Definition 6. *Let μ be a measure on a measurable space $(E, \mathcal{E})$. We say that the measure μ is **finite** if $\mu(E) < \infty$. We say that μ is **sigma-finite** if there exists a countable family $(E_n)_{n\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ such that $\mu(E_n) < \infty$ for every $n \in \mathbb{N}$ and $E = \cup_{n\in\mathbb{N}} E_n$.*

If $(E, \mathcal{E}, \mu)$ is a measured space, we say that a property holds for μ-almost every $x \in E$ if there exists a set $N \in \mathcal{E}$ such that $\mu(N) = 0$ and the property holds for every $x \in E \setminus N$.

A useful technical tool to prove equality of two measures is Proposition 2 below, which is a consequence of the monotone class lemma, that we present now.

Bonus

Definition 7. *A **monotone class** is a subset $\mathcal{M} \subset \mathcal{P}(E)$ such that:*

- *The whole set belongs to it: $E \in \mathcal{M}$;*
- *For every $A, B \in \mathcal{M}$, if $A \subset B$ then $B \setminus A \in \mathcal{M}$;*
- *It is stable by non-decreasing countable union, that is to say, if $(A_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of elements in $\mathcal{M}$ (i.e., such that $A_n \subset A_{n+1}$ for every $n \in \mathbb{N}$), then $\cup_{n\in\mathbb{N}} A_n \in \mathcal{M}$.*

Exercise 1. *Can the second assumption in the above definition be replaced with "for every set $A \in \mathcal{M}$, we have $E \setminus A \in \mathcal{M}$"? Prove that it gives an equivalent definition, or build a counter-example.*

Note that an intersection of monotone classes is also a monotone class, which allows to define:

Definition 8. *If $\mathcal{F} \subset \mathcal{P}(E)$, the* ***monotone class generated by*** *$\mathcal{F}$ is defined as*

$$\mathcal{M}(\mathcal{F}) = \bigcap_{\mathcal{M} \text{ monotone class } : \, \mathcal{M} \supset \mathcal{F}} \mathcal{M}.$$

It is the smallest monotone class which contains $\mathcal{F}$.

We can now state the monotone class lemma:

Theorem 1 (Monotone class lemma)**.** *If $\mathcal{C} \subset \mathcal{P}(E)$ is stable by finite intersections (i.e., if for every A, $B \in \mathcal{C}$ we have $A \cap B \in \mathcal{C}$), then $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

Note that a σ-algebra is also a monotone class, so $\sigma(\mathcal{C})$ is a monotone class, so the inclusion $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$ comes for free from the definition. Hence, the interesting part of this theorem is that if $\mathcal{C}$ is stable by finite intersections, then $\mathcal{M}(\mathcal{C})$ is also a σ-field, whence the inclusion $\sigma(\mathcal{C}) \subset \mathcal{M}(\mathcal{C})$.

We will rarely use directly the monotone class theorem, but we often use the following important consequence, to show equality of two measures.

Proposition 2. *Let μ, ν be measures on a measurable space $(E, \mathcal{E})$. Assume that $\mathcal{C} \subset \mathcal{E}$ is stable under finite intersections, satisfies $\sigma(\mathcal{C}) = \mathcal{E}$ and that $\mu(A) = \nu(A)$ for all $A \in \mathcal{C}$.*

Assume moreover that $\mu(E) = \nu(E) < \infty$ (note that this condition is always satisfied if μ and ν are two probability measures). Then, we have $\mu = \nu$.

The conclusion also holds if, instead of $\mu(E) = \nu(E) < \infty$, we assume that there exists a sequence $(E_n)_{n \in \mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ such that $\cup_{n \in \mathbb{N}} E_n = E$ and $\mu(E_n) = \nu(E_n) < \infty$ for every $n \in \mathbb{N}$.

This allows for instance to prove uniqueness of the Lebesgue measure, or that two finite measures on $\mathbb{R}$ coincide if and only if they agree on sets of the form $(-\infty, a]$, for $a \in \mathbb{R}$.

Exercise 2. *Does the result remain true if instead of $\mu(E) = \nu(E) < \infty$ we only assume that the two measures are finite? or sigma-finite with $\mu(E) = \nu(E) = \infty$?*

I.1.c Integrals

An important use of measures is that they allow to define integrals of measurable functions.

Proposition/Definition 1. *Let $(E, \mathcal{E}, \mu)$ be a measured space. Then there exists a unique way of defining, for every measurable and non-negative function $f : E \to \mathbb{R}_+$, a quantity $\int_E f d\mu \in [0, +\infty]$ (also written $\int_E f(x) d\mu(x)$ or $\int_E f(x) \mu(dx)$ and omitting the set E and/or the measure μ if there is no ambiguity), such that:*

- *If $A \in \mathcal{E}$, then $\int 1_A d\mu = \mu(A)$;*

- *Linearity: for any measurable functions f, $g \geq 0$ and $a, b \in \mathbb{R}_+$, it holds $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$;*
- *Monotonicity: for any measurable functions f, g with $0 \leq f \leq g$, we have $\int f d\mu \leq \int g d\mu$ (note that in fact this point follows from linearity combined with the fact that $\int f d\mu \geq 0$ for every $f \geq 0$).*

Proof. Here is a sketch of proof of existence. First, we define the integral of indicator functions using the first point above. Then, we define the integral of non-negative simple functions (i.e., of finite linear combinations of indicator functions, with non-negative coefficients), using linearity. Lastly, for a general non-negative measurable function f, we define the integral of f as the supremum of the integrals of all non-negative simple functions g which are below f. $\square$

More generally, if f is a measurable function taking values in $\mathbb{R}$, we then say that f is integrable if $\int |f| d\mu < +\infty$, and we then define $\int f d\mu = \int f_+ d\mu - \int f_- d\mu$, where $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$. Similarly, if f takes values in $\mathbb{C}$, we say that f is integrable if its real part and its imaginary part are integrable, and we define the integral of f as $\int f d\mu = \int (\mathrm{Re}\, f) d\mu + i \int (\mathrm{Im}\, f) d\mu$.

Given a measured space $(E, \mathcal{E}, \mu)$ and a measurable $f \geq 0$, we can always define a new measure ν on $(E, \mathcal{E})$ by

$$\nu(A) = \int_E f 1_A \, d\mu \, .$$

We say that f is the density of ν with respect to μ, also written $f = d\nu/d\mu$ or $d\nu = f d\mu$.

The three following limit theorems are very useful.

Theorem 2. ***(Monotone convergence)*** *Let $(f_n)_{n \in \mathbb{N}}$ be a non-decreasing sequence of measurable and non-negative functions, let f be a measurable function, and assume that for almost every $x \in E$, we have $f_n(x) \to f(x)$. Then it holds that*

$$\lim_{n \to \infty} \int f_n d\mu = \int f d\mu \, .$$

(Fatou's lemma) *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of measurable and non-negative functions, then it holds that*

$$\int \left(\liminf_{n \to \infty} f_n \right) d\mu \leq \liminf_{n \to \infty} \int f_n d\mu \, .$$

To recall the direction of the inequality, you can have in mind the example of $f_n = 1_{[n, n+1)}$, with the Lebesgue measure on $\mathbb{R}$.

(Dominated convergence) *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of measurable functions and f be a measurable function such that for μ-almost every $x \in E$, we have $f_n(x) \to f(x)$, and assume that there exists an integrable function g such that for every $n \in \mathbb{N}$, for μ-almost every $x \in E$, we have $|f_n(x)| \leq g(x)$. Then, we have*

$$\lim_{n \to \infty} \int f_n d\mu = \int f d\mu \, .$$

Fatou's lemma can be deduced from the monotone convergence theorem, and the dominated convergence theorem can be proved using Fatou's lemma. It is a good exercise to write these proofs.

Bonus

The two following results follow from the dominated convergence theorem, and are useful to study functions defined using an integral.

Proposition 3 (Continuity theorem)**.** *Let $(E, \mathcal{E}, \mu)$ be a measured space, let (F, d) be a metric space and let $f : E \times F \to \mathbb{R}$ be such that:*

- *For every $y \in F$, the map $x \in E \mapsto f(x, y)$ is $\mathcal{E}$-measurable and μ-integrable (so that $\int_E f(x, y)\, d\mu(x)$ is well defined);*
- *For μ-almost every $x \in E$, the map $y \in F \mapsto f(x, y)$ is continuous on F;*
- *There exists a μ-integrable function $g : E \to \mathbb{R}$ such that for μ-almost every $x \in E$, for all $y \in F$, we have $|f(x, y)| \le g(x)$.*

Then the function $h : y \in F \mapsto \int_E f(x, y)\, d\mu(x)$ is continuous on F.

Proposition 4 (Derivation under the integral)**.** *Let $(E, \mathcal{E}, \mu)$ be a measured space, let I be an interval of $\mathbb{R}$ and let $f : (x, y) \in E \times I \mapsto f(x, y) \in \mathbb{R}$ be such that:*

- *For every $y \in I$, the map $x \in E \mapsto f(x, y)$ is $\mathcal{E}$-measurable and μ-integrable;*
- *For μ-almost every $x \in E$, the map $y \in I \mapsto f(x, y)$ is differentiable on I (we denote by $\partial f/\partial y$ its derivative);*
- *There exists a μ-integrable function $g : E \to \mathbb{R}$ such that for μ-almost every $x \in E$, for all $y \in I$, we have $|\frac{\partial f}{\partial y}(x, y)| \le g(x)$.*

Then the function $h : y \in I \mapsto \int_E f(x, y) d\mu(x)$ is differentiable on I and for every $y \in I$ we have

$$h'(y) \;=\; \int_E \frac{\partial f}{\partial y}(x, y)\, d\mu(x)\,.$$

I.1.d Product spaces

Definition 9. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. We can define the* ***product σ-algebra*** *$\mathcal{E} \otimes \mathcal{F}$, which is a σ-algebra on $E \times F$ given by*

$$\mathcal{E} \otimes \mathcal{F} \;=\; \sigma(\mathcal{E} \times \mathcal{F}) \;=\; \sigma\big(\{A \times B\,,\; A \in \mathcal{E},\, B \in \mathcal{F}\}\big)\,.$$

Definition 10. *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measured spaces such that μ and ν are sigma-*

*finite. The **product measure** $\mu\otimes\nu$ is the unique measure on $(E\times F,\ \mathcal{E}\otimes\mathcal{F})$ such that for all $A\in\mathcal{E}$ and $B\in\mathcal{F}$,*

$$(\mu\otimes\nu)(A\times B) \;=\; \mu(A)\,\nu(B)\,.$$

Theorem 3 (Fubini for non-negative functions)**.** *Let $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},\nu)$ be two measured spaces such that both μ and ν are sigma-finite. Let $f: E\times F\to\mathbb{R}$ be a measurable and non-negative function with respect to $\mathcal{E}\otimes\mathcal{F}$. Then the maps $x\in E\mapsto \int_F f(x,\,y)d\nu(y)$ and $y\in F\mapsto\int_E f(x,\,y)d\mu(x)$ are measurable and*

$$\int_{E\times F} f\,d(\mu\otimes\nu) \;=\; \int_F\left(\int_E f(x,y)\mu(dx)\right)\nu(dy) \;=\; \int_E\left(\int_F f(x,y)\nu(dy)\right)\mu(dx)\,. \tag{I.1}$$

Theorem 4 (Fubini for integrable functions)**.** *Let $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},\nu)$ be two measured spaces such that both μ and ν are sigma-finite. Let $f: E\times F\to\mathbb{R}$ be a measurable function with respect to $\mathcal{E}\otimes\mathcal{F}$. Then the following are equivalent:*

1. *f is integrable with respect to $\mu\otimes\nu$;*
2. *$\int\left(\int|f(x,y)|\,\mu(dx)\right)\nu(dy)<\infty$;*
3. *$\int\left(\int|f(x,y)|\,\nu(dy)\right)\mu(dx)<\infty$,*

and if this holds, then the maps $x\in E\mapsto\int_F f(x,\,y)d\nu(y)$ and $y\in F\mapsto\int_E f(x,\,y)d\mu(x)$ are integrable and (I.1) *holds.*

I.2 Probability: random variables, inequalities, independence

I.2.a Random variables

We now fix a **probability space**, namely a measured space $(\Omega,\ \mathcal{F},\ \mathbb{P})$ where $\mathbb{P}$ is a **probability measure**, i.e., it satisfies $\mathbb{P}(\Omega)=1$.

In this context, familiar objects from measure theory are given new names:

Definition 11. *An **event** is a measurable set $A\in\mathcal{F}$.*

*An event A holds **almost surely** (abbreviated **a.s.**) if $\mathbb{P}(A)=1$.*

*If $(E,\ \mathcal{E})$ is a measurable space, a (E-valued) **random variable** X is a measurable map from $(\Omega,\ \mathcal{F})$ to $(E,\ \mathcal{E})$. When the arrival set E is not specified, this usually means $E=\mathbb{R}$ with the Borel σ-algebra, and in this case the variable is said to be **scalar**).*

*The **expectation** or **mean** of a non-negative or integrable random variable X is $\mathbb{E}[X]=\int_\Omega X(\omega)\mathbb{P}(d\omega)$.*

*An integrable variable is said to be **centered** if its mean is 0.*

*The **variance** of an integrable variable X is*

$$\operatorname{Var}(X) \;=\; \mathbb{E}\big[(X-\mathbb{E}[X])^2\big] \;=\; \mathbb{E}\big[X^2\big]-\mathbb{E}[X]^2 .$$

*The **law** of a E-valued random variable X is the measure image on E, defined for $A \in \mathcal{E}$ by*

$$\mathcal{L}_X(A) \;=\; \mathbb{P}\circ X^{-1}(A) \;=\; \mathbb{P}\big(\{\omega\in\Omega \,:\, X(\omega)\in A\}\big) .$$

Note that in probability theory, the underlying set Ω is typically unimportant (and is often not specified). The important objects are random variables and their properties (such as their laws).

Proposition 5. *Let X be a random variable taking values in some measurable space $(E,\,\mathcal{E})$, let $h : E \to \mathbb{R}$ be measurable and such that $h(X)$ (this notation means $h\circ X$) is integrable. Then*

$$\mathbb{E}\left[h(X)\right] = \int_E h(x)\,\mathcal{L}_X(dx) .$$

For instance, if X is real-valued and its law admits a density f with respect to Lebesgue measure, then

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)\,dx .$$

Given a random variable X taking values in $(E,\,\mathcal{E})$, the **σ-algebra generated by** X, denoted $\sigma(X)$, is

$$\sigma(X) \;=\; \left\{X^{-1}(A),\, A\in\mathcal{E}\right\} .$$

It is the smallest σ-algebra on Ω for which X is measurable, that is to say, if $\mathcal{G}$ is a σ-algebra on Ω then X is a measurable map from $(\Omega,\mathcal{G})$ to $(E,\mathcal{E})$ if and only if $\mathcal{G}\supset\sigma(X)$.

A σ-algebra can be interpreted as a certain quantity of information: in view of this, the σ-algebra $\sigma(X)$ consists of all the information that one can deduce from the value of X.

Proposition 6. *Let X be a random variable taking values in some measurable space $(E,\,\mathcal{E})$ and let Y be a $\sigma(X)$-measurable real random variable, then there exists a measurable map $f : E\to\mathbb{R}$ such that $Y = f(X)$.*

It is a good exercise to prove the above proposition.

Definition 12 (L^p spaces)**.** *Fix $1\le p<\infty$. Given a random variable X, its L^p norm is defined by*

$$\|X\|_{L^p} \;=\; \mathbb{E}\big[|X|^p\big]^{1/p} .$$

We then define the set

$$\mathcal{L}^p(\Omega,\,\mathcal{F},\,\mathbb{P}) \;=\; \big\{X \,:\, \Omega\to\mathbb{R} \text{ measurable such that } \|X\|_{L^p}<\infty\big\} .$$

The L^p norm is not a norm on this set because $\|X\|_{L^p} = 0$ only implies that $X = 0$ almost surely, so we define the space $L^p(\Omega, \mathcal{F}, \mathbb{P})$ which is obtained by taking the quotient by the equivalence relation $X \sim Y$ if and only if $X = Y$ almost surely. Equipped with the L^p norm, it is a Banach space (i.e., a complete normed vector space).

We can also define the L^∞ norm (and the corresponding L^∞ space)

$$\|X\|_{L^\infty} \;=\; \operatorname{ess\,sup} |X| \;=\; \inf\big\{c \in \mathbb{R} \;:\; \mathbb{P}(|X| \le c) = 1\big\}\,.$$

I.2.b Inequalities

We record the following important inequalities for expectations of random variables.

Proposition 7. ***(Jensen)*** *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex and X is a real random variable, then*

$$\phi\left(\mathbb{E}[X]\right) \;\le\; \mathbb{E}\left[\phi(X)\right]$$

(as long as the two expectations above are well defined).

(Cauchy-Schwarz) *For any two real random variables X and Y,*

$$\mathbb{E}\big[|XY|\big] \;\le\; \|X\|_{L^2}\,\|Y\|_{L^2}\;.$$

(Hölder) *Fix $1 \le p,\, q \le \infty$ such that $1/p + 1/q = 1$. Then for any two real random variables X and Y,*

$$\mathbb{E}\big[|XY|\big] \;\le\; \|X\|_{L^p}\,\|Y\|_{L^q}\;.$$

A simple but often very efficient way of measuring probabilities is given by the following proposition:

Proposition 8. *Let $U : \mathbb{R} \to \mathbb{R}$ be a non-decreasing positive function. Then for any real random variable X such that the expectation below makes sense, for any $a \in \mathbb{R}$, it holds that*

$$\mathbb{P}\left(X \ge a\right) \;\le\; \frac{\mathbb{E}[U(X)]}{U(a)}\,.$$

This implies the following classical special cases:

$$\forall\, a > 0 \qquad \mathbb{P}(|X| \ge a) \;\le\; \frac{\mathbb{E}[|X|]}{a} \qquad \text{(Markov)},$$

$$\forall\, a > 0 \qquad \mathbb{P}\big(|X - \mathbb{E}[X]| \ge a\big) \;\le\; \frac{\operatorname{Var}(X)}{a^2} \qquad \text{(Bienaymé-Tchebychev)},$$

$$\forall \lambda > 0 \quad \forall\, a \in \mathbb{R} \qquad \mathbb{P}(X \ge a) \;\le\; \mathbb{E}\big[e^{\lambda X}\big]\, e^{-\lambda a} \qquad \text{(Chernoff)}.$$

I.2.c Independence

Definition 13. *Two events $A, B \in \mathcal{F}$ are* ***independent*** *if*

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

A family $(A_i)_{i\in I}$ of events is independent if for every $n \geq 1$, for any choice of indices $i_1, \ldots, i_n \in I$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) \;=\; \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_n})\,.$$

Similarly, two random variables X and Y taking values respectively in $(E,\, \mathcal{E})$ and $(E',\, \mathcal{E}')$ are independent if for every $A \in \mathcal{E}$ and every $B \in \mathcal{E}'$ we have

$$\mathbb{P}(X \in A,\, Y \in B) \;=\; \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B)\,.$$

A family $(X_i)_{i\in I}$ of random variables taking values in whatever measurable spaces $(E_i,\, \mathcal{E}_i)$ is independent if for every $n \geq 1$, for any choice of indices $i_1, \ldots, i_n \in I$ and of measurable sets $A_1 \in \mathcal{E}_{i_1}, \ldots, A_n \in \mathcal{E}_{i_n}$, it holds that

$$\mathbb{P}(X_{i_1} \in A_1, \ldots, X_{i_n} \in A_n) \;=\; \mathbb{P}(X_{i_1} \in A_1) \cdots \mathbb{P}(X_{i_n} \in A_n)\,.$$

A family $(\mathcal{G}_i)_{i\in I}$ of σ-algebras is independent if any family $(A_i)_{i\in I}$ of events with $A_i \in \mathcal{G}_i$ for every $i \in I$ is independent.

Note that a family of variables is independent if and only if the family of σ-algebras generated by these variables is independent.

Note that independence of a family is stronger than pairwise independence of its elements.

Note that it follows from the definition that two variables X and Y are independent if and only if the law of (X, Y) is the product measure $\mathcal{L}_X \otimes \mathcal{L}_Y$. A similar result is true for families of random variables.

In particular, in conjunction with Proposition 5 and Fubini's theorem, this implies that if X and Y are independent and f and g are two measurable functions such that the expectations below make sense, then

$$\mathbb{E}[f(X)g(Y)] \;=\; \mathbb{E}[f(X)]\mathbb{E}[g(Y)]\,.$$

Given a sequence $(A_n)_{n\in\mathbb{N}}$ of events, we define

$$\limsup_n A_n \;=\; \bigcap_{n\in\mathbb{N}} \left(\bigcup_{k\geq n} A_k \right) \;=\; \big\{\omega \in \Omega \;:\; \omega \text{ belongs to infinitely many events } A_n\big\}\,.$$

and

$$\begin{aligned}\liminf_n A_n \;&=\; \bigcup_{n\in\mathbb{N}} \left(\bigcap_{k\geq n} A_k \right) \\ &=\; \big\{\omega \in \Omega \;:\; \omega \text{ belongs to all events } A_n \text{ except a finite number of them}\big\}\,.\end{aligned}$$

Theorem 5 (Borel-Cantelli's Lemma)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space.*

1. *Let* $(A_n)_{n\in\mathbb{N}}$ *be a sequence of events such that* $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$. *Then* $\mathbb{P}(\limsup A_n) = 0$.
2. *Let* $(A_n)_{n\in\mathbb{N}}$ *be an independent sequence of events such that* $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$. *Then we have* $\mathbb{P}(\limsup A_n) = 1$.

I.3 Convergence of random variables

I.3.a The various modes of convergence

Let X, and X_n, $n \in \mathbb{N}$ be some random variables defined on the same probability space.

Definition 14. *We say that* X_n *converges to* X*:*

in probability *if for every* $\varepsilon > 0$, *we have* $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$.

in L^p *(for a given* $p \geq 1$*) if* $\|X_n - X\|_{L^p} \to 0$.

almost surely *if* $\mathbb{P}(X_n \to X) = 1$.

in law *or* ***in distribution*** *if for every bounded and continuous function* ϕ, *we have* $\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$.

Note that convergence in law is really a property of the laws of the random variables, not of the random variables themselves, unlike the other modes of convergence, and that for convergence in law the variables in fact do not need to be defined on the same probability space.

Proposition 9. *We have the following implications between the different modes of convergence:*

1. *If* $X_n \to X$ *almost surely then* $X_n \to X$ *in probability.*
2. *If* $X_n \to X$ *in* L^p *and* $q \leq p$ *then* $X_n \to X$ *in* L^q.
3. *If* $X_n \to X$ *in* L^1 *then* $X_n \to X$ *in probability.*
4. *If* $X_n \to X$ *in probability then* $X_n \to X$ *in law.*
5. *If* $X_n \to X$ *in* L^∞ *then* $X_n \to X$ *almost surely.*
6. *If* $X_n \to X$ *in law and the variable* X *is constant, then* $X_n \to X$ *in probability.*
7. *If* $X_n \to X$ *in probability, then there exists a subsequence* $(X_{\varphi(n)})_{n\in\mathbb{N}}$ *which converges almost surely to* X.
8. *(Dominated convergence theorem) If* $X_n \to X$ *almost surely and there exists an integrable variable* Z *such that* $|X_n| \leq Z$ *for every* $n \in \mathbb{N}$, *then* $X_n \to Z$ *in* L^1.

I.3.b More about convergence in law

We now spend more time on the convergence in law.

Definition 15. *The **characteristic function** of (the law of) a real random variable X is defined by*

$$\phi_X \ : \ t \in \mathbb{R} \ \mapsto \ \mathbb{E}[e^{itX}] \, .$$

Note that for any random variable X, it is a continuous and bounded function on $\mathbb{R}$.

It can be linked to moments of X in the following way:

Proposition 10. *Assume that $\mathbb{E}|X|^k < \infty$. Then ϕ_X is C^k on $\mathbb{R}$, and $\phi^{(k)}(0) = i^k \mathbb{E}[X^k]$.*

Proof. Exercise (use differentiation under $\mathbb{E}$) □

A main use of characteristic functions comes from the following result.

Theorem 6 ((Lévy's Continuity Theorem))**.** *The following are equivalent:*

(1) $X_n \to X$ in law,

(2) For every $t \in \mathbb{R}$, $\phi_{X_n}(t) \to \phi_X(t)$.

To study convergence in law, we can also use the cumulative distribution function:

Definition 16. *The **cumulative distribution function** of a scalar random variable X is the function $x \in \mathbb{R} \mapsto \mathbb{P}(X \leq x)$. By a remark above, it fully characterizes the law of X.*

Theorem 7. *$X_n \to X$ in law if and only if for every $x \in \mathbb{R}$, if F_X is continuous at x then $F_{X_n}(x) \to F_X(x)$.*

I.4 The Law of Large Numbers and the Central Limit Theorem

I.4.a The two laws of large numbers

In this section, we consider a sequence $(X_n)_{n \geq 1}$ of independent real random variables which all have the same law. This is called an **i.i.d.** sequence (for "independent and identically distributed").

Theorem 8 (Weak Law of Large Numbers)**.** *Assume that $(X_n)_{n \geq 1}$ is an i.i.d. sequence of real integrable variables, and let $m = \mathbb{E}[X_1]$. Then*

$$\frac{X_1 + \cdots + X_n}{n} \overset{n \to \infty}{\longrightarrow} m \qquad \textit{in probability.}$$

Proof. See exercise sessions. □

We also have the stronger law, which has the same hypotheses.

Theorem 9 (Strong Law of Large Numbers)**.** *Under the same hypotheses than in Theorem 8, we have*

$$\frac{X_1 + \cdots + X_n}{n} \overset{n\to\infty}{\longrightarrow} m \qquad \textit{almost surely.}$$

One of the many different proofs of this result is the subject of an exercise in the exercise sheet.

I.4.b The Central Limit Theorem

Definition 17. *The standard Gaussian measure, denoted by $\mathcal{N}(0,1)$, is the probability measure on $\mathbb{R}$ with density function given by $x \mapsto e^{-x^2/2}/\sqrt{2\pi}$.*

Exercise: check that the above is a well-defined probability measure (hint: compute $\int e^{-x^2-y^2}dxdy$ via polar coordinates). Further check that if Z has law $\mathcal{N}(0,1)$, then $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^2] = 1$, and the characteristic function is given by $\phi_Z(t) = e^{\frac{-t^2}{2}}$ (hint: use integration by parts to show that $\phi_Z'(t) = -t\phi_Z(t)$).

Theorem 10 (Central Limit Theorem)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of variables with $\mathbb{E}|X_1|^2 < \infty$, and let $m = \mathbb{E}[X_1]$ and $\sigma^2 = Var(X_1)$. For every $n \geq 1$, we write $S_n = X_1 + \cdots + X_n$. Then we have the convergence in law*

$$\frac{S_n - mn}{\sigma\sqrt{n}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)\,,$$

which shows the desired convergence.

Proof. Left as an exercise (use the characteristic function). □

Chapter II

Conditional expectation

The goal of this chapter is to construct the conditional expectation and to give its main properties. In all what follows, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space.

II.1 Definition

The question we address in this chapter is the following: given a sub-σ-field $\mathcal{G} \subset \mathcal{F}$, we want to estimate a random variable X knowing the (partial) information contained in $\mathcal{G}$. The mathematical concept which corresponds to this idea is the *conditional expectation* of X knowing $\mathcal{G}$, which, intuitively, is the $\mathcal{G}$-measurable random variable which approximates X best.

Before giving the general definition, let us start with some particular cases to motivate the concept.

If $A, B \in \mathcal{F}$ are two events with $\mathbb{P}(B) > 0$, we can define the conditional probability

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

More generally, if $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, we can define the conditional expectation of X knowing B:

$$\mathbb{E}[X|B] \;=\; \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\,.$$

Note that if $X = \mathbf{1}_A$ then we recover $\mathbb{P}(A|B)$.

Now, instead of an event B, we want to generalize this definition and to construct the conditional expectation with respect to a random variable.

First, if $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ is an integrable variable and $Y : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ is a random variable taking values in a finite or countable set E, equipped with the σ-algebra $\mathcal{E} = \mathcal{P}(E)$, then we define

$$\mathbb{E}[X|Y] \;=\; \sum_{y \in E \,:\, \mathbb{P}(Y=y)>0} \mathbb{E}[X|Y=y]\mathbf{1}_{\{Y=y\}} \;=\; \phi(Y)\,,$$

where ϕ is the function given by

$$\phi\ :\ \begin{cases} E \longrightarrow \mathbb{R} \\ y \longmapsto \begin{cases} \mathbb{E}[X|Y=y] & \text{if } \mathbb{P}(Y=y)>0\,, \\ 0 & \text{otherwise.} \end{cases} \end{cases}$$

Note that $\mathbb{E}[X|Y]$ is not a number but a random variable defined on Ω, because its value depends on the value of $Y = Y(\omega)$.

Note that when we write $\mathbb{E}[X|Y] = \phi(Y)$, more formally, this means $\mathbb{E}[X|Y] = \phi \circ Y$, that is to say, for every $\omega \in \Omega$, $\mathbb{E}[X|Y](\omega) = \phi(Y(\omega))$.

The above definition breaks down if, instead of a discrete random variable Y, we consider a continuous variable, for example a real variable which admits a density with respect to Lebesgue's measure, because in this case we have $\mathbb{P}(Y = y) = 0$ for every $y \in \mathbb{R}$, so that $\mathbb{E}[X|Y = y]$ is not well defined.

Before constructing the conditional expectation with respect to a more general random variable, we construct conditional expectation with respect to a σ-algebra. We consider a random variable $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and we let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$.

A σ-field can be seen a certain quantity of information: if $\mathcal{G}$ is the information known by someone, it means that, for every event $A \in \mathcal{G}$, this person knows whether the event A is realized or not, whereas, for events $A \notin \mathcal{G}$, the person does not know whether the event is realized, but can estimate the probability of this event A given the available information.

For example, knowing $\mathcal{G} = \mathcal{P}(\Omega)$ means that you know everything: if, for every subset $A \subset \Omega$, you know whether $\omega \in A$ or not, this means that you know precisely the value of ω, so that this ω is not random anymore.

On the contrary, knowing $\mathcal{G} = \{\varnothing, \Omega\}$ means that you know nothing, because then you only know that $\omega \in \Omega$ and $\omega \notin \varnothing$.

If $\mathcal{G} = \sigma(A) = \{\varnothing, \Omega, A, \Omega \setminus A\}$ for some event $A \in \mathcal{F}$, then knowing $\mathcal{G}$ means that you only know if A is realized or not.

Reminder: if $\mathcal{E} \subset \mathcal{P}(\Omega)$, then

$$\sigma(\mathcal{E}) \ = \bigcap_{\mathcal{G}\ \sigma\text{-algebra}\ :\ \mathcal{G} \supset \mathcal{E}} \mathcal{G}$$

is the σ-field generated by $\mathcal{E}$. It is the smallest σ-field which contains $\mathcal{E}$.

Assume that $(A_n)_{n\in\mathbb{N}}$ is a countable partition of Ω, with $\mathbb{P}(A_n) > 0$ for every $n \in \mathbb{N}$, and consider the σ-algebra $\mathcal{G} = \sigma(\{A_n,\, n \in \mathbb{N}\})$. Note that we then have $\mathcal{G} = \{\cup_{n\in I} A_n,\, I \subset \mathbb{N}\}$. Then, we define the conditional expectation of X with respect to $\mathcal{G}$ as

$$\mathbb{E}[X|\mathcal{G}] \ = \ \sum_{n=0}^{\infty} \mathbb{E}[X|A_n]\, \mathbf{1}_{A_n}\,.$$

This may be rewritten as

$$\mathbb{E}[X|\mathcal{G}] \,:\, \begin{cases} \Omega \longrightarrow \mathbb{R} \\ \omega \in A_n \longmapsto \mathbb{E}[X|A_n]\,. \end{cases}$$

The interpretation is that knowing $\mathcal{G}$ means knowing in which set A_n is ω, and if we know that $\omega \in A_n$ then $\mathbb{E}[X|A_n]$ is the best prediction of X we can make. Note that the above definition also works in the case of a finite partition.

Example 3. *Let us consider the particular example of the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, where* $\Omega = [0,\, 1[$*,* $\mathcal{F}$ *is the Borel* σ*-algebra* $\mathcal{B}([0,\, 1[)$*, and* $\mathbb{P} = \lambda$*, the Lebesgue measure on* $[0,\, 1[$*. In this setting, a random variable is simply a measurable function* $X : [0,\, 1[\to \mathbb{R}$*. Let* $X : [0,\, 1[\to \mathbb{R}$ *be measurable and integrable, and let* $\mathcal{G} = \{\varnothing,\, [0,\, 1[\,,\, [0,\, 1/2[\,,\, [1/2,\, 1[\,\}$*. Then, up to almost sure equality,* $\mathbb{E}[X|\mathcal{G}]$ *is equal to the random variable*

$$Y \,:\, \begin{cases} [0,\, 1[\longrightarrow \mathbb{R} \\ \omega < \dfrac{1}{2} \longmapsto \mathbb{E}\left[X \,\middle|\, \omega < \dfrac{1}{2}\right] \;=\; \dfrac{\int_{[0,\,1[} X\mathbf{1}_{[0,\,1/2[} d\lambda}{\lambda([0,\, 1/2[)} \;=\; 2\displaystyle\int_0^{1/2} X d\lambda\,, \\ \omega \geq \dfrac{1}{2} \longmapsto \mathbb{E}\left[X \,\middle|\, \omega \geq \dfrac{1}{2}\right] \;=\; \dfrac{\int_{[0,\,1[} X\mathbf{1}_{[1/2,\,1[} d\lambda}{\lambda([1/2,\, 1[)} \;=\; 2\displaystyle\int_{1/2}^{1} X d\lambda\,. \end{cases}$$

Bibliography

[1] Jean-Fran¸cois Le Gall. *Measure theory, probability, and stochastic processes*, volume 295 of *Graduate Texts in Mathematics*. Springer, Cham, 2022.