PROBABLITY AND MEASURE

PIERRE RAPHAËL

1. σ-algebra and the monotone class Lemma

Definition 1.1 (σ-algebra). *Let E be a set. A σ-algebra (or tribe) on E is a family $\mathcal{A} \subset \mathcal{P}(E)$ such that*
(i) $E \in \mathcal{A}$
(ii) $A \in \mathcal{A} \Rightarrow A^c = E \backslash A \in \mathcal{A}$
(iii) $A_n \in \mathcal{A} \Rightarrow \cup_{n \geq 0} A_n \in \mathcal{A}$

Definition 1.2 (Generating set). *Let $\mathcal{C} \subset \mathcal{P}(E)$, we let $\sigma(\mathcal{C})$ be the smallest σ-algebra containing $\mathcal{C}$.*

Definition 1.3 (Borel σ-algebra). *If E is a topological space, we let $\mathcal{B}(E)$ be the σ-algebra generated by the open sets of E.*

Definition 1.4 (Product σ-algebra). *Given $(E_1, \mathcal{A}_1)$, $(E_2, \mathcal{A}_2)$, we let*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma\left(\{A_1 \times A_2, A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}\right).$$

Proposition 1.5. *If E_1, E_2 are separable metric spaces, then $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2) = \mathcal{B}(E_1 \times E_2)$.*

Definition 1.6 (Monotone class). *A family $\mathcal{M} \subset \mathcal{P}(E)$ is a monotone class if:*
(i) $E \in \mathcal{M}$
(ii) $A, B \in \mathcal{M}$ and $A \subset B$ imply $B \backslash A \in \mathcal{M}$.
(iii) $A_n \in \mathcal{M}$ increasing family implies $\cup_{n \geq 0} A_n \in \mathcal{M}$.

Definition 1.7. *Let $\mathcal{C} \subset \mathcal{P}(E)$, we let $\mathcal{M}(\mathcal{C})$ be the smallest monotone class containing $\mathcal{C}$.*

Lemma 1.8. *A monotone class is a σ-algebra iff it is stable by finite intersection.*

Proposition 1.9 (Monotone class Lemma). *Let $\mathcal{C} \subset P(E)$ stable by finite intersection, then $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

2. Measure

Definition 2.1 (Measure). *A (positive) measure on $(E, \mathcal{A})$ is a map $\mu : \mathcal{A} \to [0, \infty]$ with:*
(i) $\mu(\emptyset) = 0$
(ii) $A_n \in \mathcal{A}$ two by two disjoint then $\mu(\sqcup_{n \geq 0} A_n) = \sum_{n \geq 0} \mu(A_n)$.

Proposition 2.2. *Let $(E, \mathcal{A}, \mu)$ be a measure space.*
(i) Let A_n be an increasing family $A_n \subset A_{n+1}$ then $\mu\left(\cup_{n \geq 0} A_n\right) = \lim_{n \to \infty} \mu(A_n)$.
(ii) Let B_n be a decreasing family $B_{n+1} \subset B_n$ with $\mu(B_0) < +\infty$ then $\mu\left(\cap_{n \geq 0} B_n\right) = \lim_{n \to \infty} \mu(B_n)$.

Definition 2.3 (Probability measure). *We say μ is finite if $\mu(E) < +\infty$. If $\mu(E) = 1$, we call μ a probability.*

Definition 2.4 (σ-finite)**.** *We say μ is σ-finite if there exists an increasing sequence E_n with $\cup_{n\geq 0} E_n = E$ and $\forall n$, $\mu(E_n) < +\infty$.*

Definition 2.5 (Negligible set)**.** *A set $N \subset E$ is negligible if there exists $A \in \mathcal{A}$ with $\mu(A) = 0$ such that $N \subset A$. A property is said to hold a.e if the set of points $x \in E$ where it does not hold is negligible.*

Definition 2.6 (Completeness)**.** *We say μ is complete if $\mathcal{A}$ contains all the negligible sets.*

Proposition 2.7 (Uniqueness of measures)**.** *Let μ, ν be two measures on $(E, \mathcal{A})$. Assume:*
(i) $\mathcal{A} = \sigma(\mathcal{C})$ with $\mathcal{C}$ stable by finite intersection;
(ii) $\forall A \in \mathcal{C}$, $\mu(A) = \nu(A)$;
then we can conclude that $\mu = \nu$ on the full σ-algebra $\mathcal{A}$ whenever one of the following two conditions holds:
(i) $\mu(E) = \nu(E) < +\infty$
(ii) $\exists$ an increasing sequence $E_n \subset \mathcal{C}$ with $E = \cup_{n\geq 0} E_n$ and $\forall n$, $\mu(E_n) < +\infty$.

3. Measurable functions

Definition 3.1 (Measurable spaces)**.** *Let $(E, \mathcal{A})$, $(F, \mathcal{B})$ be two measurable spaces. A map $f : E \to F$ is measurable if $\forall B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{A}$.*

Proposition 3.2. *The composition of two measurable functions is measurable.*

Proposition 3.3. *Assume $\mathcal{B} = \sigma(\mathcal{C})$, then f is measurable iff $\forall B \in \mathcal{C}$, $f^{-1}(B) \in \mathcal{A}$.*

Proposition 3.4. *A continuous function $f : (E, \mathcal{B}(E)) \to (F, \mathcal{B}(F))$ is measurable.*

Lemma 3.5. *Let $f_1 : (E, \mathcal{A}) \mapsto (F_1, \mathcal{B}_1)$ and $f_2 : (E, \mathcal{A}) \mapsto (F_2, \mathcal{B}_2)$. Then $f : (E, \mathcal{A}) \mapsto (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ defined by $f(x) = (f_1(x), f_2(x))$ is measurable iff each component f_1, f_2 is measurable.*

Lemma 3.6. *Let $f, g : (E, \mathcal{A}) \to (\mathcal{R}, \mathcal{B}(\mathcal{R}))$ be two measurable functions, then the following functions are measurable:*
(i) any linear combination of (f, g);
(ii)fg;
(iii) $f^+ = \max(f, 0)$, $\quad f^- = \max(-f, 0)$.

Definition 3.7 ($\limsup$, $\liminf$)**.** *We let*

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \downarrow \left(\sup_{k\geq n} a_k\right) = \inf_{n\geq 0} \left(\sup_{k\geq n} a_k\right)$$

and

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \uparrow \left(\inf_{k\geq n} a_k\right) = \sup_{n\geq 0} \left(\inf_{k\geq n} a_k\right).$$

Proposition 3.8. *Let f_n be a sequence a measurable functions from $(E, \mathcal{A})$ to $\left(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\right)$, then:*
(i) the functions

$$\sup_{n\geq 0} f_n, \quad \inf_{n\geq 0} f_n, \quad \limsup_{n\to\infty} f_n, \quad \liminf_{n\to\infty} f_n$$

are also measurable. In particular, ($f_n \to f$ pointwise $\Rightarrow f$ measurable);
(ii) the set $\{x \in E;\ \ f_n(x)\ \text{ has a limit as }\ n \to \infty\}$ is measurable.

Definition 3.9 (Measure tranported by an application). *Let $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ be a measurable function and μ a measure on $(E, \mathcal{A})$. The image measure of μ by f, noted $f_*(\mu)$ (sometimes $f(\mu)$) is the measure defined by*

$$\forall B \in \mathcal{B}, \quad f_*\mu(B) = \mu(f^{-1}(B)).$$

4. Integration

4.1. Integration of positive functions.

Definition 4.1 (Simple function). *Let $(E, \mathcal{A}, \mu)$ measure space. A function $f : (E, \mathcal{A}) \to \mathbb{R}$ is called simple if*

$$f = \sum_{j=1}^{N} \alpha_j \mathbf{1}_{A_j}, \quad A_j \in \mathcal{A} \quad and \quad \sqcup_{j=1}^{n} A_j = E$$

for some $\alpha_j \in \mathbb{R}$. If the values α_j are distinct, this is the canonical representation of f.

Definition 4.2 (Integral of simple functions). *Let f simple valued in $\mathbb{R}_+$, we define*

$$\int f d\mu = \sum_{j=1}^{n} \alpha_j \mu(A_j)$$

with the convention: $(\alpha_j = 0, \mu(A_j) = +\infty) \Rightarrow \alpha_j \mu(A_j) = 0$. This number does not depend on the representation of f. We let $\mathcal{E}_+$ be the set of positive (≥ 0) simple functions.

Definition 4.3 (Integral of a positive function). *Let $(E, \mathcal{A}, \mu)$ measure space and $f : (E, \mathcal{A}) \to \overline{\mathbb{R}}_+$ a positive measurable function. We define*

$$\int f d\mu = \sup_{h \in \mathcal{E}_+, h \leq f} \int h d\mu.$$

Lemma 4.4 (Basic properties). *There holds:*
(i) $f \geq g \Rightarrow \int f d\mu \geq \int g d\mu$.
(ii) $\mu(\{x \in E, \quad f(x) > 0\}) = 0 \Rightarrow \int f d\mu = 0$.

Theorem 4.5 (Monotone convergence). *Let $(E, \mathcal{A}, \mu)$ measure space and $f_n : (E, \mathcal{A}) \to \overline{\mathbb{R}}_+$ an* increasing *sequence of* positive *(≥ 0) measurable functions. Let $f = \lim_n \uparrow f_n$, then*

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

Proposition 4.6 (Approximation by simple functions). *Let f be measurable positive, then there exists an increasing sequence f_n of positive simple functiona such that $\forall x \in E$, $\lim_{n\to\infty} f_n(x) = f(x)$.*

Proposition 4.7 (Basic properties). *All functions are assumed to be measurable positive on $(E, \mathcal{A}, \mu)$ measure space.*
(i) $\forall a, b \geq 0$, $\int (af + bg) d\mu = a \int f d\mu + b \in g d\mu$.
(ii) $\int \left(\sum_{n \geq 0} f_n \right) d\mu = \sum_{n \geq 0} \int f_n d\mu$.
(iii) (Markov) $\mu(\{x \in E, f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu$.
(iv) $\int f d\mu < \infty \Rightarrow f < \infty$ *a.e.*
(v) $\int f d\mu = 0 \Rightarrow f = 0$ *a.e.*
(vi) $f = g$ *a.e* $\Rightarrow \int f d\mu = \int g d\mu$.

Theorem 4.8 (Fatou)**.** *Let $(E, \mathcal{A}, \mu)$ measure space and f_n measurable positive, then*

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

4.2. **Integration of real valued functions.**

Definition 4.9 (Integral of real valued functions)**.** *Let $(E, \mathcal{A}, \mu)$ measure space and $f : E \to \mathbb{R}$. We 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 $f^+ = \max\{f, 0\}$, $f^- = \max\{-f, 0\}$. We note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ the space of real valued integrable functions.

Proposition 4.10 (Linearity)**.** *$\mathcal{L}^1(E, \mathcal{A}, \mu)$ is a vector space and $f \to \int f d\mu$ is a linear form.*

Proposition 4.11 (Basic properties)**.** *Let $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$.*
(i) $\left|\int f d\mu\right| \leq \int |f| d\mu$.
(ii) $f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$.
(iii) $f = g$ a.e $\Rightarrow \int f d\mu = \int g d\mu$.

Theorem 4.12 (Lebesgue's dominated convergence)**.** *Let $f_n \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ such that:*
(i) $\exists f$ real valued measurable with $\lim_{n\to\infty} f_n(x) = f(x)$ a.e.;
(ii) $\exists g : E \to \mathbb{R}_+ \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ such that $\forall n$, $|f_n(x)| \leq g(x)$ a.e..
Then $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ and

$$\lim_{n\to\infty} \int |f_n - f| = 0.$$

4.3. **Integral depending on a parameter.**

Theorem 4.13 (Continuity below the integral sign)**.** *Let $(E, \mathcal{A}, \mu)$ a measure space and (U, d) a metric space. Let $f : U \times E \to \mathbb{R}$ and $u_0 \in U$. Assume:*
(i) $\forall u \in U$, the map $x \mapsto f(u, x)$ is measurable;
(ii) a.e $x \in E$, the map $u \mapsto f(x, u)$ is continuous at u_0;
(iii) $\exists g \in \mathcal{L}^1_+(E, \mathcal{A}, \mu)$ such that $\forall u \in U$, $|f(u, x)| \leq g(x)$ a.e..
Then the function $F(u) = \int f(u, x) d\mu$ is well defined for all $u \in U$ and continuous at u_0.

Theorem 4.14 (Derivability below the integral sign)**.** *Let $(E, \mathcal{A}, \mu)$ a measure space and $I \subset \mathbb{R}$ an open interval. Let $f : I \times E \to \mathbb{R}$ and $u_0 \in I$. Assume:*
(i) $\forall u \in I$, the map $x \mapsto f(u, x)$ is in $\mathcal{L}^1(E, \mathcal{A}, \mu)$;
(ii) a.e $x \in E$, the map $u \mapsto f(x, u)$ is differentiable at u_0 with derivative noted $\frac{\partial f}{\partial u}(u_0, x)$;
(iii) $\exists g \in \mathcal{L}^1_+(E, \mathcal{A}, \mu)$ such that $\forall u \in I$, $|f(u, x) - f(u_0, x)| \leq g(x)|u - u_0|$ a.e..
Then the function $F(u) = \int f(u, x) d\mu$ is differentiable at u_0 with

$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x) d\mu.$$

5. **Construction of measures**

Proof is this section are non examinable, but statements are examinable.

5.1. **Lebesgue measure.**

Definition 5.1 (Outer measure). *Let E be a set. A map $\mu^* : \mathcal{P}(E) \to [0, \infty]$ is called an outer measure if:*
(i) $\mu^*(\emptyset) = 0$
(ii) μ^* is increasing: $A \subset B \Rightarrow \mu^*(A) \leq \mu^*(B)$.
(iii) μ^* is σ sub additive: $\forall A_k \subset E$, $\mu^*(\cup_{k\geq 0} A_k) \leq \sum_{k\geq 0} \mu^*(A_k)$.

Definition 5.2 (Measurability). *A set $B \subset E$ is called μ^*-measurable if*

$$\forall A \subset E, \quad \mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c).$$

We let $\mathcal{M}(\mu^*)$ be the set of μ^*-measurable sets.

Theorem 5.3 (Construction of a measure from an outer measure). *$\mathcal{M}(\mu^*)$ is a σ-algebra in E and the restriction of μ^* to $\mathcal{M}(\mu^*)$ is a complete measure.*

Theorem 5.4 (Lebesgue measure on $\mathbb{R}$). *We define*

$$\forall A \subset \mathbb{R}, \quad \lambda^*(A) = \inf \left\{ \sum_{i\in\mathbb{N}} (b_i - a_i), \quad A \subset \cup_{i\in\mathbb{N}}]a_i, b_i[\right\}.$$

Then:
(i) λ^* is an outer measure on $\mathbb{R}$.
(ii) $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}(\lambda^*)$.
(iii) $\forall a \leq b$, $\lambda^*(]a, b[) = \lambda^*([a, b]) = b - a$.
The restriction of λ^* to $\mathcal{M}(\lambda^*)$ is called the Lebesgue measure on $\mathbb{R}$.

Definition 5.5. *Let $(E, \mathcal{A}, \mu)$ be a measure space, we let $\overline{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N})$ where $\mathcal{N}$ are the μ-negligible sets of E.*

Proposition 5.6. *The Lebesgue tribe $\mathcal{M}(\lambda^*)$ is identical to $\overline{\mathcal{B}(\mathbb{R})}$.*

Proposition 5.7. *Let $f : \mathbb{R} \to \mathbb{R}$ Borelian. Let $g : \mathbb{R} \to \mathbb{R}$ such that $g = f$ λ-a.e, then g is measurable for $\overline{\mathcal{B}(\mathbb{R})}$.*

5.2. **Link to Riemann.**

Definition 5.8 ([Riemann integrability). *Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ bounded. Let*

$$\left| \begin{array}{l} S_-(f) = \sup\{\int_I h(x)dx, \quad h \quad stair, \quad h \leq f\} \\ S_+(f) = \inf\{\int_I h(x)dx, \quad h \quad stair, \quad h \geq f\}. \end{array} \right.$$

We say that f is Riemann integrable if

$$S_+(f) = S_-(f)$$

and then the Riemann integral of f is $S(f) = S_\pm(f)$.

Proposition 5.9 (Relation to Riemann). *Let $f : I \to \mathbb{R}$ Riemann integrable, then f is measurable for $\overline{\mathcal{B}(I)}$ and $S(f) = \int_I f d\lambda$.*

5.3. **An example of non measurable set.**

Proposition 5.10. *Let $\mathbb{R}/\mathbb{Q}$ be the set of equivalence class for the relation $x \sim y \Leftrightarrow x - y \in \mathbb{Q}$. Let $F = \{x_a, a \in \mathbb{R}/\mathbb{Q}\}$ where $x_a \in [0, 1]$ is a representant of the equivalence class of a. Then F is not Lebesgue measurable.*

Remark 5.11. The existence of F requires the axiom of choice because $\mathbb{R}/\mathbb{Q}$ is uncountable, and this is necessary to construct Lebesgue non measurable sets.

5.4. **Lebesgue-Stieltjes measures.**

Proposition 5.12. *Let (E,d) be a metric space and μ a* finite *measure on $(E,\mathcal{B}(E))$. Then for all $A\in\mathcal{B}(E)$:*

$$\mu(A)=\inf\{\mu(U),\quad U \text{ open } \ , A\subset U\}=\sup\{\mu(F),\quad F \text{ closed } \ , F\subset A\}$$

Theorem 5.13 (Lebesgue-Stieltjes)**.** *The following hold.*
(i) Let μ be a finite *measure on $(\mathbb{R},\mathcal{B}(\mathbb{R})$, then its distribution function*

$$\forall x\in\mathbb{R},\quad F_\mu(x)=\mu(]-\infty,x])$$

is increasing, bounded, continuous on the right and $\lim_{x\to-\infty}F_\mu(x)=0$.
(ii) Given $F:\mathbb{R}\to\mathbb{R}$ increasing, bounded, continuous on the right and with $\lim_{x\to-\infty}F(x)=0$, then there exists a unique finite Borelian measure μ such that $F_\mu=F$, and then

$$F(b)-F(a)=\mu(]a,b]).$$

6. L^p spaces

Definition 6.1 ($L^p(E,\mathcal{A},\mu), 1\le p<\infty$)**.** *Let $(E,\mathcal{A},\mu)$ be a measure space.*
(i) *Given $p\in[1,\infty[$, we define*

$$\mathcal{L}^p(E,\mathcal{A},\mu)=\{f:E\to\mathbb{R} \ \text{ measurable with } \int|f|^pd\mu<+\infty.\}$$

(ii) *We define $\|f\|_p=\left(\int|f|^pd\mu\right)^{\frac1p}$.*
(iii) *We let $L^p(E,\mathcal{A},\mu)=\mathcal{L}^p(E,\mathcal{A},\mu)/\sim$ for $f\sim g\Leftrightarrow f=g$ μa.e.*

Definition 6.2 ($L^\infty(E,\mathcal{A},\mu), 1\le p<+\infty$)**.** *Let $(E,\mathcal{A},\mu)$ be a measure space.*
(i) *We define*

$$\mathcal{L}^\infty(E,\mathcal{A},\mu)=\{f:E\to\mathbb{R} \ \text{ measurable}, \ \ \exists C\in\mathbb{R}_+,\ \ |f|\le C \ \ \mu-a.e.\}$$

(ii) *We define $\|f\|_\infty=\inf\{C\in\mathbb{R}_+,\ \ |f|\le C\ \ \mu-a.e.\}$.*
(iii) *We let $L^\infty(E,\mathcal{A},\mu)=\mathcal{L}^\infty(E,\mathcal{A},\mu)/\sim$ for $f\sim g\Leftrightarrow f=g$ μa.e.*

Proposition 6.3 (Hölder)**.** *The following hold.*
(i) $\forall 1\le p,p'\le\infty$, $\int|fg|d\mu\le\|f\|_p\|g\|_{p'}$ for $\frac1p+\frac1{p'}=1$.
(ii) $\forall 1\le p,p_i\le\infty$, $\|\Pi_{i=1}^Nf_i\|_p\le\Pi_{i=1}^N\|f_i\|_{p_i}$ for $\frac1p=\sum_{i=1}^N\frac1{p_i}$.
(iii) $\forall 1\le p_1,p_2\le\infty$, $\|f\|_p\le\|f\|_{p_1}^\alpha\|f\|_{p_2}^{1-\alpha}$ for $\frac1p=\frac{\alpha}{p_1}+\frac{1-\alpha}{p_2}$.

Theorem 6.4 (Riesz-Fischer)**.** *For all $1\le p\le+\infty$, $(L^p(E,\mathcal{A},\mu),\|\cdot\|_p)$ is a Banach space (ie a complete normed vector space). For $p=2$, $\left(L^2(E,\mathcal{A},\mu),\langle\cdot,\cdot\rangle_2\right)$ is a Hilbert space (ie a complete space equipped with a scalar product) for the scalar product $\langle f,g\rangle_2=\int_Efgd\mu$.*

Proposition 6.5. *Let $f_n,f\in L^p(E,\mathcal{A},\mu)$ with $\lim_{n\to+\infty}\|f_n-f\|_p=0$. Then there exists a subsequence $\phi(n)$ strictly increasing such that $\lim_{n\to+\infty}f_{\phi(n)}(x)=f(x)$ for μ-a.e $x\in E$.*

Proposition 6.6 (Dense subsets)**.** *The following holds.*
(i) Let $1\le p\le\infty$, then simple functions are dense in $L^p(E,\mathcal{A},\mu)$.
(ii) Let $1\le p<\infty$, then $\mathcal{C}_c^\infty(\mathbb{R}^d)$ is dense in $L^p(R^d,\overline{\mathcal{B}(\mathbb{R}^d)},\lambda)$ where λ is the Lebesgue measure (and this is obviously false for $p=\infty$).

7. Product measure and Fubbini's theorem

7.1. **Product measure and Fubbini.** We let $(E_1, \mathcal{A}_1)$, $(E_2, \mathcal{A}_2)$ be two measurable spaces.

Definition 7.1 (Horizontal and vertical slices of sets). *Let* $B \subset E_1 \times E_2$.
(i) For $x \in E_1$, *the vertical slice is* $B_x = \{y \in E_2, (x, y) \in B\}$.
(i) For $y \in E_2$, *the horizontal slice is* $B^y = \{x \in E_1, (x, y) \in B\}$.

Definition 7.2 (Horizontal and vertical slices of functions). *Let* $f : E_1 \times E_2 \to F$.
(i) For $x \in E_1$, *we let* $f_x = f(x, \cdot) : E_2 \to F$.
(i) For $y \in E_2$, *we let* $f^y = f(\cdot, y) : E_1 \to F$.

Theorem 7.3 (Fubini measurability). *The following hold.*
(i) Pick $B \in \mathcal{A}_1 \otimes \mathcal{A}_2$, *then* $\forall x \in E_1$, $B_x \in \mathcal{A}_2$ *and* $\forall y \in E_2$, $B^y \in \mathcal{A}_1$.
(i) Let $f(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (F, \mathcal{B})$ *measurable, then* $\forall x \in E_1$, $f_x : (E_2, \mathcal{A}_2) \to (F, \mathcal{B})$ *is measurable, and* $\forall y \in E_2$, $f^y : (E_1, \mathcal{A}_1) \to (F, \mathcal{B})$ *is measurable.*

Theorem 7.4 (Construction of the product measure). *Let* μ, ν *be two* σ*-finite measures on resp.* $(E, \mathcal{A})$ *and* $(F, \mathcal{B})$.
(i) There exists a unique measure m *on* $(E \times F, \mathcal{A} \otimes \mathcal{A})$ *such that*

$$\forall A \in \mathcal{A}, \quad \forall B \in \mathcal{B}, \quad m(A \times B) = \mu(A)\nu(B)$$

with the usual convention $0 \cdot \infty = 0$. *This measure is called the product measure and is* σ*-finite, we note it* $\mu \otimes \nu$.
(ii) For all $C \in \mathcal{A} \otimes \mathcal{B}$,

$$\mu \otimes \nu(C) = \int_E \nu(C_x)\mu(dx) = \int_F \mu(C^y)\nu(dy).$$

Theorem 7.5 (Fubini-Tonnelli). *Let* μ, ν *be two* σ*-finite measures on resp.* $(E, \mathcal{A})$ *and* $(F, \mathcal{B})$. *Let* $f : E \times F \to [0, \infty]$ *a measurable function.*
(i) The functions $x \in E \mapsto \int_F f(x, y)\nu(dy)$ *andy* $\in F \mapsto \int_F f(x, y)\mu(dx)$ *are resp.* $\mathcal{A}$*-measurable and* $\mathcal{B}$*-measurable.*
(ii) There holds

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

Theorem 7.6 (Fubini-Lebesgue). *Let* μ, ν *be two* σ*-finite measures on resp.* $(E, \mathcal{A})$ *and* $(F, \mathcal{B})$. *Let* $f \in \mathcal{L}^1(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$.
(i) For μ *a.e* $x \in E$, $f_x \in \mathcal{L}^1(F, \mathcal{B}, \nu)$ *and* ν *a.e* $y \in F$, $f^y \in \mathcal{L}^1(E, \mathcal{A}, \mu)$.
(ii) The function $x \mapsto \int_F f(x, y)\nu(dy)$ *is well defined for* μ *a.e* $x \in E$ *and belongs to* $\mathcal{L}^1(E, \mathcal{A}, \mu)$. *The function* $y \mapsto \int_E f(x, y)\mu(dx)$ *is well defined for* ν *a.e* $y \in F$ *and belongs to* $\mathcal{L}^1(F, \mathcal{B}, \nu)$.
(iii) There holds

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

7.2. **Applications.**

Definition 7.7 (Convolution). *Let* $f, g : \mathbb{R}^d \to \mathbb{R}$ *measurable. For* $x \in \mathbb{R}^d$, *we define the convolution product by*

$$f \star g(x) = \int f(x - y)g(y)dy.$$

Proposition 7.8. *Let $f, g \in L^1(\mathbb{R}^d, \overline{\mathcal{B}(\mathbb{R}^d)}, \lambda)$, then for λ a.e $x \in \mathbb{R}^d$, the convolution $f \star g(x)$ is well defined. Moreover, $f \star g \in L^1(\mathbb{R}^d, \overline{\mathcal{B}(\mathbb{R}^d)}, \lambda)$ with*

$$\|f \star g\|_{L^1} \leq \|f\|_{L^1} \|g\|_{L^1}.$$

Proposition 7.9 (Young's inequality)**.** *Ler $1 \leq r, p, q \leq +\infty$ with $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, then*

$$\|f \star g\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Definition 7.10 (Approximation of identity)**.** *We say a sequence of functions $\phi_n : \mathbb{R}^d \to \mathbb{R}$ is an approximation of identity if:*
(i) $\phi_n \geq 0$;
(ii) there exists $K \subset \mathbb{R}^d$ compact such that $Supp \phi_n \subset K$;
(iii) $\forall \delta > 0$, $\lim_{n \to \infty} \int_{|x| < \delta} \phi_n(x) dx = 0$.

Proposition 7.11 (Density of smooth functions in $L^p(\mathbb{R}^d)$)**.** *Let ϕ_n be an approximation of identity. Let $1 \leq p < \infty$. Then for all $f \in L^p(\mathbb{R}^d, \overline{\mathcal{B}(\mathbb{R}^d)}, \lambda)$, $\phi_n \star f \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ and $\lim_{n \to \infty} \|f_n - f\|_{L^p} = 0$.*

8. Properties of the Lebesgue measure

Theorem 8.1 (Change of variables formula)**.** *Let $\phi : U \to D$ $\mathcal{C}^1$ diffeomorphism. Then for all Borelian function $f : D \to \mathbb{R}_+$ or integrable $f \in \mathcal{L}^1(D, \lambda)$, there holds*

$$\int_D f(x) dx = \int_U f[\phi(u)] |J_\phi(u)| du.$$

Theorem 8.2 (Measure on the sphere)**.** *Let λ_d be the Lebesgue measure λ_d on $\mathbb{R}^d$.*
(i) For every Borelian $A \in \mathcal{B}(S^{d-1})$,

$$w_d(A) = d\lambda_d [\Gamma(A)]$$

where $\Gamma(A) = \{rz; z \in A, r \in [0,1]\}$, defines a finite measure on S^{d-1}, invariante by the linear isometries of $\mathbb{R}^d$, and with

$$w_d(S^{d-1}) = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}.$$

(ii) For every Borelian positive (or integrable) function f on $\mathbb{R}^d$, there holds

$$\int_{\mathbb{R}^d} f(x) d\lambda(x) = \int_0^\infty \int_{S^{d-1}} f(rx) r^{d-1} dr dw_d(z).$$

9. Random variables

9.1. Basic definitions.

We fix once and for all $(\Omega, \mathcal{A}, P)$ probability space.

Definition 9.1 (Random variable)**.** *Let $(\Omega, \mathcal{A})$ and $(E, \mathcal{E})$. A measurable map $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ is called a random variable with value in E. If $E = \mathbb{R}$, we speak of real random variable,*

Definition 9.2 (Law)**.** *The law of a random variable X is the probablity image of P by X:*

$$\forall B \in \mathcal{E}, \quad P_X(B) = P(X \in B) = P\left(X^{-1)}(B)\right).$$

Remark 9.3. If E is countable (ie the r.v is discrete), then

$$P_X = \sum_{x \in E} P(X = x)\delta_x$$

where δ_x is for $x \in E$ the measure

$$\forall B \in \mathcal{E}, \quad \delta_x(B) = \left| \begin{array}{ll} 1 & \text{if } \ x \in B \\ 0 & \text{otherwise} \end{array} \right.$$

Definition 9.4 (Density of a r.v)**.** *A r.v with value in* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *is said to have a law with density* p *if*

$$dP_X = p(x)dx \Leftrightarrow \forall B\mathcal{E}, \quad P_X(B) = \int_B p(x)dx.$$

(dx *is Lebesgue)*

Definition 9.5 (Expectation)**.** *Let* X *be a real r.v. We define its expectation by*

$$\mathbb{E}[X] = \int_\Omega X(w)P(dw).$$

Proposition 9.6. *Let* X *be a r.v with value in* $(E, \mathcal{E})$ *and* $f : (E, \mathcal{E}) \to [0, \infty]$*, then*

$$\mathbb{E}\left[f(X)\right] = \int_E f(x)P_X(dx).$$

Proposition 9.7 (Marginals)**.** *Let* $X = (X_1, \ldots, X_d)$ *be a r.v. with value in* $\mathbb{R}^d$*. Assume that the law of* X *has density* $p(x_1, \ldots, x_d)$*, then* $\forall j \in \{1, \ldots, d\}$*, the law of* X_j *has density*

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_d)dx_1 \ldots d_{x_{j-1}}dx_{j+1} \ldots dx_d.$$

9.2. **Classical laws.**

Definition 9.8 (Discrete laws)**.** *The following are classical examples of discrete laws.*

(i) Uniform law. *Let* E *be a set with* $\#E = n$*, a r.v has uniform law if* $\forall x \in E$*,* $P(X = x) = \frac{1}{n}$.

(ii)Bernoulli. *Let* $p \in [0, 1]$*, this is the law of a r.v* X *with value in* $\{0, 1\}$ *such that*

$$P(X = 1)p, \quad P(X = 0) = 1 - p.$$

It is intepreted as the result after one throw of a biaised coin which falls on head with probability p*.*

(iii) Binomial $\mathcal{B}(n, p)$, $n \in \mathbb{N}^*$, $p \in [0, 1]$ *It is the law of a r.v with values in* $\{1, \ldots, n\}$ *such that*

$$P(X = k) = (1 - p)p^k.$$

It is interpreted as the number of heads obtained after n *throws of the biaised coin.*

(iv) Poisson's law of parameter $\lambda > 0$. *It is the law of a r.v with value in* $\mathbb{N}$ *and*

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

It corresponds to the number of rare events during a long period. Mathematically, if X_n follows the law $\mathcal{B}(n, p_n)$ and $np_n \to \lambda$ as $n \to +\infty$, then

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Definition 9.9 (Continuous laws)**.** *The following are classical examples of continuous laws. Let X be a r.v with value in $\mathbb{R}$ and density $p(x)$.*

(i) Uniform law on $[a, b]$, $a < b$*:* $p(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$.

(ii) Exponential law with parameter $\lambda > 0$*:* $p(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}$.*. Exponential laws have the "no memory" property*

$$P(X > a + b) = P(X > a)P(X > b).$$

(iii) Gaussian law $\mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma > 0$*:*

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right).$$

The parameters m, σ are interpreted as

$$m = \mathbb{E}[X], \quad \sigma^2 = \mathbb{E}[(X - m)^2].$$

By convention, we say that a r.v. constant equal to m follows the Gaussian law $\mathcal{N}(m, 0)$. If X follows $\mathcal{N}(m, \sigma^2)$,

$$\forall \lambda, \mu \in \mathbb{R}, \quad \lambda X + \mu \ \textit{ follows } \ \mathcal{N}(\lambda m + \mu, \lambda^2\sigma^2).$$

9.3. **Structural definitions.**

Definition 9.10 (Distribution function)**.** *Let X be a real r.v, its distribution function is $F_X : \mathbb{R} \to [0, 1]$ given by*

$$F_X(t) = P(X \le t) = P_X(]-\infty, t]), \quad \forall t \in \mathbb{R}.$$

It is increasing, continuous on the right with $\lim_{t\to-\infty} F_X(t) = 0$, $\lim_{t\to+\infty} F_X(t) = 1$. Moreover,

$$\left| \begin{array}{l} P(a \le X \le b) = F_X(b) - F_X(a^-) \ \ \textit{for} \ \ a \le b \\ P(a < X < b) = F_X(b^-) - F_X(a) \ \ \textit{for} \ \ a < b. \end{array} \right.$$

Definition 9.11. *Let X be a r.v valued in $(E, \mathcal{E})$, the σ-algebra generated by X is*

$$\sigma(X) = \{X^{-1}(B), \quad B \in \mathcal{E}\}.$$

If $(X_i)_{i\in I}$ is a family of r.v valued in $(E_i, \mathcal{E}_i)$, the σ-algebra generated by the family is the smallest σ-algebra which makes all X_i measurable ie

$$\sigma\left[(X_i)_{i\in I}\right] = \sigma\left(\left\{X_i^{-1}(B_i), B_i \in \mathcal{E}_i\right\}\right).$$

Proposition 9.12. *Let X r.v valued in $(E, \mathcal{E})$ and Y real r.v. TFAE:*
(i) Y is $\sigma(X)$ measurable;
(ii) there exists $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable such that $Y = f(X)$.

9.4. **Moments and variance.**

Remark 9.13. Theorems of integration in probababilistic language:
(i) Monotone convergence: $X_n \geq 0$, $X_n \uparrow X \Rightarrow \mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.
(ii) Fatou: $X_n \geq 0 \Rightarrow \mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.
(iii) Lebesgue: $|X_n| \leq Z, \mathbb{E}[Z] < +\infty$, $X_n \to X$ a.e $\Rightarrow \mathbb{E}[X_n] \to \mathbb{E}[X]$.
(iv) Hölder: $\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} \left(\mathbb{E}[|X|^{p'}]\right)^{\frac{1}{p'}}$ for $\frac{1}{p} + \frac{1}{p}' = 1$, $1 \leq p, p' \leq +\infty$.
(v) L^p embeddings: $\|X\|_r \leq \|X\|_p$ for $1 \leq r \leq p \leq +\infty$.

Definition 9.14 (variance)**.** *Let $X \in L^2(\Omega, \mathcal{A}, P)$. The variance is*

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

and the standard deviation is

$$\sigma_X = \sqrt{\mathrm{var}(X)}.$$

Lemma 9.15. $\mathrm{var}(X) = \inf_{a \in \mathbb{R}} E[(X-a)^2]$.

Remark 9.16. Classical inequalities in probabilistic language:
(i) Markov: $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.
(ii) Bienayme-Tchebicheff: if $X \in L^2(\Omega, \mathcal{A}, P)$, $P\left(|X - \mathbb{E}(X)| \geq a\right) \leq \frac{\mathrm{var}(X)}{a^2}$.

Definition 9.17 (covariance)**.** *Let $X, Y \in L^2(\Omega, \mathcal{A}, P)$, their covariance is*

$$\mathrm{cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])\right] \mathbb{E}\left[(Y - \mathbb{E}[Y])\right] = [XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

If $X = (X_1, \dots, X_d)$ has coordinates in $X_i \in L^2(\Omega, \mathcal{A}, P)$, then the matrix covariance is

$$K_X = \left(\mathrm{cov}(X_i, X_j)\right)_{1 \leq i,j \leq n}.$$

Proposition 9.18 (Linear regression)**.** *Let $(X, Y_1, \dots, Y_n)$ be real r.v in $L^2(\Omega, \mathcal{A}, P)$. Then*

$$\inf_{\beta_0, \dots, \beta_n \in \mathbb{R}} \mathbb{E}\Big[\,(X - (\beta_0 + \beta_1 Y_1 + \dots + \beta_n Y_n))^2\,\Big] = \mathbb{E}\left[(X - Z)^2\right]$$

where

$$Z = \mathbb{E}[X] + \sum_{j=1}^{n} \alpha_j (Y_j - \mathbb{E}[Y_j])$$

and $(\alpha_j)_{1 \leq j \leq n}$ is any solution to the system

$$\sum_{j=1}^{n} \alpha_j \mathrm{cov}(Y_j, Y_k) = \mathrm{cov}(X, Y_k), \quad 1 \leq k \leq n.$$

9.5. **Characteristic function.**

Definition 9.19. *Let $f \in L^1(\mathbb{R}^d)$, we let $\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{-ix\cdot\xi} f(x)dx$.*

Lemma 9.20 (Inverse Fourier transform)**.** *Let $f \in \mathcal{C}_c^\infty(\mathbb{R}^d)$, then*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ix\cdot\xi} \hat{f}(\xi) d\xi.$$

Lemma 9.21 (Plancherel)**.** *Let $f, g \in \mathcal{C}_c^\infty(\mathbb{R}^d)$, then*

$$\int_{\mathbb{R}^d} f(x)\overline{g(x)}dx = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\xi)\overline{\hat{g}(\xi)}d\xi.$$

This allows us to uniquely extend the Fourier as a continuous isomorphism of L^2.

Definition 9.22 (Characteristic function). *Let X be a r.v valued in $\mathbb{R}^d$, its characteristic function is the function $\Phi_X : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\Phi_X(\xi) = \mathbb{E}\left[e^{ix\cdot\xi}\right] = \int_{\mathbb{R}^d} e^{ix\cdot\xi} P_X(dx).$$

Lemma 9.23 (Characteristic function of one dimensional Gaussian). *Let X be a real r.v with las $\mathcal{N}(0,\sigma^2)$, then*

$$\Phi_X(\xi) = e^{-\frac{\sigma^2\xi^2}{2}}.$$

Theorem 9.24. *The characteristic function of a r.v valued in $\mathbb{R}^d$ characterizes the law of this r.v. Equivalently, the Fourier transform defined on the space of probability measure on $\mathbb{R}^d$ is injective.*

The proof relies on the following (density related) statement.

Proposition 9.25. *Let μ, ν be two probability measures on $\mathbb{R}^d$. Assume that for all test function $\phi \in \mathcal{C}_b(\mathbb{R}^d)$ (ie continuous bounded and real valued),*

$$\int_{\mathbb{R}^d} \phi d\mu = \int_{\mathbb{R}^d} \phi d\nu$$

then $\mu = \nu$.

Proposition 9.26. *Let $X = (X_1, \dots, X_n)$ r.v valued in $\mathbb{R}^d$ and square integrable. Then Φ_X is $\mathcal{C}^2$ and*

$$\Phi_X(\xi) = 1 + i\sum_{j=1}^n \xi_j \mathbb{E}[X_j] - \frac{1}{2}\sum_{j=1}^d\sum_{k=1}^d \xi_j\xi_k \mathbb{E}[X_jX_k] + o(|\xi|^2).$$

Definition 9.27 (Generating function). *Let X r.v valued in $\mathbb{N}$. The generating function of X is the function $g_X : [0,1] \to \mathbb{R}_+$ given by*

$$g_X(r) = \mathbb{E}[r^X] = \sum_{n\geq 0} P(X=n) r^n.$$

10. Independance

10.1. Definitions.

Definition 10.1 (Independance of events). *We say n events $A_1, \dots, A_n$ are independent iff for all $\{j_1, \dots, j_p\} \subset \{1, \dots, n\}$,*

$$P\left(A_{j_1} \cap \dots \cap A_{j_p}\right) = P(A_{j_1}) \dots P(A_{j_p}).$$

Lemma 10.2. *The n events $A_1, \dots, A_n$ are independent iff*

$$P(B_1 \cap \dots \cap B_n) = P(B_1) \dots P(B_n)$$

whenever $B_i \in \sigma(A_i) \equiv \{\emptyset, A_i, A_i^c, \Omega\}$, $\forall i \in \{1, \dots, n\}$.

Definition 10.3 (Independance of tribes). *Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ be sub σ-algebra of $\mathcal{A}$. We say they are independent iff*

$$\forall A_1 \in \mathcal{B}_1, \dots, A_n \in \mathcal{B}_n, \quad P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n).$$

Definition 10.4 (Independance of re.v). *Let $X_1, \dots, X_n$ be r.v with values in $(E_1, \mathcal{E}_1)$, ..., $(E_n, \mathcal{E}_n)$. We say they are independent iff*

$$\forall F_1 \in \mathcal{E}_1, \dots, F_n \in \mathcal{E}_n, \;\; P\left(\{X_1 \in F_1\} \cap \dots \cap \{X_n \in F_n\}\right) = P(X_1 \in F_1) \dots P(X_n \in F_n).$$

Remark 10.5. (i) Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ family of independent sub σ-algebra, then if X_i $\mathcal{B}_i$-measurable r.v, then $(X_1, \dots, X_n)$ are independent.
(ii) Events $A_1, \dots, A_n$ are independent iff the sub σ-algebra $(\sigma(A_1), \dots, \sigma(A_n))$ are.

Definition 10.6. *Let $(\mathcal{B}_i)_{i\in I}$ an arbitrary family of sub σ-algebra, we say that the family is independent iff $\forall\{i_1, \dots, i_p\}$, $(\mathcal{B}_1, \dots, \mathcal{B}_{i_p})$ are independent. We say a an arbitrary family $(X_i)_{i\in I}$ of r.v is independent iff $(\sigma(X_i))_{i\in I}$ is.*

Corollary 10.7. *Let $(X_n)_{n\in\mathbb{N}}$ a family of independent r.v, then $\forall p \geq 1$,*

$$\mathcal{B}_1 = \sigma(X_1, \dots, X_p), \quad \mathcal{B}_2 = \sigma(X_{p+1}, X_{p+2}, \dots)$$

are independent.

10.2. **Characterization of independance.**

Theorem 10.8 (Characterization with law). *The r.v $(X_1, \dots, X_n)$ are independent iff the law of $X = (X_1, \dots, X_n)$ is the product of the law of the X_i's:*

$$P_X = P_{X_1} \otimes \cdots \otimes P_{X_n}.$$

In this case, for all $f_i \geq 0$ measurable on $(E_i, \mathcal{E}_i)$, $1 \leq i \leq n$:

$$\mathbb{E}\Big[\Pi_{i=1}^n f_i(X_i)\Big] = \Pi_{i=1}^n \mathbb{E}\left[f_i(X_i)\right].$$

Corollary 10.9. *Let X_1, X_2 be two L^2 integrable r.v, then* $\operatorname{cov}(X_1, X_2) = 0$.

Theorem 10.10 (Characterization with densities). *Let $(X_1, \dots, X_n)$ real r.v.*
(i) Assume that $\forall i \in \{1, \dots, n\}$, the law of X_i has density p_i, and $(X_1, \dots, X_n)$ are independant. Then the law of $X = (X_1, \dots, X_n)$ is

$$p(x_1, \dots, x_n) = \Pi_{i=1}^n p_i(x_i).$$

(ii) Conversely, assume that the law of $X = (X_1, \dots, X_n)$ has a density of the form

$$p(x_1, \dots, x_n) = \Pi_{i=1}^n q_i(x_i)$$

for some Borelian positive functions q_i. Then $(X_1, \dots, X_n)$ are independant and $\forall i \in \{1, \dots, n\}$, the law of X_i has density $p_i = C_i p_i$ for some constant $C_i > 0$.

Remark 10.11. Let $(X_1, \dots, X_n)$ real r.v, then TFAE:
(i) $(X_1, \dots, X_n)$ are independent.
(ii) $\forall a_1, \dots, a_n \in \mathbb{R}$, $P(X_1 \leq a_1, \dots, X_n \leq a_n) = \Pi_{i=1}^n P(X_i \leq a_i)$.
(iii) For all $f_1, \dots, f_n$ continuous bounded from $\mathbb{R}$ to $\mathbb{R}_+$, $\mathbb{E}\Big[\Pi_{i=1}^n f_i(X_i)\Big] = \Pi_{i=1}^n \mathbb{E}\left[f_i(X_i)\right]$.
(iv) The characteristic function of $X = (X_1, \dots, X_n)$ is $\Phi_X(\xi_1, \dots, \xi_n) = \Pi_{i=1}^n \Phi_{X_i}(\xi_i)$.

Proposition 10.12. *Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ be sub σ-algebra of $\mathcal{A}$. Assume:*
(i) $\forall i \in \{1, \dots, n\}$, there exists $\mathcal{C}_i \subset \mathcal{B}_i$ monotone class stable by finite intersection with $\sigma(\mathcal{C}_i) = \mathcal{B}_i$;
(ii) $\forall C_1 \in \mathcal{C}_1, \dots, C_n \in \mathcal{C}_n$, $P(C_1 \cap \cdots \cap C_n) = P(C_1) \dots P(C_n)$.
Then $\mathcal{B}_1, \dots, \mathcal{B}_n$ are independent.

Corollary 10.13 (Regrouping). *Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ independent sub σ-algebra. Then for all $0 < n_1 < \cdots < n_p = n$, the sub σ-algebra*

$$\left|\begin{array}{l} \mathcal{D}_1 = \mathcal{B}_1 \vee \cdots \vee \mathcal{B}_{n_1} \equiv \sigma(\mathcal{B}_1 \vee \cdots \vee \mathcal{B}_{n_1}) \\ \mathcal{D}_2 = \mathcal{B}_{n_1+1} \vee \cdots \vee \mathcal{B}_{n_2} \\ \dots \\ \mathcal{D}_p = \mathcal{B}_{n_{p-1}+1} \vee \cdots \vee \mathcal{B}_{n_p} \end{array}\right.$$

are independent.

Corollary 10.14. *If $(X_1,\dots,X_n)$ are independent, then the r.v*

$$Y_1=(X_1,\dots,X_{n_1}),\dots,Y_p=(X_{n_{p-1}+1},\dots,X_{n_p})$$

are also independent.

Definition 10.15 (Independance for an infinite family)**.** *(i) Let $(\mathcal{B}_i)_{i\in I}$ be an arbitrary sub family of tribes of $\mathcal{A}$. We say that this family is independent if for all $\{i_1,\dots,i_p\}$, $(\mathcal{B}_{i_1},\dots,\mathcal{B}_{i_p})$ is an independent family.*
(ii) We say an arbitrary family $(X_i)_{i\in I}$ is independent of the family of tribes $(\sigma(X_i))_{i\in I}$ is.

Proposition 10.16. *Let $(X_n)_{n\geq 1}$ be a family of independent r.v. Then for all integer $p\in\mathbb{N}$, the tribes $\mathcal{B}_1=\sigma(X_0,\dots,X_p)$ and $\mathcal{B}_2=\sigma(X_{p+1},X_{p+2},\dots)$ are independent.*

10.3. **Borel Cantelli.**

Lemma 10.17 (Borel Cantelli)**.** *Let $(A_n)_{n\in\mathbb{N}}$ a family of events, and define*

$$\left|\begin{array}{l}\limsup A_n=\cap_{n=0}^{\infty}\left(\cup_{k=n}^{\infty}A_k\right)\\ \liminf A_n=\cup_{n=0}^{\infty}\left(\cap_{k=n}^{\infty}A_k\right)\end{array}\right.$$

(i) If $\sum_{n\in\mathbb{N}}P(A_n)<+\infty$, then $P(\limsup A_n)=0$. Equivalently,

$$\text{a.s},\quad \{n\in\mathbb{N}, w\in A_n\}\quad \textit{is finite.}$$

(ii)If $\sum_{n\in\mathbb{N}}P(A_n)=+\infty$ and the events A_n are independent, then $P(\limsup A_n)=1$. Equivalenlty,

$$\text{a.s},\quad \{n\in\mathbb{N}, w\in A_n\}\quad \textit{is infinite.}$$

10.4. **Sum of independent random variables.**

Definition 10.18 (Convolution of measures)**.** *Let μ,ν be two probability measures on $\mathbb{R}^d$, then $\mu\star\nu$ is the image of $\mu\otimes\nu$ by the map $(x,y)\to x+y$. Equivalently, for any positive measurable ϕ:*

$$\int_{\mathbb{R}^d}\phi(z)\mu\star\nu(dz)=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\phi(x+y)\nu(dx)\nu(dy).$$

Proposition 10.19. *Let X,Y be two independent r.v with value in $\mathbb{R}^d$.*
(i) The law of $X+Y$ is $P_X\star P_Y$. In particular, if P_X has density p_X and P_Y has density p_y, then P_{X+Y} has density $p_X\star p_Y$.
(ii) The characteristic function of $X+Y$ is $\Phi_{X+Y}(\xi)=\Phi_X(\xi)\Phi_Y(\xi)$. (Equivalently, if μ,ν are two probability measures on $\mathbb{R}^d$, then $\widehat{\mu\star\nu}=\hat{\mu}\hat{\nu}$.).
(iii) If X,Y are square integrable, then $K_{X+Y}=K_X+K_Y$. In particular, for $d=1$, $\mathrm{var}(X+Y)=\mathrm{var}(X)+\mathrm{var}(Y)$.

Theorem 10.20 (Weak law of large numbers)**.** *Let $(X_n)_{n\geq 1}$ be a family of real valued independent r.v with same law. If $E(X_1^2)<+\infty$, then*

$$\frac{X_1+\dots+X_n}{n}\xrightarrow[n\to\infty]{L^2}\mathbb{E}[X_1].$$

Proposition 10.21. *Let $(X_n)_{n\geq 1}$ be a family of real valued independent r.v with same law. If $\mathbb{E}(X_1^4)<+\infty$, then*

$$\text{a.s}\ \ \frac{X_1+\dots+X_n}{n}\to\mathbb{E}[X_1].$$

Definition 10.22 (Convolution semi group)**.** *A family of probability measures* $(\mu_t)_{t\in I}$ *is said to be a convolution semi group if*

$$\left| \begin{array}{l} \mu_0 = \delta_0 \\ \mu_t \star \mu_{t'} = \mu_{t+t'}, \quad \forall t, t' \in I \end{array} \right.$$

Lemma 10.23. *For* $(\mu_t)_{t\in I}$ *to be a convolution semi group, it is enough that:*
(i) if $I = \mathbb{N}$*,* $\hat{\mu_t}(\xi) = [\phi(\xi)]^t$*,* $\forall t \in I$*.*
(ii) if $I = \mathbb{R}$*,* $\hat{\mu_t}(\xi) = e^{-t\phi(\xi)}$*,* $\forall t \in I$*.*

Examples.
(i) For $I = \mathbb{N}$ and $n > 0$, let μ_n be the binomial law $B(n,p)$ (where $p \in [0,1]$ has been fixed), then $\mu_{n+m} = \mu_n \star \mu_m$ can be seen by computing $\hat{\mu}_n(\xi) = (pe^{i\xi} + 1 - p)^n$.
(ii) For $I = \mathbb{R}_+$, $t \in \mathbb{R}_+$, let μ_t be the Poisson law of parameter t, then

$$\hat{\mu_t}(\xi) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} e^{ik\xi} e^{-t} = e^{-t(1-e^{i\xi})},$$

(iii) For $I = \mathbb{R}_+$, $t > 0$, let μ_t be the Gaussian law $\mathcal{N}(0,t)$, then

$$\hat{\mu_t}(\xi) = e^{-\frac{t\xi^2}{2}}.$$

Important consequences. Let X, Y be two independent real r.v.
(i) If X follows Poisson of parameter λ and Y follows Poisson of parameter λ', then $X + Y$ follows Poisson of parameter $\lambda + \lambda'$.
(ii) If X follows the Gaussian law $\mathcal{N}(m, \sigma^2)$ and X' follows the Gaussian law $\mathcal{N}(m', (\sigma')^2)$, then $X + X'$ follows the Gaussian law $\mathcal{N}(m + m', \sigma^2 + (\sigma')^2)$.

11. Convergence of random variables

11.1. **Convergence in probability.** We have already introduced

$$X_n \xrightarrow[n\to\infty]{a.s} X \Leftrightarrow P\left(\{x \in \Omega : \lim_{n\to\infty} X_n(w) = X(x)\}\right) = 1$$

and for $1 \leq p < +\infty$

$$X_n \xrightarrow[n\to\infty]{L^p} \Leftrightarrow \lim_{n\to\infty} \mathbb{E}\left[|X - X_n|^p\right] = 0.$$

Definition 11.1 (Convergence in probability)**.** *We say* X_n *converges to* X *in probability*

$$X_n \xrightarrow[n\to\infty]{(P)} X$$

ifs

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0.$$

Proposition 11.2 (Completeness)**.** *Let* $\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ *be the quotient of the space of all r.v with value in* $\mathbb{R}^d$ *quotiented by the equivalence relation* $X \sim Y \Leftrightarrow X = Y$ a.s.. *Then:*
(i) $d(X,Y) = \mathbb{E}\left[|X - Y| \wedge 1\right]$ *defines a distance on* $\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$*, and this metric space is complete.*
(ii) $\lim_{n\to\infty} d(X_n, X) = 0 \Leftrightarrow X_n \xrightarrow[n\to\infty]{(P)} X$*.*

Lemma 11.3. *If* X_n *converges a.s to* X *(or in* L^p*), then it also converges in probability. Conversely, if* X_n *converges in probability to* X*, then there exists a subsequence* X_{n_k} *which converges a.s to* X*.*

11.2. **Strong law of large numbers.**

Theorem 11.4 (Kolmogorov's 0-1 law). *Let $(X_n)_{n\geq 1}$ be a sequence of independent r.v. Let for $n \geq 1$*

$$\mathcal{B}_n = \sigma\left(X_k, k \geq n\right),$$

then the asymptotic tribe

$$\mathcal{B}_\infty = \cap_{n\geq 1}\mathcal{B}_n$$

is rough in the sense that

$$\forall B \in \mathcal{B}_\infty, \quad P(B) \in \{0,1\}.$$

Theorem 11.5 (Strong law of large numbers, a.e version). *Let $(X_n)_{n\geq 1}$ be a sequence of real independent r.v. with same law and $\mathbb{E}[|X_1|] < +\infty$, then*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{a.s} \mathbb{E}[X_1].$$

11.3. **Convergence in law.**

Definition 11.6 (Test function). *(i) We let $\mathcal{C}_b(\mathbb{R}^d)$ be the space of continuous bounded functions from $\mathbb{R}^d \to \mathbb{R}$ equipped with the sup norm*

$$\|\phi\|_{L^\infty} = \sup_{x\in\mathbb{R}^d} |\phi(x)|.$$

(ii) We let $\mathcal{C}_c(\mathbb{R}^d)$ the subset of $\mathcal{C}_b(\mathbb{R}^d)$ of continuous functions with compact support (where we recall $\mathrm{Supp}(f) = \overline{\{x \in \mathbb{R}^d, f(x) \neq 0\}}$.)

Definition 11.7 (Convergence in law). *(i) We say a sequence $(\mu_n)_{n\geq 1}$ of probability measures over $\mathbb{R}^d$ converges in distribution to a probability measure μ ($\mu_n \xrightarrow[n\to\infty]{(D)} \mu$) iff*

$$\forall \phi \in \mathcal{C}_b(\mathbb{R}^d), \quad \lim_{n\to\infty} \int \phi d\mu_n = \int \phi d\mu.$$

(ii) We say a sequence of r.v. $(X_n)_{n\geq 1}$ with value in $\mathbb{R}^d$ converges in law (or in distribution) to X iff $P_{X_n} \xrightarrow[n\to\infty]{(D)} P_X$. Equivalently,

$$\forall \phi \in \mathcal{C}_b(\mathbb{R}^d), \quad \lim_{n\to\infty} \mathbb{E}\left[\phi(X_n)\right] = \mathbb{E}\left[\phi(X)\right].$$

Lemma 11.8. *Let (μ_n, μ) be probability measures on $\mathbb{R}^d$. TFAE:*
(i) $\mu_n \xrightarrow[n\to\infty]{(D)} \mu$;
(ii) $\forall G$ open, $\liminf \mu_n(G) \geq \mu(G)$;
(iii) $\forall F$ closed, $\limsup \mu_n(F) \leq \mu(F)$;
(iv) $\forall B$ Borelian with $\mu(\partial B) = 0$, $\lim_{n\to\infty} \mu_n(B) = \mu(B)$ where $\partial B = \overline{B}\backslash\mathring{B}$.

Proposition 11.9. *A sequence of r.v. $(X_n)_{n\geq 1}$ with value in $\mathbb{R}^d$ converges in law to X iff the distribution functions $F_{X_n}(x)$ converges to $F_X(x)$ at every point x where F_X is continuous.*

Proposition 11.10 (Weakening test functions). *Let $(\mu_n)_{n\geq 1}, \mu$ be probality measures over $\mathbb{R}^d$. Let H be a subspace of $\mathcal{C}_b(\mathbb{R}^d)$ which closure (for the sup norm) contains $\mathcal{C}_c(\mathbb{R}^d)$. Then TFAE:*
(i) $\mu_n \xrightarrow[n\to\infty]{(D)} \mu$.
(ii) $\forall \phi \in \mathcal{C}_c(\mathbb{R}^d)$, $\lim_{n\to\infty} \int \phi d\mu_n = \int \phi d\mu$.
(iii) $\forall \phi \in H$, $\lim_{n\to\infty} \int \phi d\mu_n = \int \phi d\mu$.

Theorem 11.11 (Levy)**.** *(i) A sequence of probability measures $(\mu_n)_{n\geq 1}$ converges in distribution to a probability measure μ iff*

$$\forall \xi \in \mathbb{R}^d, \quad \lim_{n\to\infty} \widehat{\mu_n}(\xi) = \hat{\mu}(\xi).$$

(ii) A sequence of r.v. $(X_n)_{n\geq 1}$ with value in $\mathbb{R}^d$ converges in law (or in distribution) to X iff

$$\forall \xi \in \mathbb{R}^d, \quad \lim_{n\to\infty} \Phi_{X_n}(\xi) = \Phi_X(\xi).$$

11.4. **Central limit theorem.**

Theorem 11.12 (Scalar central limit theorem)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of real r.v independent with same law, $X_1 \in L^2$. Let $\sigma^2 = \mathrm{var}(X_1)$, then*

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{(law)}} \mathcal{N}(0, \sigma^2).$$

Equivalently, $\forall a < b$,

$$\lim_{n\to\infty} \mathbb{P}\left(X_1 + \cdots + X_n \in (n\mathbb{E}[X_1] + a\sqrt{n}, n\mathbb{E}[X_1] + b\sqrt{n})\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2\sigma^2}} dx.$$

Definition 11.13 (Gaussian centered vector)**.** *Let C be a square $d \times d$ symmetric positive matrix, with real valued entries. A square integrable r.v valued in $\mathbb{R}^d$ is said to be a centered Gaussian vector with covariance C if*

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_X(\xi) = e^{-\frac{1}{2}{}^t\xi C \xi}.$$

We also say that X follows the law $\mathcal{N}(0, C)$.

Remark 11.14. $X \sim \mathcal{N}(0, C) \Rightarrow (\mathbb{E}[X] = 0, K_X = C)$.

Lemma 11.15 (Existence of Gaussian centered vectors)**.** *Let $A = \sqrt{C}$ and $(Y_i)_{1\leq i\leq d}$ independent real r.v following the law $\mathcal{N}(0,1)$, then $X = AY$ follows $\mathcal{N}(0, C)$.*

Theorem 11.16 (Vectorial central limit theorem)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of r.v valued in $\mathbb{R}^d$ independent with same law, $X_1 \in L^2$, $C = K_{X_1}$, then*

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{(law)}} \mathcal{N}(0, C).$$

12. **Ergodic theory**

12.1. **Basic results.**

Definition 12.1. *Let $(E, \mathcal{E}, \Gamma)$ be a measure space.*
(i) A map $\theta : E \to E$ is called measure preserving if: $\forall A \in \mathcal{E}$, $\Gamma(\theta^{-1})(A)) = \Gamma(A)$. In this case,

$$\forall f \in L^1, \quad \int_E f d\Gamma = \int_E f \circ \theta d\Gamma.$$

(ii) A measurable function f is called θ-invariant if $f = f \circ \theta$.
(iii) A set $A \in \mathcal{E}$ is called θ-invariant if $\theta^{-1}(A) = A$. The family $\mathcal{E}_\theta$ of θ-invariant sets is a σ-algebra and f invariant iff f is $\mathcal{E}_\theta$-measurable.

Definition 12.2. *The map θ is called ergodic if*

$$A \in \mathcal{E}_\theta \Rightarrow (\Gamma(A) = 0 \ \ or \ \ \Gamma(E \backslash A) = 0).$$

Remark 12.3. f θ-invariant for θ ergodic implies f constant Γ-a.e.

Theorem 12.4 (Birkhoff). *Let $(E,\mathcal{E},\Gamma)$ be σ-finite,. Let θ be a measure preserving map. Given $f\in L^1$, let*

$$\left| \begin{array}{l} S_0(f)=0 \\ S_n(f)=\sum_{k=0}^{n-1} f\circ\theta^k, \end{array} \right.$$

then $\exists \overline{f}\in L^1$ θ-invariant such that

$$\frac{S_n(f)}{n}\xrightarrow[n\to\infty]{}\overline{f}\quad \Gamma \quad a.e.$$

Lemma 12.5 (Ergodic Lemma). *Under the assumptions of Birkhoff's theorem, let $S^*(f)=\sup_{n\geq 0} S_n(f)$, then*

$$\int_{S^*(f)>0} f d\Gamma \geq 0.$$

Theorem 12.6 (Von Neumann's). *Assume $\Gamma(E)<+\infty$. Pick $1\leq p<+\infty$, θ-meaure preserving and $f\in L^p$. Then*

$$\frac{S_n(f)}{n}\xrightarrow[n\to\infty]{L^p}\overline{f}$$

for some $\overline{f}\in L^p$ θ-invariant.

12.2. **Application to iid.** Consider the infinite product space

$$E=\mathbb{R}^{\mathbb{N}}=\{(x_n)_{n\geq 1},\quad x_n\in\mathbb{R}\}$$

equipped with the cylindrical σ-algebra $\sigma(\mathcal{C})$ generated by cylinders

$$\mathcal{C}=\{A=\Pi_{n=1}^{+\infty}A_n,\quad A_n\in\mathcal{B},\quad A_n=\mathbb{R}\ \text{ for some }\ n\geq N\}.$$

Given $(X_n)_{n\geq 1}$ real valued independent r.v on $(\Omega,\mathcal{F},\mathbb{P})$ with law m, the map

$$X:\Omega\to E,\quad X(w)=(X_1(w),X_2(w),\dots)$$

is measurable and the image measure $\Gamma=P_X$ is the unique measure $\sigma(\mathcal{C})$ satisfying

$$\Gamma(\Pi_{n=1}^{+\infty}A_n)=\Pi_{n=1}^{+\infty}m(A_n).$$

Definition 12.7. *The product space $(E=\mathbb{R}^{\mathbb{N}},\sigma(\mathcal{C}),\Gamma)$ is called the canonical model associated to the sequence of iid $X(n)_{n\geq 0}$.*

Lemma 12.8. *The shit map*

$$\theta:E\to E,\quad \theta(x_1,x_2,\dots)=(x_2,x_3,\dots)$$

is Γ-measure preserving and ergodic.

By applying Von Neuman to

$$f:E\to\mathbb{R},\quad f(x)=x_1,$$

we obtain the L^1 version of the strong law of large numbers.

Theorem 12.9 (Strong Law of Large Numbers, L^1 version). *Let $(X_n)_{n\geq 1}$ be real valued independent r.v with same law and $\mathbb{E}[|X_1|]<+\infty$, then*

$$\mathbb{E}\left[\left|\frac{X_1+\dots+X_n}{n}-\mathbb{E}[X_1]\right|\right]\xrightarrow[n\to\infty]{}0.$$

Department of Pure Mathematics and Mathematical Statistics, Cambridge, UK
E-mail address: pr463@cam.ac.uk