Probability and Measure

Shing Tak Lam

May 31, 2023

Contents

1 Measures 1
- 1.1 Definitions 1
- 1.2 Dynkin's lemma 2
- 1.3 Carathéodory extension and uniqueness of measures 3
- 1.4 Borel and Lebesgue measure 4
- 1.5 Measurable functions 5
- 1.6 Image measures 6
- 1.7 Convergence of measurable functions 7

2 Probability theory 8
- 2.1 Independence 8
- 2.2 Borel-Cantelli lemmas 9
- 2.3 Random variables 9
- 2.4 Convergence of random variables 11
- 2.5 Tail events 11

3 Integration 12
- 3.1 Definition of the integral 12
- 3.2 Monotone convergence theorem 13
- 3.3 Fatou and dominated convergence 14
- 3.4 Densities and image measure 15
- 3.5 Products 16

4 L^p spaces 18
- 4.1 Inequalities 19
- 4.2 Banach and Hilbert space structure 20
- 4.3 Convergence in $L^1(\mathbb{P})$ and uniform integrability 22

5 Fourier transforms 24
- 5.1 Definitions 24
- 5.2 Gaussian convolutions and Fourier inversion 25
- 5.3 Fourier-Plancherel 27
- 5.4 Weak convergence and characteristic functions 28

6 Ergodic theory 29
- 6.1 Birkhoff and von Neumann ergodic theorems 29
- 6.2 Shift map 31

7 Limit theorems 31

1 Measures

1.1 Definitions

Definition 1.1 (σ-algebra, measurable space)
Let E be a set. A collection $\mathcal{E}$ of subsets of E is called a σ-algebra if

- $\varnothing \in \mathcal{E}$,
- for all $A \in \mathcal{E}$, $A^{\complement} = E \setminus A \in \mathcal{E}$,
- for all $(A_n) \subseteq \mathcal{E}$,
$$\bigcup_n A_n \in \mathcal{E}$$

The pair $(E, \mathcal{E})$ is called a measurable space, and $A \in \mathcal{E}$ is called a measurable set.

Definition 1.2 (measure)
A measure μ on $(E, \mathcal{E})$ is a function $\mu : \mathcal{E} \to [0, \infty]$ such that

- $\mu(\varnothing) = 0$,
- for all $(A_n) \subseteq \mathcal{E}$, with the A_n pairwise disjoint, we have that
$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$$

The triple $(E, \mathcal{E}, \mu)$ is called a measure space.

Definition 1.3 (generated σ-algebra)
Let $\mathcal{A}$ be a collection of subsets of E. Then define

$$\sigma(\mathcal{A}) = \{A \subseteq E \mid A \in \mathcal{E} \text{ for any } \sigma\text{-algebra } \mathcal{E} \text{ containing } \mathcal{A}\}$$

Which is called the σ-algebra generated by $\mathcal{A}$.

1.2 Dynkin's lemma

Definition 1.4 (π-system)
Let $\mathcal{A}$ be a set of subsets of E. Then we say that $\mathcal{A}$ is a π-system if

- $\varnothing \in \mathcal{A}$,
- for any $A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$.

Definition 1.5 (d-system)
Let $\mathcal{A}$ be a set of subsets of E. We say that $\mathcal{A}$ is a d-system if

- $E \in \mathcal{A}$,
- for all $A, B \in \mathcal{A}$ with $A \subseteq B$, we have that $B \setminus A \in \mathcal{A}$,
- for any increasing sequence $(A_n) \subseteq \mathcal{A}$,
$$\bigcup_n A_n \in \mathcal{A}$$

Lemma 1.6 (Dynkin's lemma)**.** Let $\mathcal{A}$ be a π-system. Then any d-system containing $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.

Proof. Let $\mathcal{D}$ be the intersection of all d-systems containing $\mathcal{A}$. Then $\mathcal{D}$ is a d-system. We will show that $\mathcal{D}$ is also a π-system, so it is a σ-algebra. Consider

$$\mathcal{D}' = \{B \in \mathcal{D} \mid B \cap A \in D \text{ for all } A \in \mathcal{A}\}$$

As $\mathcal{A}$ is a π-system, we must have that $\mathcal{A} \subseteq \mathcal{D}'$. We'll now show that $\mathcal{D}'$ is a d-system. Suppose $A, B \in \mathcal{D}'$ with $A \subseteq B$, and $C \in \mathcal{A}$. Then

$$(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in \mathcal{D}$$

So $B \setminus A \in \mathcal{D}'$. Now suppose $(B_n) \subseteq \mathcal{D}'$, with $B_n \nearrow B$, then for $\mathcal{A} \in A$, we have that $B_n \cap A \nearrow B \cap A$, so $B \cap A \in \mathcal{D}$, and $B \in \mathcal{D}'$. So we must have that $\mathcal{D} \subseteq \mathcal{D}' \subseteq \mathcal{D}$, so $\mathcal{D} = \mathcal{D}'$. Now define

$$\mathcal{D}'' = \{B \in \mathcal{D} \mid B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{D}\}$$

Then $\mathcal{D}''$ is a d-system, with $\mathcal{A} \subseteq \mathcal{D}''$, so $\mathcal{D}'' = \mathcal{D}$, and $\mathcal{D}$ is a π-system. □

1.3 Carathéodory extension and uniqueness of measures

Definition 1.7 (ring)

A collection $\mathcal{A}$ of subsets of E is called a ring if

- $\varnothing \in \mathcal{A}$,
- for any $A, B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$ and $B \setminus A \in \mathcal{A}$.

Definition 1.8 (algebra)

A collection $\mathcal{A}$ of subsets of E is called an algebra if

- $\varnothing \in \mathcal{A}$,
- for any $A \in \mathcal{A}$, $A^{\complement} \in \mathcal{A}$,
- for all $A, B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$.

That is, a σ-algebra is an algebra where we allow countable unions.

Definition 1.9 (outer measure)

Let $\mathcal{A}$ be a ring of subsets of E, $\mu : \mathcal{A} \to [0, \infty]$ be countably additive. Then for any $B \subseteq E$, define the outer measure of E to be

$$\mu^*(E) = \inf \left\{ \sum_n \mu(A_n) \,\middle|\, (A_n) \subseteq \mathcal{A}, B \subseteq \bigcup_n A_n \right\}$$

where we set $\inf \varnothing = \infty$. Let

$$\mathcal{M} = \left\{ A \subseteq \mathcal{E} \,\middle|\, \mu^*(B) = \mu^*(A \cap B) + \mu^*(A^{\complement} \cap B) \text{ for all } B \subseteq E \right\}$$

denote the set of μ^*-measurable sets.

Theorem 1.10 (Carathéodory extension)**.** Let $\mathcal{A}$ be a ring of subsets of $\mathcal{E}$, $\mu : \mathcal{A} \to [0, \infty]$ be a countably additive set function. Then μ extends to a measure on $\sigma(\mathcal{A})$.

More precisely, we will show that $\mathcal{M}$ is a σ-algebra containing $\mathcal{A}$, and μ^* restricts to a measure on

$\mathcal{M}$ extending μ.

Proof. Proof is non-examinable. □

Theorem 1.11. Let μ_1, μ_2 be measures on $(E, \mathcal{E})$, with $\mu_1(E) = \mu_2(E) < \infty$. Suppose $\mu_1 = \mu_2$ on a π-system $\mathcal{A}$ generating $\mathcal{E}$. Then $\mu_1 = \mu_2$.

Proof. Let $\mathcal{D} = \{A \in \mathcal{E} \mid \mu_1(A) = \mu_2(A)\}$. Then by hypothesis, $E \in \mathcal{E}$ and for any $A, B \in \mathcal{D}$, with $A \subseteq B$, we have that

$$\mu_i(A) + \mu_i(B \setminus A) = \mu_i(B) < \infty \quad \text{for} \quad i = 1, 2$$

So we must have thta $B \setminus A \in \mathcal{D}$. Furthermore, if $(A_n) \subseteq \mathcal{D}$ with $A_n \nearrow A$, then

$$\mu_1(A) = \lim_n \mu_1(A_n) = \lim_n \mu_2(A_n) = \mu_2(A)$$

So $A \in \mathcal{D}$. Therefore $\mathcal{D}$ is a d-system containing $\mathcal{A}$, so by Dynkin's lemma, $\mathcal{E} \subseteq \mathcal{D}$. □

1.4 Borel and Lebesgue measure

Definition 1.12 (Borel σ-algebra and Borel, Radon measures)

Let E be a Hausdorff topological space[a]. Then the σ-algebra generated by the open sets of $\mathcal{E}$ is called the Borel σ-algebra of $\mathcal{E}$, and is denoted $\mathcal{B}(E)$. We write $\mathcal{B} = \mathcal{B}(\mathbb{R})$.

A measure on $(E, \mathcal{B}(E))$ is called a Borel measure. If μ is a Borel measure with $\mu(K) < \infty$ for all $K \subseteq E$ compact, then we call μ a Radon measure.

[a] Strictly speaking Hausdorff is not necessary, however as we are only looking at the open sets of a topological space, we don't lose any generality by requiring the space to be Hausdorff.

Definition 1.13 (finite and σ-finite measures)

A measure μ on $E, \mathcal{E}$ is finite if $\mu(E) < \infty$. μ is σ-finite if there exists sets $(E_n)_{n \in \mathbb{N}} \subseteq E$ such that $\bigcup_n E_n = E$ and $\mu(E_n) < \infty$ for all n.

Theorem 1.14 (Lebesgue measure)**.** There exists a unique Borel measure μ on $\mathbb{R}$ such that for all $a < b$,

$$\mu((a, b]) = b - a \tag{*}$$

μ is called the Lebesgue measure on $\mathbb{R}$.

Proof. **Existence:** Consider the ring

$$\mathcal{A} = \{(a_1, b_1] \cup \cdots \cup (a_n, b_n] \text{ disjoint intervals}\}$$

which generates $\mathcal{B}$. Then for such $A \in \mathcal{A}$, define

$$\mu(A) = \sum_{i=1}^{n} (b_i - a_i)$$

It is easy to check that μ is well defined and additive. By Carathéodory's extension theorem, suffices to show that μ is countably additive on $\mathcal{A}$. By additivity, suffices to show if $A \in \mathcal{A}$, and $A_n \nearrow A$ is an increasing sequence in $\mathcal{A}$, then $\mu(A_n) \nearrow \mu(A)$. Set $B_n = A \setminus A_n$, then $B_n \in \mathcal{A}$, and $B_n \searrow \varnothing$. By additivity, suffices, to show $\mu(B_n) \searrow 0$. Suppose not. Then there exists $\varepsilon > 0$ such that $\mu(B_n) \geq 2\varepsilon$ for all n. For each n, we can find $C_n \in \mathcal{A}$ such that $C_n \subseteq B_n$, and $\mu(B_n \setminus C_n) \leq 2^{-n}\varepsilon$. Then

$$\mu(B_n \setminus (C_1 \cap \cdots \cap C_n)) \leq \mu((B_1 \setminus C_1) \cap \cdots \cap (B_n \setminus C_n) \leq \sum_n 2^{-n}\varepsilon = \varepsilon$$

Since $\mu(B_n) \geq 2\varepsilon$, we must then have that $\mu(C_1 \cap \cdots \cap C_n) \geq \varepsilon$ for all n, which means that $K_n = \overline{C_1} \cap \cdots \cap \overline{C_n} \neq \varnothing$ for all n. In this case, we then have that (K_n) is a decreasing sequence of compact sets, so $\varnothing \neq \bigcap_n K_n \subseteq \bigcap_n B_n$. Contradiction.

Uniqueness: Let λ be any measure on $\mathcal{B}$ satisfying (*). Fix $n \in \mathbb{Z}$ and define

$$\mu_n(A) = \mu((n, n+1] \cap A) \quad \text{and} \quad \lambda_n(A) = \lambda((n, n+1] \cap A)$$

Then μ_n, λ_n are probability measures on $\mathcal{B}$ and $\lambda_n = \mu_n$ on the π-system of intervals of the form $(a, b]$ which generates $\mathcal{B}$. Therefore, by uniqueness of measures, $\mu_n = \lambda_n$ on $\mathcal{B}$. Hence for all $A \in \mathcal{B}$, we have that

$$\mu(A) = \sum_n \mu_n(A) = \sum_n \lambda_n(A) = \lambda(A)$$

□

Corollary 1.15. The Lebesgue measure is translation invariant, if we define for $x \in \mathbb{R}$, $B \in \mathcal{B}$,

$$\mu_x(B) = \mu(B + x)$$

Then $\mu_x = \mu$.

1.5 Measurable functions

Definition 1.16 (measurable function)

Let $(E, \mathcal{E}), (G, \mathcal{G})$ be measurable spaces, then $f : E \to G$ is $\mathcal{E}$-$\mathcal{G}$-measurable if for any $A \in \mathcal{G}$, $f^{-1}(A) \in \mathcal{E}$. If $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$, we say that f is $\mathcal{E}$-measurable, and if $(G, \mathcal{G}) = ([0, \infty], \mathcal{B}([0, \infty]))$, we say that f is a nonnegative measurable function. Moreover, if E is a topological space, $\mathcal{E} = \mathcal{B}(E)$, then we say that f is Borel measurable.

Lemma 1.17. For any $f : E \to G$, $\left\{f^{-1}(A) \mid A \in \mathcal{G}\right\}$ is a σ-algebra over E, and $\left\{A \mid f^{-1}(A) \in \mathcal{E}\right\}$ is a σ-algebra over G.

Lemma 1.18. Suppose $\mathcal{G} = \sigma(\mathcal{A})$, $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A}$, then f is measurable.

Proof. $\{A \mid f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra containing $\mathcal{A}$, so it contains $\mathcal{G}$. Hence f is measurable. □

Corollary 1.19. $f : E \to \mathbb{R}$ is measurable if and only if $\{x \mid f(x) \leq y\}$ is measurable for all y.

Corollary 1.20. If E is a topological space, $f : E \to \mathbb{R}$ continuous, then f is measurable.

Definition 1.21 (σ-algebra generated by functions)

Given any family of functions $f_i : E \to G$, $i \in I$, we can make them all measurable by taking

$$\mathcal{E} = \sigma(f_i^{-1}(A) \mid A \in \mathcal{G}, i \in I)$$

Then $\mathcal{E}$ is the σ-algebra generated by $(f_i)_{i\in I}$.

Proposition 1.22. The sum, product, composition, lim, lim inf, lim sup, ... of measurable functions are measurable.

Theorem 1.23 (Monotone class theorem)**.** Let $(E, \mathcal{E})$ be a measurable space, $\mathcal{A}$ a π-system generating $\mathcal{E}$. Let $\mathcal{V}$ be a vector space of bounded functions $f : E \to \mathbb{R}$ such that

1. $1 \in \mathcal{V}$ and $1_A \in \mathcal{V}$ for all $A \in \mathcal{A}$,
2. if $(f_n) \subseteq \mathcal{V}$, f bounded and $0 \leq f_n \nearrow f$, then $f \in \mathcal{V}$.

Then $\mathcal{V}$ contains every bounded measurable function.

Proof. Consider $\mathcal{D} = \{A \in \mathcal{E} \mid 1_A \in \mathcal{V}\}$. Then $\mathcal{D}$ is a d-system containing $\mathcal{A}$, so $\mathcal{D} = \mathcal{E}$. Since $\mathcal{V}$ is a vector space, it contains all finite linear combinations of indicator functions of measurable sets. If f is bounded nonnegative and measurable, then

$$f_n = 2^{-n} \lfloor 2^n f \rfloor \in \mathcal{V} \quad \text{with} \quad 0 \leq f_n \nearrow f$$

So $f \in \mathcal{V}$. Finally, any bounded measurable function f can be written as $f = g - h$, where g, h are bounded nonnegative and measurable. □

1.6 Image measures

Definition 1.24 (image measure)

Let $(E, \mathcal{E}), (G, \mathcal{G})$ be measurable spaces, $f : E \to G$ be a measurable function and μ a measure on $\mathcal{E}$. Then the image measure on $\mathcal{G}$ is defined by $\mu \circ f^{-1}$, where

$$(\mu \circ f^{-1})(A) = \mu(f^{-1}(A))$$

Lemma 1.25. Let $g : \mathbb{R} \to \mathbb{R}$ be increasing right continuous. Then set $g(\pm\infty) = \lim_{x\to\pm\infty} g(x)$ and write $I = (g(-\infty), g(\infty))$. Define $f : I \to \mathbb{R}$ by

$$f(x) = \inf\{y \in \mathbb{R} \mid x \leq g(y)\}$$

Then f is increasing left continuous, and for all $x \in I, y \in \mathbb{R}$,

$$f(x) \leq y \iff x \leq g(y)$$

We call f the generalised inverse of g.

Proof. Define $J_x = \{y \in \mathbb{R} \mid x \leq g(y)\}$. Since $x > g(-\infty)$, J_x is nonempty and bounded below. Thus, $f(x) = \inf J_x$ exists. Since g is increasing, if $y \in J_x$, $y' \geq y$, then $y' \in J_x$. As g is right continuous, if $y_n \in J_x$, with $y_n \searrow y$, then $y \in J_x$. So $J_x = [f(x), \infty)$. Furthermore, $x \leq g(y)$ if and only if $f(x) \leq y$.

For $x \leq x'$, we have that $J_x \supseteq J_{x'}$, so $f(x) \leq f(x')$. Finally, if $x_n \nearrow x$, then $J_{x_n} \searrow J_x$, and so $f(x_n) \nearrow f(x)$. □

Theorem 1.26. Let $g : \mathbb{R} \to \mathbb{R}$ be a increasing right continuous function. Then there exists a unique Radon measure μ_g on $\mathbb{R}$ such that

$$\mu_g((a,b]) = g(b) - g(a)$$

for all $a < b$. Furthermore, any Radon measure on $\mathbb{R}$ can be obtained this way.

Proof. Let f be the generalised inverse of g, then $f^{-1}((-\infty, z]) = (g(-\infty), g(z)]$ is measurable, so f is Borel measurable. Thus, the image measure $\mu_g = \mu \circ f^{-1}$ satisfies

$$\mu_g((a,b]) = \mu(\{x \mid a < f(x) \leq b\}) = \mu([g(a), g(b)) = g(b) - g(a))$$

This uniquely determines the measure, by the same argument as for the Lebesgue measure. Finally, if ν is any Radon measure on $\mathbb{R}$, then define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(y) = \begin{cases} \nu((0,y]) & \text{if } y \geq 0 \\ -\nu((0,y]) & \text{if } y < 0 \end{cases}$$

Then $\nu((a,b]) = g(b) - g(a)$, so $\nu = \mu_g$ by uniqueness. □

1.7 Convergence of measurable functions

Definition 1.27 (almost everywhere)
Let $(E, \mathcal{E}, \mu)$ be a measure space, a property P holds almost everywhere if

$$\mu(\{x \mid \text{not } P(x)\}) = 0$$

Definition 1.28 (almost everywhere convergence)
A sequence of measurable functions f_n converges to f almost everywhere if

$$\mu(\{x \mid f_n(x) \not\to f(x)\}) = 0$$

Definition 1.29 (convergence in measure)
A sequence of measurable functions f_n converges to f in measure if

$$\mu(\{x \mid |f_n(x) - f(x)| > \varepsilon\}) = 0$$

for all $\varepsilon > 0$.

Theorem 1.30. Let (f_n) be a sequence of measurable functions, then

(i) If $\mu(E) < \infty$ and $f_n \to 0$ a.e., then $f_n \to 0$ in measure.

(ii) If $f_n \to 0$ in measure, then $f_{n_k} \to 0$ a.e. for some subsequence (n_k).

Proof. (i) Fix $\varepsilon > 0$, we have that

$$\mu(|f_n| \leq \varepsilon) \geq \mu\left(\bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\right)$$

and

$$\mu\left(\left\{\bigcap_{m\geq n}\{|f_m|\leq\varepsilon\}\right\}\right)\nearrow\mu\left(\bigcup_n\bigcap_{m\geq n}\{|f_m|\leq\varepsilon\}\right)$$
$$=\mu\left(|f_m|\leq\varepsilon \text{ eventually}\right)$$
$$\geq\mu(f_n\to 0 \text{ as } n\to\infty)$$
$$=\mu(E)$$

Hence $\liminf_n \mu(|f_n|\leq\varepsilon)\geq\mu(E)$, so $\limsup_n \mu(|f_n|>\varepsilon)\leq 0$, and so $\mu(|f_n|>\varepsilon)\to 0$ as $n\to\infty$.

(ii)By hypothesis, for fixed k and $\varepsilon>0$, for n large, we have that $\mu(|f_n|>1/k)<\varepsilon$. Choosing $\varepsilon=1/k^2$, we have that along some subsequence,

$$\mu(|f_{n_k}|>1/k)<1/k^2$$

Thus, $\sum_k \mu(|f_{n_k}|>1/k)<\infty$, so by the first Borel-Cantelli lemma, we have that

$$\mu\left(|f_{n_k}|>1/k \text{ i.o.}\right)=0$$

So $f_{n_k}\to 0$ almost everywhere. □

2 Probability theory

Definition 2.1 (probability measure)

A measure μ on $E,\mathcal{E}$ is called a probability measure if $\mu(E)=1$, and $(E,\mathcal{E},\mu)$ is called a probability space. We often write $(\Omega,\mathcal{F},\mathbb{P})$ for a probability space.

We can think of Ω as the set of outcomes, $\mathcal{F}$ the set of events, and $\mathbb{P}(A)$ is the probability of an event.

2.1 Independence

Definition 2.2 (independence of events, σ-algebra)

Let I be a countable set. Then $(A_i)_{i\in I}\subseteq\mathcal{F}$ is independent if for all $J\subseteq I$ finite,

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

A family $(\mathcal{A}_i)$ of sub-σ-algebras of $\mathcal{F}$ is independent if (A_i) is independent whenever $A_i\in\mathcal{A}_i$ for all i.

Theorem 2.3. Let $\mathcal{A}_1,\mathcal{A}_2$ be π-systems contained in $\mathcal{F}$, and suppose that

$$\mathbb{P}(A_1\cap A_2)=\mathbb{P}(A_1)\mathbb{P}(A_2)$$

whenever $A_1\in\mathcal{A}_1$ and $A_2\in\mathcal{A}_2$. Then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

Proof. Fix $A_1\in\mathcal{A}_1$, and define for $A\in\mathcal{F}$,

$$\mu(A)=\mathbb{P}(A_1\cap A)\quad\text{and}\quad\nu(A)=\mathbb{P}(A_1)\mathbb{P}(A)$$

Then μ and ν are measures which agree on the π-system $\mathcal{A}_2$, with $\mu(\Omega)=\nu(\Omega)<\infty$, hence by uniqueness of measures, $\mu=\nu$. Thus, for all $A_2\in\sigma(\mathcal{A}_2)$, we have that

$$\mathbb{P}(A_1\cap A_2)=\mu(A_2)=\nu(A_2)=\mathbb{P}(A_1)\mathbb{P}(A_2)$$

Now fix $A_2\in\sigma(\mathcal{A}_2)$, and repeat the argument with

$$\mu'(A)=\mathbb{P}(A\cap A_2)\quad\text{and}\quad\nu'(A)=\mathbb{P}(A)\mathbb{P}(A_2)$$

to show that for all $A_1 \in \sigma(\mathcal{A}_1)$, we have that

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

□

2.2 Borel-Cantelli lemmas

Definition 2.4 (lim inf and lim sup of events)
Given a sequence (A_n) of events, set

$$\liminf_n A_n = \bigcup_n \bigcap_{m \geq n} A_m \quad \text{and} \quad \limsup_n A_n = \bigcap_n \bigcup_{m \geq n} A_m$$

We write $\{A_n \text{ infinitely often}\} = \limsup_n A_n$ and $\{A_n \text{ eventually}\} = \liminf_n A_n$.

Lemma 2.5 (First Borel-Cantelli)**.** If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 0$.

Proof. As $n \to \infty$, we have that

$$\mathbb{P}(A_k \text{ i.o.}) \leq \mathbb{P}\left(\bigcup_{m \geq n} A_m\right) \leq \sum_{m \geq n} \mathbb{P}(A_m) \to 0$$

□

Lemma 2.6 (Second Borel-Cantelli)**.** If the events (A_n) are independent, and $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. We will use the inequality $1 - a \leq e^{-a}$. Note the events $(A_n^\complement)$ are also independent. Set $a_n = \mathbb{P}(A_n)$. Fix $n \in \mathbb{N}$, taking $n \leq N \to \infty$, we have that

$$\mathbb{P}\left(\bigcap_{m=n}^{N} A_m^\complement\right) = \prod_{m=n}^{N}(1 - a_n) \leq \exp\left(-\sum_{m=n}^{N} a_n\right) \to 0$$

Hence $\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^\complement\right) = 0$ for all n. Hence

$$\mathbb{P}(A_n \text{ i.o.}) = 1 - \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^\complement\right) = 1$$

□

2.3 Random variables

Definition 2.7 (random variable)
If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $(E, \mathcal{E})$ is a measurable space, then a measurable function $X : \Omega \to E$ is called a random variable in E. If $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ we just call X a random variable.

Definition 2.8 (law, distribution)
If X is a random variable with value in E, then the image measure $\mu_X = \mathbb{P} \circ X^{-1}$ is called the law or distribution of X.

Definition 2.9 (distribution function)
If X is a random variable, we define the distribution function of X by

$$F_X(x) = \mu_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

Proposition 2.10. The distribution function determines the distribution of a real valued random variable.

Proof. The set of intervals $(-\infty, x]$ forms a π-system, so we can use uniqueness of measures. □

Proposition 2.11. For a random variable X, the distribution function F_X is increasing and right continuous, with

$$\lim_{x \searrow -\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x \nearrow \infty} F_X(x) = 1$$

Conversely, any F satisfying these properties is the distribution function of a random variable.

Proof. For the converse, let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}((0, 1))$ and $\mathbb{P}$ be the Lebesgue measure, then

$$X(\omega) = \inf \{x \mid \omega \leq F(x)\}$$

defines a random variable with distribution function F. □

Definition 2.12 (independence)
A countable collection $(X_i)_{i \in I}$ of random variables is independent if the σ-algebras

$$\sigma(X_i^{-1}(A) \mid A \in \mathcal{E})$$

are independent.

Definition 2.13 (Rademacher functions)
Let $\Omega = (0, 1)$, then the Rademacher functions are defined by

$$R_n(\omega) = \omega_n$$

where we write

$$\omega = \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}$$

$\omega_n \in \{0, 1\}$ and we forbid infinitely many zeroes.

Proposition 2.14. The Rademacher functions are independent Ber(1/2) random variables.

Proposition 2.15. There exists a sequence (Y_n) of iid Unif$[0, 1]$ random variables.

Proof. Fix a bijection $m : \mathbb{N}^2 \to \mathbb{N}$, and define new random variables $Y_{n,k} = R_{m(n,k)}$, which are still independent. Define random variables

$$Y_n = \sum_k \frac{Y_{n,k}}{2^k}$$

which is a bounded monotone sequence so converges. Furthermore, then Y_n are still independent. Finally, to determine the distribution of Y_n, consider the π-system of intervals $(i/2^n, (i+1)/2^n]$ which generate the Borel σ-algebra. Then

$$\mathbb{P}\left(Y_n \in (i/2^n, (i+1)/2^n]\right) = 1/2^n$$

□

Corollary 2.16. Let (F_n) be a sequence of distribution functions, then there exists a sequence of independent random variables (X_n) with distribution functions (F_n).

2.4 Convergence of random variables

In the special case of probability spaces, we give different names to some of the concepts.

- almost surely := almost everywhere
- converges in probability := converges in measure

Definition 2.17 (Convergence in distribution)
A sequence X_n of random variables converges to X in distribution if

$$F_{X_n}(x) \to F_X(x)$$

for all $x \in \mathbb{R}$ where F_X is continuous.

2.5 Tail events

Definition 2.18 (Tail σ-algebra)
Let (X_n) be a sequence of random variables, define

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$$

and the tail σ-algebra is

$$\mathcal{T} = \bigcap_n \mathcal{T}_n$$

Theorem 2.19 (Kolmogorov's zero-one law)**.** Suppose (X_n) is a sequence of independent random variables, then the tail σ-algebra $\mathcal{T}$ contains only events of probability 0 or 1. Moreover, if $Y : (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B})$ is measurable, then Y is constant almost surely.

Proof. Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, then $\mathcal{F}_n$ is generated by the π-system of events

$$A = \{X_1 \leq x_1, \dots, X_n \leq x_n\}$$

whereas $\mathcal{T}_n$ is generated by the π-system of events,

$$B = \{X_{n+1} \leq x_{n+1}, \dots, X_{n+k} \leq x_{n+k}\} \quad \text{for} \quad k \in \mathbb{N}$$

By independence, we have that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all such A, B. Thus, $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent. Now $\bigcup_n \mathcal{F}_n$ is a π-system which generates the σ-algebra $\mathcal{F}_\infty = \sigma(X_n \mid n \in \mathbb{N})$. Thus, $\mathcal{F}_\infty$ and $\mathcal{T}$ are independent. But $\mathcal{T} \subseteq \mathcal{F}_\infty$, so if $A \in \mathcal{T}$,

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

so $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Finally, if Y is $\mathcal{T}$-measurable, then

$$\mathbb{P}(Y = c) = 1$$

where $c = \inf \{y \mid F_Y(y) = 1\}$. □

3 Integration

3.1 Definition of the integral

Definition 3.1 (simple function)

A simple function is a function of the form

$$f = \sum_{k=1}^{m} a_k 1_{A_k}$$

where $0 \leq a_k < \infty$ and A_k measurable for all k.

We define the integral of a simple function to be

$$\int f \mathrm{d}\mu = \mu(f) = \sum_{k=1}^{m} a_k \mu(A_k)$$

where $0 \cdot \infty = 0$. For a nonnegative measurable function f, define

$$\int_f \mathrm{d}\mu = \sup \left\{ \int g \mathrm{d}\mu \mid 0 \leq g \leq f, g \text{ simple} \right\}$$

Definition 3.2 (positive and negative part)

For a function f, define

$$f^+ = \max\{f, 0\} \quad \text{and} \quad f^- = \max\{-f, 0\}$$

Proposition 3.3.

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-$$

and if f is measurable, so is f^+, f^-.

Definition 3.4 (integrable)

A function $f : E \to \mathbb{R}$ is integrable if $\mu(|f|) < \infty$.

For an integrable f, define

$$\int f \mathrm{d}\mu = \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu$$

3.2 Monotone convergence theorem

Theorem 3.5 (monotone convergence)**.** Let f be a nonnegative measurable function, (f_n) a sequence of nonnegative measurable functions. Suppose $f_n \nearrow f$. Then $\mu(f_n) \nearrow \mu(f)$.

Proof. We perform a sequence of approximations.

Step 1: $f_n = 1_{A_n}, f = 1_A$. In this case, the proof is obvious from the axioms of a measure.

Step 2: f_n simple, $f = 1_A$. Fix $\varepsilon > 0$, and set $A_n = \{f_n > 1 - \varepsilon\}$. Then $A_n \nearrow A$, and

$$(1-\varepsilon)1_{A_n} \leq f_n \leq 1_A$$

so we have that

$$(1-\varepsilon)\mu(A_n) \leq \mu(f_n) \leq \mu(A)$$

But $A_n \nearrow A$ and $\varepsilon > 0$ was arbitrary.

Step 3: f_n and f simple. Write

$$f = \sum_{k=1}^{m} a_k 1_{A_k}$$

with $a_k > 0$ for all k and the A_k disjoint. Then $f_n \nearrow f$ implies that

$$a_k^{-1} 1_{A_k} f_n \nearrow 1_{A_k}$$

so by step 2,

$$\mu(f_n) = \sum_k \mu(1_{A_k} f_n) \nearrow \sum_k a_k \mu(A_k) = \mu(f)$$

Step 4: f_n simple, $f \geq 0$ measurable. Let g be simple with $g \leq f$. Then $f_n \nearrow f$ implies that $f_n \wedge g \nearrow g$, so by step 3,

$$\mu(f_n) \geq \mu(f_n \wedge g) \nearrow \mu(g)$$

Since g was arbitrary we are done.

Step 5: $f_n, f \geq 0$ measurable. Set $g_n = (2^{-n}\lfloor 2^n f_n \rfloor) \wedge n$, then g_n is simple and $g_n \leq f_n \leq f$, so

$$\mu(g_n) \leq \mu(f_n) \leq \mu(f)$$

But $f_n \nearrow f$ forces $g_n \nearrow f$, so $\mu(g_n) \nearrow f$ by step 4. Hence $\mu(f_n) \nearrow \mu(f)$. □

Proposition 3.6.

(i) for f, g nonnegative measurable, $\alpha, \beta \geq 0$, we have

$$\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$$

(ii) for $f \leq g$ nonnegative measurable, $\mu(f) \leq \mu(g)$,

(iii) $\mu(f) = 0$ if and only if $f = 0$ a.e.

Proof. (i) Define simple functions f_n, g_n by

$$f_n = (2^{-n}\lfloor 2^n f \rfloor) \wedge n \quad \text{and} \quad g_n = (2^{-n}\lfloor 2^n g \rfloor) \wedge n$$

Then $f_n \nearrow f$ and $g_n \nearrow g$, so $\alpha f_n + \beta g_n \nearrow \alpha f + \beta g$. By the monotone convergence theorem, $\mu(f_n) \nearrow \mu(f)$, $\mu(g_n) \nearrow \mu(g)$ and $\mu(\alpha f_n + \beta g_n) \nearrow \mu(\alpha f + \beta g)$. But $\mu(\alpha f_n + \beta g_n) = \alpha\mu(f_n) + \beta\mu(g_n)$, so we are done.

(ii) is immediate from the definition of the integral.

(iii) If $f = 0$ a.e. then $f_n = 0$ a.e., so $\mu(f) = 0$. Conversely, if $\mu(f) = 0$ then $\mu(f_n) = 0$ for all n, i.e. $f_n = 0$ a.e. for all n, so $f = 0$ a.e. □

Theorem 3.7. Let $f, g : E \to \mathbb{R}$ be integrable, then

(i) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$,

(ii) $g \leq f$ implies $\mu(g) \leq \mu(f)$,

(iii) $f = 0$ a.e. implies $\mu(f) = 0$.

Proof. Follows from the case for nonnegative functions. □

3.3 Fatou and dominated convergence

Lemma 3.8 (Fatou)**.** Let $f_n : E \to \mathbb{R}$ be measurable nonnegative functions. Then

$$\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n)$$

Proof. For $k \geq n$, we have that

$$\inf_{m \geq n} f_m \leq f_k$$

which means that

$$\mu\left(\inf_{m \geq n} f_m\right) \leq \inf_{k \geq n} \mu(f_k) \leq \liminf_{\ell} \mu(f_\ell)$$

But as $n \to \infty$,

$$\inf_{m \geq n} f_m \nearrow \liminf_{\ell} f_\ell$$

so by monotone convergence,

$$\mu\left(\inf_{m \geq n} f_m\right) \nearrow \mu\left(\liminf_{\ell} f_\ell\right)$$

□

Theorem 3.9 (dominated convergence)**.** Let f, f_n be measurable functions, $f_n \to f$ pointwise, and there exists g integrable such that $|f_n| \leq g$. Then f and f_n are integrable, with $\mu(f_n) \to \mu(f)$.

Proof. As $|f_n| \leq g$, each f_n is integrable. Furthermore, $|f| \leq g$, so f is integrable as well. Furthermore, $0 \leq g \pm f_n \to g \pm f$, so $\liminf(g \pm f_n) = g \pm f$. By Fatou's lemma, we have that

$$\begin{aligned}
\mu(g) + \mu(f) &= \mu\left(\liminf_n (g + f_n)\right) \leq \liminf_n \mu(g + f_n) = \mu(g) + \liminf_n \mu(f_n) \\
\mu(g) - \mu(f) &= \mu\left(\liminf_n (g - f_n)\right) \leq \liminf_n \mu(g - f_n) = \mu(g) - \limsup_n \mu(f_n)
\end{aligned}$$

Since $\mu(g) < \infty$, we get that

$$\mu(f) \leq \liminf_n \mu(f_n) \leq \limsup_n \mu(f_n) \leq \mu(f)$$

So $\lim \mu(f_n) = \mu(f)$. □

Theorem 3.10 (differentiation under the integral)**.** Let $U \subseteq \mathbb{R}$ be open, $(E, \mathcal{E}, \mu)$ measure space, $f : U \times E \to \mathbb{R}$ such that

(i) $f(t, \cdot)$ is measurable for all t,

(ii) $f(\cdot, x)$ is differentiable for all x, with a μ-integrable g such that

$$\left| \frac{\partial f(t,x)}{\partial t} \right| \leq g(x) \quad \text{for all} \quad t \in U$$

Define

$$F(t) = \int_E f(t,x) \mathrm{d}x$$

Then F is differentiable, with derivative

$$F'(t) = \int_E \frac{\partial f(t,x)}{\partial t} \mathrm{d}x$$

Proof. Set

$$g_h(x) = \frac{f(t+h,x) - f(t,x)}{h} - \frac{\partial f}{\partial t}(t,x)$$

By the mean value theorem, we have that

$$g_h(x) = \frac{\partial f}{\partial t}(\tilde{t}, x) - \frac{\partial f}{\partial t}(t,x)$$

for some $\tilde{t} \in (t, t+h)$ (or $(t+h, t)$ if $h < 0$). In particular, this means that $|g_h| \leq 2g$, so g_h is integrable for all h. Furthermore, $g_h \to 0$ as $h \to 0$, so applying the dominated convergence theorem, we have that $\mu(g_h) \to 0$. But

$$\left| \frac{F(t+h) - F(t)}{h} - F'(t) \right| = \left| \int_E \frac{f(t+h,x) - f(t,x)}{h} - \frac{\partial f}{\partial t}(t,x) \mathrm{d}x \right| \to 0$$

□

3.4 Densities and image measure

Definition 3.11 (density)

Suppose $f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ measurable and nonnegative, then we can define a new measure

$$\nu_f(A) = \mu(f 1_A)$$

Proposition 3.12. For any $g : E \to \mathbb{R}$ measurable, we have that

$$\nu_f(g) = \mu(fg)$$

Therefore, we call f the density of ν_f with respect to μ.

Proposition 3.13. For $f : E \to G$ measurable, $g : G \to \mathbb{R}$ nonnegative measurable, we have that

$$\mu \circ f^{-1}(g) = \mu(g \circ f)$$

Proposition 3.14. If $g : G \to \mathbb{R}$ measurable, X is a G-valued random variable, then

$$\mathbb{E}(g(X)) = \int_\Omega g(X(\omega)) \mathrm{d}\mathbb{P}(\omega)$$

3.5 Products

Definition 3.15 (product σ-algebra)
Let $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$, then define the π-system of rectangles

$$\mathcal{A} = \{A_1 \times A_2 \mid A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$$

and the product σ-algebra

$$\mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(\mathcal{A})$$

Proposition 3.16. If E_1, E_2 are second countable Hausdorff spaces, $\mathcal{E}_i = \mathcal{B}(E_i)$ are the Borel σ-algebras, then

$$\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$$

where we give $\mathcal{E}_1 \times \mathcal{E}_2$ the product topology.

Lemma 3.17. For fixed $x_2 \in E_2$, the canonical injection $\iota : E_1 \hookrightarrow E_1 \times E_2$ is measurable, where $\iota(x_1) = (x_1, x_2)$. Furthermore, the canonical projections $\pi : E_1 \times E_2 \twoheadrightarrow E_1$ are measurable as well.

Proof. For ι, suffices to check on the generating π-system. Let $A_1 \times A_2 \in \mathcal{A}$, then $\iota^{-1}(A_1 \times A_2) = A_1 \in \mathcal{E}_1$. For π, note that $\pi^{-1}(A_1) = A_1 \times E_2 \in \mathcal{A} \subseteq \mathcal{E}_1 \otimes \mathcal{E}_2$. □

Lemma 3.18. Let f be a bounded (resp. nonnegative) measurable function on $E_1 \times E_2$, where μ_2 is a finite measure on E_2. Define for $x_1 \in E_1$,

$$f_1(x_1) = \int_{E_2} f(x_1, x_2) \mathrm{d}\mu_2(x_2)$$

Then f_1 is measurable. If f is bounded (resp. nonnegative), then so is f_1.

Proof. In the bounded case, define a vector space

$$\mathcal{V} = \left\{ f : E \to \mathbb{R} \mid f \text{ bounded measurable, } \int_{E_2} f(\cdot, x_2) \mathrm{d}\mu_2(x_2) \text{ bounded measurable} \right\}$$

Then $1_E \in \mathcal{V}$[1], and $1_A \in \mathcal{V}$ for any $A \in \mathcal{A}$. Now take $0 \le f_n \nearrow f$, where the $f_n \in \mathcal{V}$. By the monotone convergence theorem,

$$\int_{E_2} f(x_1, x_2) \mathrm{d}\mu_2(x_2) = \lim_{n \to \infty} \int_{E_2} f_n(x_1, x_2) \mathrm{d}\mu_2(x_2)$$

This is $\mathcal{E}_1$-measurable and bounded as it is the limit of bounded measurable functions. So $f \in \mathcal{V}$. By the monotone class theorem, $\mathcal{V}$ contains all bounded measurable functions.

In the nonnegative case, set $f_n = f \wedge n$, use the bounded case and the monotone convergence theorem. □

[1] As $\mu_w(E_2) < \infty$.

Theorem 3.19 (product measure). Suppose $(E_1, \mathcal{E}_1, \mu_1)$, $(E_2, \mathcal{E}_2, \mu_2)$ be finite measure spaces, then there exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

for all $A_1 \in \mathcal{E}_1$, $A_2 \in \mathcal{E}_2$.

Proof. Since we defined μ on a π-system generating the σ-algebra, suffices to show that it is a well defined measure. Define

$$\mu(A) = \int_{E_1} \left(\int_{E_2} 1_A(x_1, x_2) \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1)$$

Then

$$\mu(A_1 \times A_2) = \int_{E_1} \left(\int_{E_2} 1_{A_1}(x_1) 1_{A_2}(x_2) \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) = \mu_1(A_1)\mu_2(A_2)$$

and $\mu(\varnothing) = 0$, so suffices to show μ is countably additive. Let (A_n) be a disjoint sequence of elements of $\mathcal{E}_1 \otimes \mathcal{E}_2$. Then

$$1_{\bigcup_n A_n} = \sum_n 1_{A_n} = \lim_{N \to \infty} \sum_{n=1}^{N} 1_{A_n}$$

Thus, we have that

$$\begin{aligned}
\mu\left(\bigcup_n A_n\right) &= \int_{E_1} \left(\int_{E_2} \lim_{N\to\infty} \sum_{n=1}^{N} 1_{A_n}(x_1, x_2) \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \int_{E_1} \left(\int_{E_2} 1_{A_n}(x_1, x_2) \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \\
&= \sum_n \mu(A_n)
\end{aligned}$$

where we swap the limit and integrals by the monotone convergence theorem. □

Theorem 3.20 (Fubini). Let $(E, \mathcal{E}, \mu) = (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$, where $\mu_i(E_i) < \infty$. Then

(i) Let f be nonnegative measurable, then

$$\mu(f) = \int_E f \mathrm{d}\mu = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) = \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2)$$

(ii) Suppose f is μ-integrable, then

$$\mu_1 \left(\left\{ x_1 \in E_1 \mid \int_{E_2} f(x_1, x_2) \mathrm{d}\mu_2(x_2) = \infty \right\} \right) = 0$$

and

$$\int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) = \int_E f(x) \mathrm{d}\mu(x)$$

Proof. (i) We proceed by a series of approximations. By definition of μ on rectangles, we see that the result holds in the case $f = 1_A$. By linearity of the integral, the result holds for f simple. Taking an approximating sequence $0 \le f_1 \le \cdots \le f$, we get the result using the monotone convergence theorem.

(ii) Define $h(x_1) = \int_{E_2} |f(x_1, x_2)| \mathrm{d}\mu_2(x_2)$, then by (i), $\mu_1(h) < \mu(f) < \infty$, so h is integrable, hence h is finite a.e. The final part follows from splitting into the positive and negative parts, and using (i). □

Remark 3.21. The results in this subsection generalises to σ-finite measures by splitting into finite parts.

Proposition 3.22. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E}) = (\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i)$. Consider $X : \Omega \to E$, $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$, then the following are equivalent:

(i) the X_i are independent,

(ii) $\mu_X = \bigotimes_{i=1}^n \mu_{X_i}$,

(iii) for all $f_i : E_i \to \mathbb{R}$, bounded measurable,

$$\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}\left(f_i(X_i)\right)$$

Proof. (i) $\implies$ (ii). For rectangles $A_1 \times \cdots \times A_n$, we have that

$$\mu_X(A_1 \times \dots A_n) = \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i) = \prod_{i=1}^n \mu_{X_i}(A_i)$$

Result follows by uniqueness of a measure on the generating π-system.
(ii) $\implies$ (iii). By Fubini's theorem, we have that

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) &= \mu_X\left(\prod_{i=1}^n f_i(x_i)\right) \\
&= \int_E \prod_{i=1}^n f_i(x_i)\, \mathrm{d}(\mu_1 \otimes \dots \mu_n)(x_1, \dots, x_n) \\
&= \int_{E_1} f_1(x_1) \mathrm{d}\mu_1(x_1) \cdots \int_{E_n} f_n(x_n) \mathrm{d}\mu_n(x_n) \\
&= \prod_{i=1}^n \mathbb{E}\left(f_i(X_i)\right)
\end{aligned}$$

(iii) $\implies$ (i). Take $f_i = 1_{A_i}$, then

$$\mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n) = \prod_{i=1}^n \mathbb{E}\left(1_{A_i}\right) = \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n)$$

So $X_1, \dots, X_n$ are independent. ∎

4 L^p spaces

Definition 4.1 (L^p-norm)
For a measure space $(E, \mathcal{E}, \mu)$, and $1 \le p < \infty$, define the L^p-norm of f by

$$\left\|f\right\|_p = \left(\int_E |f(x)|^p \mathrm{d}x\right)^{1/p}$$

Define the L^∞ norm by

$$\left\|f\right\|_\infty = \operatorname*{ess\,sup}_{x \in E} |f(x)| = \inf\left\{\lambda \mid |f| \le \lambda \text{ a.e.}\right\}$$

Definition 4.2 (L^p and $\mathcal{L}^p$ spaces)
For $1 \le p \le \infty$, define the L^p space by

$$L^p(\mu) = L^p(E, \mathcal{E}, \mu) = \left\{ f : E \to \mathbb{R} \mid \|f\|_p < \infty \right\}$$

and the corresponding $\mathcal{L}^p$ space by

$$\mathcal{L}^p = \frac{L^p}{f = g \text{ a.e.}}$$

4.1 Inequalities

Proposition 4.3 (Chebyshev). Suppose f nonnegative measureable, $\lambda \ge 0$, then

$$\lambda \mu(f \ge \lambda) \le \mu(f)$$

Proof. $\lambda 1_{f \ge \lambda} \le f$ pointwise, integrating gives the required result. □

Corollary 4.4 (tail estimate). Suppose $f \in L^p$ with $p < \infty$, then

$$\mu(|g| \ge \lambda) = \mathcal{O}\left(\lambda^{-p}\right)$$

as $p \to \infty$.

Proof.

$$\mu(|g| \ge \lambda) = \mu(|g|^p \ge \lambda^p) \le \lambda^{-p}\mu(|g|^p) < \infty$$

□

Lemma 4.5. Let $I \subseteq \mathbb{R}$ be an interval, $c : I \to \mathbb{R}$ be convex, $m \in \text{Int}(I)$. Then there exists $a, b \in \mathbb{R}$ such that $c(x) \ge ax + b$, with equality at $x = m$.

Proof. For $m, x, y \in I$, with $x < m < y$, we have

$$\frac{c(m) - c(x)}{m - x} \le \frac{c(y) - c(m)}{y - m}$$

So there exists $a \in \mathbb{R}$ such that for all $x < m$ and $y < m$, we have that

$$\frac{c(m) - c(x)}{m - x} \le a \le \frac{c(y) - c(m)}{y - m}$$

Then $c(x) \ge a(x - m) + c(m)$ for all $x \in I$. □

Theorem 4.6 (Jensen's inequality). Let X be an integrable random variable with values in an interval I, $c : I \to \mathbb{R}$ be convex. Then $\mathbb{E}(c(X))$ is well defined, and

$$\mathbb{E}\left(c(X)\right) \ge c\left(\mathbb{E}(X)\right)$$

Proof. The case where X is almost everywhere constant is easy. Otherwise, $m = \mathbb{E}(X)$ must be in the interior of I. Choose a, b as in the lemma. Then $c(X) \ge aX + b$, so $\mathbb{E}\left(c(X)^{\pm}\right) \le |a|\mathbb{E}(X) + |b| < \infty$, so $c(X)$ is integrable. Moreover,

$$\mathbb{E}\left(c(X)\right) \ge a\mathbb{E}(x) + b = am + b = cm = c\left(E(X)\right)$$

□

Theorem 4.7 (Hölder's inequality). Let $p, q \in (1, \infty)$ be conjugate indices, then for all measurable functions f, g, we have

$$\mu(|fg|) \leq \|f\|_p \|g\|_q$$

Proof. The case when $\|f\|_p = 0$ or $\|f\|_p = \infty$ is clear. Then without loss of generality, we can assume $\|f\|_p = 1$. Define a probability measure $\mathbb{P}$ on $\mathcal{E}$ by

$$\mathbb{P}(A) = \int_A |f|^p \mathrm{d}\mu$$

For measurable functions $X \geq 0$,

$$\mathbb{E}(X) = \mu(X|f|^p) \quad \text{and} \quad \mathbb{E}(X) \leq \mathbb{E}(X^q)^{1/q}$$

Now $q(p-1) = p$, so

$$\begin{aligned}
\mu(|fg|) &= \mu\left(\frac{|g|}{|f|^{p-1}} 1_{\{|f|>0\}} |f|^p\right) \\
&= \mathbb{E}\left(\frac{|g|}{|f|^{p-1}} 1_{\{|f|>0\}}\right) \\
&= \mathbb{E}\left(\frac{|g|^q}{|f|^p} 1_{\{|f|>0\}}\right) \\
&\leq \mu(|g|^q)^{1/q} \\
&= \|g\|_q
\end{aligned}$$

□

Theorem 4.8 (Minkowski).

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p$$

Proof. The cases when $p = 1, p = \infty$, $\|f\|_p = \infty$, $\|g\|_p = \infty$ or $\|f+g\|_p = 0$ are clear. Otherwise, since $|f+g|^p \leq 2^p(|f|^p + |g|^p)$, we have

$$\mu(|f+g|^p) \leq 2^p \mu(|f|^p + |g|^p) < \infty$$

As

$$\left\||f+g|^{p-1}\right\|_q = \mu\left(|f+g|^p\right)^{1/q} = \mu(|f+g|^p)^{1-1/p}$$

By Hölder's inequality we have that

$$\mu(|f+g|^p) \leq \mu(|f|\|f+g\|^{p-1}) + \mu(|g|\|f+g\|^{p-1}) \leq (\|f\|_p + \|g\|_p)\left\||f+g|^{p-1}\right\|_q$$

As $\left\||f+g|^{p-1}\right\|_q > 0$, dividing through we are done. □

4.2 Banach and Hilbert space structure

Theorem 4.9. $\mathcal{L}^p$ is complete.

Proof. The case $p = \infty$ is clear. From now on, assume $p < \infty$, then choose a subsequence (n_k) such that

$$S = \sum_{k=1}^{\infty} \left\| f_{k+1} - f_k \right\|_p < \infty$$

By Minkowski's inequality, for any $K \in \mathbb{N}$,

$$\left\| \sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}| \right\|_p \leq S < \infty$$

By monotone convergence, the result also holds for $K = \infty$. Thus we have that

$$\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty \text{ a.e.}$$

So by completeness of $\mathbb{R}$, (f_{n_k}) converges a.e. Define f by

$$f(x) = \begin{cases} \lim_k f_{n_k}(x) & \text{if the limit exists} \\ 0 & \text{otherwise} \end{cases}$$

Now given $\varepsilon > 0$, we can find N such that for all $n \geq N$,

$$\mu(|f_n - f_m|^p) \leq \varepsilon \quad \text{for all} \quad m \geq n$$

In particular, $\mu(|f_n - f_{n_k}|^p) \leq \varepsilon$ for sufficiently large k. Thus by Fatou's lemma, for $n \geq N$,

$$\mu(|f_n - f|^p) = \mu\left(\liminf_k |f_n - f_{n_k}|^p\right) \leq \liminf_k \mu(|f_n - f_{n_k}|^p) \leq \varepsilon$$

Hence $f \in L^p$ by Minkowski, and as $\varepsilon > 0$ was arbitrary, $f_n \to f$. □

Theorem 4.10. $\mathcal{L}^2$ is a Hilbert space.

Proof. Since we have already shown completeness, all we need to do is define the inner product. In this case, we define

$$\langle f, g \rangle = \int_E f g \, \mathrm{d}\mu$$

Note however in the complex valued case we will need to put in an appropriate complex conjugate. □

Corollary 4.11 (Pythagoras, parallelogram).

$$\left\| f + g \right\|_2^2 = \left\| f \right\|_2^2 + \left\| g \right\|_2^2 + 2\langle f, g \rangle$$

and

$$\left\| f + g \right\|_2^2 + \left\| f - g \right\|_2^2 = 2(\left\| f \right\|_2^2 + \left\| g \right\|_2^2)$$

Theorem 4.12 (orthogonal projection). Let V be a closed subspace of L^2, then for each $f \in L^2$, there exists $v \in V, u \in V^\perp$ such that $f = v + u$. Moreover,

$$\left\| f - v \right\|_2 \leq \left\| f - g \right\|_2 \quad \text{for all} \quad g \in V$$

with equality if and only if $g = v$ a.e.

Proof. Choose a sequence $(g_n) \subseteq V$ such that

$$\|f - g_n\|_2 \to d(f, V) = \inf \{\|f - g\|_2 \mid g \in V\}$$

By the parallelogram law,

$$\|2(f - (g_n + g_m)/2)\|_2^2 + \|g_n - g_m\|_2^2 = 2(\|f - g_n\|_2^2 + \|f - g_m\|_2^2)$$

But $\|2(f - (g_n + g_m)/2)\|_2^2 \geq 4d(f, V)^2$, so we must have $\|g_n - g_m\|_2^2 \to 0$ as $n, m \to \infty$, i.e. (g_n) is Cauchy. So $g_n \to g$ by completeness, where $g = v$ a.e. for some $v \in V$ as V is closed. Hence

$$\|f - v\|_2 = \lim_n \|f - g_n\|_2 = d(f, V)$$

Now for any $h \in V$ and $t \in \mathbb{R}$, we have that

$$d(f, V)^2 \leq \|f - (v + th)\|_2^2 = d(f, V)^2 - 2t\langle f - v, h\rangle + t^2\|h\|_2^2$$

So we must have $\langle f - v, h\rangle = 0$, i.e. $f - v \in V^\perp$. □

4.3 Convergence in $L^1(\mathbb{P})$ and uniform integrability

Theorem 4.13 (bounded convergence theorem)**.** Let (X_n) be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|X_n| \leq C < \infty$ and $X_n \to X$ in probability. Then $X_n \to X$ in $L^1(\mathbb{P})$.

Proof. We know that X_n converges almost surely along a subsequence, so $|X| \leq C$ almost surely. For $\varepsilon > 0$, there exists N such that for all $n \geq N$,

$$\mathbb{P}(|X_n - X| > \varepsilon/2) \leq \varepsilon/(4C)$$

Then

$$\mathbb{E}|X_n - X| = \mathbb{E}\left(|X_n - X|1_{|X_n - X| > \varepsilon/2}\right) + \mathbb{E}\left(|X_n - X|1_{|X_n - X| \leq \varepsilon/2}\right) \leq 2C(\varepsilon/(4C)) + \varepsilon/2 = \varepsilon$$

□

Lemma 4.14. Let X be an integrable random variable, and set

$$I_X(\delta) = \sup \{\mathbb{E}(|X|1_A) \mid A \in \mathcal{F}, \mathbb{P}(A) < \delta\}$$

Proof. Suppose not. Then there exists $\varepsilon > 0$, $(A_n) \subseteq \mathcal{F}$ such that $\mathbb{P}(A_n) \leq 2^{-n}$ and $\mathbb{E}(||1_{A_n}) \geq \varepsilon$ for all n. Thus by the first Borel-Cantelli lemma, $\mathbb{P}(A_n \text{ i.o.}) = 0$. But then by dominated convergence theorem,

$$\varepsilon \leq \mathbb{E}\left(|X|1_{\bigcup_{m \geq n} A_m}\right) \to \mathbb{E}(|X|1_{A_n \text{ i.o.}}) = 0$$

Contradiction. □

Definition 4.15 (uniformly integrable)

A collection $\mathcal{X} \subseteq L^1(\mathbb{P})$ is uniformly integrable if it is bounded in $L^1(\mathbb{P})$ and

$$I_{\mathcal{X}}(\delta) = \sup \{\mathbb{E}(|X|1_A) \mid X \in \mathcal{X}, A \in \mathcal{F}, \mathbb{P}(A) < \delta\} \to 0 \quad \text{as} \quad \delta \to 0$$

Lemma 4.16. If $\mathcal{X}$ is bounded in L^p for some $1 < p < \infty$, then $\mathcal{X}$ is uniformly integrable.

Proof. By Hölder's inequality,

$$\mathbb{E}\left(|X|1_A\right) \le \|X\|_p \mathbb{P}(A)^{1/q}$$

□

Lemma 4.17. Let $\mathcal{X}$ be a family of random variables, then $\mathcal{X}$ is uniformly integrable if and only if

$$\sup\left\{\mathbb{E}\left(|X|1_{|X|\ge K} \mid X \in \mathcal{X}\right)\right\} \to 0 \quad \text{as} \quad K \to \infty$$

Proof. Suppose $\mathcal{X}$ is uniformly integrable. Then given $\varepsilon > 0$, choose $\delta > 0$ such that $I_{\mathcal{X}}(\delta) < \varepsilon$, and choose $K < \infty$ such that $I_{\mathcal{X}}(1) < K\delta$. Then for $X \in \mathcal{X}$, $A = \{|X| \ge K\}$, we have that $\mathbb{P}(A) \le \delta$ by Chebyshev's inequality, so $\mathbb{E}\left(|X|1_A\right) < \varepsilon$. Hence as $K \to \infty$,

$$\sup\left\{\mathbb{E}\left(|X|1_{|X|\ge K}\right) \mid X \in \mathcal{X}\right\} \to 0$$

On the other hand, if this condition holds, then since

$$\mathbb{E}(|X|) \le K + \mathbb{E}\left(|X|1_{|X|\ge K}\right)$$

we have that $I_{\mathcal{X}}(1) < \infty$. Now given $\varepsilon > 0$, choose $K < \infty$ such that $\mathbb{E}\left(|X|1_{|X|\ge K}\right) < \varepsilon/2$ for all $X \in \mathcal{X}$. Then choose $\delta > 0$ such that $K\delta < \varepsilon/2$. Now for all $X \in \mathcal{X}$ and $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$, we have that

$$\mathbb{E}\left(|X|1_A\right) \le \mathbb{E}\left(|X|1_{|X|\ge K}\right) + K\mathbb{P}(A) < \varepsilon$$

Hence $\mathcal{X}$ is uniformly integrable. □

Theorem 4.18. Let X be a random variable, (X_n) a sequence of random variables. Then the following are equivalent.

(i) X_n and X are integrable, with $X_n \to X$ in L^1,

(ii) $\{X_n\}$ is uniformly integrable and $X_n \to X$ in probability.

Proof. Suppose (i) holds. By Chebyshev's inequality, for all $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \varepsilon^{-1}\mathbb{E}(|X_n - X|) \to 0$$

So $X_n \to X$ in probability. Moreover, given $\varepsilon > 0$, there exists N such that $\mathbb{E}(|X_n - X|) < \varepsilon/2$ for all $n \ge N$. Furthermore, as any finite set is uniformly integrable (by the dominated convergence theorem), we can find $\delta > 0$ such that $\mathbb{P}(A) \le \delta$ implies

$$\mathbb{E}\left(|X|1_A\right) \le \varepsilon/2 \quad \text{and} \quad \mathbb{E}\left(|X_n|1_A\right) \le \varepsilon \text{ for } n = 1, \dots, N$$

Then for $n \ge N$, and $\mathbb{P}(A) \le \delta$, we have that

$$\mathbb{E}(|X_n|1_A) \le \mathbb{E}(|X_n - X|1_A) + \mathbb{E}(|X|1_A) \le \mathbb{E}(|X_n - X|) + \mathbb{E}(|X|1_A) \le \varepsilon$$

Hence $\{X_n\}$ is uniformly integrable.

Now suppose (ii) holds. Then there exists a subsequence (n_k) such that $X_{n_k} \to X$ almost surely. So by Fatou's lemma,

$$\mathbb{E}(|X|) \le \liminf_k \mathbb{E}(|X_{n_k}|) < \infty$$

Now given $\varepsilon > 0$, there exists $K < \infty$ such that for all n,

$$\mathbb{E}(|X_n|1_{|X_n|\ge K}) < \varepsilon/3 \quad \text{and} \quad \mathbb{E}(|X|1_{|X|\ge K}) < \varepsilon/3$$

Now consider the uniformly bounded sequence $X_n^K = -K \vee X_n \wedge K$ and set $X^K = -K \vee X \wedge K$. Then $X_n^K \to X^K$ in probability, so by bounded convergence, there exists N such that for all $n \ge N$,

$$\mathbb{E}\left|X_n^K - X^K\right| < \varepsilon/3$$

But then for all $n \geq N$,

$$\mathbb{E}|X_n - X| \leq \mathbb{E}(|X|1_{|X_n| \geq K}) + \mathbb{E}\left|X_n^K - X^K\right| + \mathbb{E}(|X|1_{|X| \geq K}) < \varepsilon$$

□

5 Fourier transforms

In this section, for $p < \infty$, we write $L^p = L^p(\mathbb{R}^d)$ for some d fixed, for the set of *complex values* Borel measurable functions on $\mathbb{R}^d$ with finite p-norm.

5.1 Definitions

Definition 5.1 (Fourier transform of functions)
The Fourier transform of $f \in L^1$ is defined to be

$$\hat{f}(u) = \int_{\mathbb{R}^d} f(x) e^{iu \cdot x} \mathrm{d}x \quad \text{for} \quad u \in \mathbb{R}^d$$

Proposition 5.2. For $f \in L^1$, $\hat{f}$ is a countinuous bounded function on $\mathbb{R}^d$.

Proof. Continuity follows from the dominated convergence theorem. For boundedness, notice that

$$\sup_{u \in \mathbb{R}^n} \left|\hat{f}(u)\right| \leq \|f\|_1$$

□

Definition 5.3 (Fourier transform of measure)
Let μ be a finite Borel measure on $\mathbb{R}^d$, then the Fourier transform of μ is

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{iu \cdot x} \mathrm{d}\mu(x)$$

Proposition 5.4. For μ a finite Borel measure on $\mathbb{R}^d$, $\hat{\mu}$ is a continuous bounded function on $\mathbb{R}^d$. Furthermore, if μ has density f with respect to Lebesgue measure, then $\hat{\mu} = \hat{f}$.

Proof. Again continuity follows from the dominated convergence theorem. In this case, we have

$$\sup_{u \in \mathbb{R}^n} |\hat{\mu}(u)| \leq \mu(\mathbb{R}^d)$$

□

Definition 5.5 (Characteristic function)
The characteristic function ϕ_X of a random variable X in $\mathbb{R}^d$ is the Fourier transform of its law μ_X. That is,

$$\phi_X(u) = \hat{\mu_X}(u) = \mathbb{E}(e^{iu \cdot X})$$

Definition 5.6 (Convolution of functions)
For $f \in L^p$ and a probability measure ν on $\mathbb{R}^d$, define the convolution $f * \nu$ by

$$f * \nu(x) \int_{\mathbb{R}^d} f(x-y) \mathrm{d}\nu(y)$$

if the integral exists, and $f * \nu(x) = 0$ otherwise.

Proposition 5.7. The integral defining the convolution exists a.e., and we have that

$$\|f * \nu\|_p \leq \|f\|_p$$

Proof. By Jensen's inequality and Fubini,

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)| \mathrm{d}\nu(y) \right)^p \mathrm{d}x &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)|^p \mathrm{d}\nu(y) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)|^p \mathrm{d}x \mathrm{d}\nu(y) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x)|^p \mathrm{d}x \mathrm{d}\nu(y) \\
&= \|f\|_p^p
\end{aligned}$$

□

Definition 5.8 (Convolution of probability measures)
Suppose X, Y independent random variables with laws μ, ν respectively. Define $\mu * \nu$ to be the density of $X + Y$, namely

$$\mu * \nu(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_A(x+y) \mathrm{d}\mu(x) \mathrm{d}\nu(y)$$

Proposition 5.9. If μ has density f with respect to the Lebesgue measure, then $\mu * \nu$ has density $f * \nu$ with respect to the Lebesgue measure.

Proof. Fubini. □

Proposition 5.10.

(i) $\widehat{f * \nu} = \hat{f} \hat{\nu}$

(ii) $\widehat{\mu * \nu} = \hat{\mu} \hat{\nu}$

5.2 Gaussian convolutions and Fourier inversion

Definition 5.11 (Centred Gaussian)
For $t > 0$, define the centred Gaussian probability density on $\mathbb{R}^d$ with variance t by

$$g_t(x) = (2\pi t)^{-d/2} \exp\left(\frac{-||x||^2}{2t}\right)$$

Proposition 5.12.

$$\hat{g}_t(u) = \exp\left(\frac{-||u||^2 t}{2}\right) = \left(\frac{2\pi}{t}\right)^{d/2} g_{1/t}(u)$$

Proof. Let Z be a standard one-dimensional Gaussian random variable. Then as Z is integrable, ϕ_Z is differentiable, and we can differentiate under the integral sign to obtain

$$\phi_Z'(u) = \mathbb{E}\left(iZe^{iuZ}\right) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} ixe^{-x^2/2} \mathrm{d}x = -u\phi_Z(u)$$

Solving the differential equation we obtain that $\phi_Z(u) = e^{-u^2/2}$. Now consider d standard normal random variables $Z_1, \ldots, Z_d$, and set $Z = (Z_1, \ldots, Z_d)$. Then $\sqrt{t}Z$ has density g_t. So

$$\hat{g}_t(u) = \mathbb{E}\left(e^{iu\cdot\sqrt{t}Z}\right) = \mathbb{E}\left(\prod_{j=1}^{d} e^{iu_j\sqrt{t}Z_j}\right) = \prod_{j=1}^{d} \phi_Z(u_j\sqrt{t}) = e^{-||u||^2 t/2}$$

□

Corollary 5.13. The Fourier inversion formula holds for centred Gaussian densities, that is,

$$g_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{g}_t(u) e^{-iu\cdot x} \mathrm{d}u$$

Proposition 5.14. The Fourier inversion formula holds for all Gaussian convolutions, that is,

$$(f * g_t)(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f * g_t}(u) e^{-iu\cdot x} \mathrm{d}u$$

Proof. We use Fourier inversion for g_t and Fubini's theorem

$$\begin{aligned}
(2\pi)^d f * g_t(u) &= (2\pi)^d \int_{\mathbb{R}^d} f(x-y) g_t(y) \mathrm{d}y \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x-y) \hat{g}_t(u) e^{-iu\cdot y} \mathrm{d}u\mathrm{d}y \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x-y) e^{iu\cdot(x-y)} \hat{g}_t(u) e^{-iu\cdot x} \mathrm{d}u\mathrm{d}y \\
&= \int_{\mathbb{R}^d} \hat{f}(u) \hat{g}_t(u) e^{-iu\cdot x} \mathrm{d}u \\
&= \int_{\mathbb{R}^d} \widehat{f * g_t}(u) e^{-iu\cdot x} \mathrm{d}u
\end{aligned}$$

□

Lemma 5.15. Let $f \in L^p$ with $p < \infty$, then $\left\|f * g_t - f\right\|_p \to 0$ as $t \to 0$.

Proof. Given $\varepsilon > 0$, there exists $h \in C_c(\mathbb{R}^d)$ such that $\|f - h\|_p < \varepsilon$. Then

$$\|f * g_t - h * g_t\|_p = \|(f-h) * g_t\|_p \le \|f - h\|_p \le \varepsilon$$

. Thus, by a 3ε argument, suffices to prove the result for h. Set

$$e(y) = \int_{\mathbb{R}^d} |h(x-y) - h(x)|^p \mathrm{d}x$$

Then $|e(y)| \le 2^p \|h\|_p^p$ and e is continuous at 0 by the dominated convergence theorem. By Jensen's inequality and bounded convergence theorem,

$$\begin{aligned}
\|h * g_t - h\|_p^p &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (h(x-y) - h(x)) g_t(y) \mathrm{d}y \right|^p \mathrm{d}x \\
&\le \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |h(x-y) - h(x)|^p g_t(y) \mathrm{d}y \mathrm{d}x \\
&= \int_{\mathbb{R}^d} e(y) g_t(y) \mathrm{d}y \\
&= \int_{\mathbb{R}^d} e(\sqrt{t} y) g_1(y) \mathrm{d}y \to 0
\end{aligned}$$

as $t \to 0$. □

Theorem 5.16. Let $f \in L^1$, define for $t > 0$,

$$f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-\|u\| t/2} e^{-iu\cdot x} \mathrm{d}u$$

Then $\|f_t - f\|_1 \to 0$ as $t \to 0$. Moreover, if $\hat{f} \in L^1$, then

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-iu\cdot x} \mathrm{d}u$$

Proof. By Fourier inversion of Gaussian convolutions, $f_t = f * g_t$, so the convergence result follows from the previous lemmas. Suppose $\hat{f} \in L^1$. Then by the dominated convergence theorem, for all $x \in \mathbb{R}^d$, as $t \to 0$,

$$f_t(x) \to \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(u) e^{-iu\cdot x} \mathrm{d}u$$

But $f_{t_n} \to f$ a.e. for some subsequence $t_n \to 0$. Hence the Fourier inversion formula holds for f. □

5.3 Fourier-Plancherel

Theorem 5.17. Suppose $f \in L^1 \cap L^2$. Then $\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2$.

Proof. If $f, \hat{f} \in L^1$, then Fourier inversion formula holds and $f, \hat{f} \in L^\infty$, and $(x, u) \mapsto f(x)\hat{f}(u)$ is integrable on $\mathbb{R}^d \times \mathbb{R}^d$. So by Fubini's theorem, we get

$$
\begin{aligned}
(2\pi)^2\|f\|_2^2 &= \int_{\mathbb{R}^d} f(x)\overline{f(x)}\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \hat{f}(u)e^{-iu\cdot x}\mathrm{d}u \right) \overline{f(x)}\mathrm{d}x \\
&= \int_{R^d} \hat{f}(u) \overline{\left(\int_{\mathbb{R}^d} f(x)e^{iu\cdot x}\mathrm{d}x \right)} \mathrm{d}u \\
&= \int_{\mathbb{R}^d} \hat{f}(u)\overline{\hat{f}(u)}\mathrm{d}u \\
&= \left\|\hat{f}\right\|_2^2
\end{aligned}
$$

Now suppose $f \in L^1 \cap L^2$, and let $f_t = f * g_t$. By lemma, $f_t \to f$ in L^2, so $\|f_t\|_2 \to \|f\|_2$. Furthermore, $\hat{f}_t = \hat{f}\hat{g}_t$ and $\hat{g}_t(u) = e^{-\|u\|^2 t/2}$. Hence $\left\|\hat{f}_t\right\|_2 \nearrow \left\|\hat{f}\right\|_2$ by monotone convergence. But $f_t, \hat{f}_t \in L^1$ so we are done. □

Corollary 5.18. We can extend the (rescaled) Fourier transform to a unique isometry $\mathcal{F} : L^2 \to L^2$.

Proof. $L^1 \cap L^2$ is dense in L^2. □

5.4 Weak convergence and characteristic functions

Definition 5.19 (weak convergence of measures)
Let μ be a Borel probability measure on $\mathbb{R}^d$, (μ_n) a sequence of such measures. We say that $\mu_n \to \mu$ weakly if $\mu_n(f) \to \mu(f)$ for all continuous bounded functions f on $\mathbb{R}^d$.

Note that if we consider the space of (signed) measures as the dual space to the Banach space of continuous bounded functions, then "weak convergence of measures" is in fact weak-$*$ convergence in the dual space sense. However, in probability theory weak convergence is not as useful, so convention is that this is called "weak convergence".

Definition 5.20 (weak convergence of random variables)
Given a random variable X in $\mathbb{R}^d$, and a sequence of random variables (X_n) on $\mathbb{R}^d$, we say that $X_n \to X$ weakly if $\mu_{X_n} \to \mu_X$ weakly.

Theorem 5.21. Let X be a random variable on $\mathbb{R}^d$. Then the distribution μ_X of X is uniquely determined by its characteristic function ϕ_X. Furthermore, if $\phi_X \in L^1$, then μ_X has density f_X with respect to Lebesgue measure, where

$$
f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u)e^{-iu\cdot x}\mathrm{d}u
$$

Proof. Let Z be a standard Gaussian in $\mathbb{R}^d$ independent of X. Then $\sqrt{t}Z$ has density g_t, and $X + \sqrt{t}Z$ has density $f_t = \mu_X * g_t$. Then $\hat{f}_t(u) = \phi_X(u)e^{-\|u\|t/2}$, so by Fourier inversion formula,

$$
f_t(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(u)e^{-\|u\|^2 t/2}e^{-iu\cdot x}\mathrm{d}u
$$

By bounded convergence, for any g,

$$
\int_{\mathbb{R}^d} g(x)f_t(x)\mathrm{d}x = \mathbb{E}(g(X + \sqrt{t}Z)) \to \mathbb{E}(g(X)) = \int_{\mathbb{R}^d} g(x)\mathrm{d}\mu_X(x)
$$

Hence ϕ_X determines μ_X. The density statement follows from the dominated convergence theorem. □

Theorem 5.22 (Levy continuity). Suppose X_n, X random variables on $\mathbb{R}^d$ such that $\phi_{X_n} \to \phi_X$ pointwise on $\mathbb{R}^d$. Then $X_n \to X$ weakly.

Proof. By density, suffices to show that $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ for all $g \in C_c^\infty(\mathbb{R}^d)$. Fix $g \in C_c^\infty$, and let $C = \|g'\|_\infty$. Then g is C-Lipschitz. Fix $\varepsilon > 0$, choose $t > 0$ such that $\sqrt{t} C \mathbb{E}|X| \le \varepsilon$. On the other hand, by Fourier inversion and dominated convergence theorem, we have that

$$\begin{aligned}
\mathbb{E}(g(X_n + \sqrt{t}Z)) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x)\phi_{X_n}(u) e^{-\|u\|^2 t/2} e^{-ix\cdot u} \mathrm{d}u\mathrm{d}x \\
&\to \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x)\phi_{X}(u) e^{-\|u\|^2 t/2} e^{-ix\cdot u} \mathrm{d}u\mathrm{d}x \\
&= \mathbb{E}(g(X + \sqrt{t}Z))
\end{aligned}$$

Hence $|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| < \varepsilon$ for sufficiently n. □

6 Ergodic theory

Definition 6.1 (measure preserving transformation)
Suppose $(E, \mathcal{E}, \mu)$ is a σ-finite measure space, $\theta : E \to E$ measurable. Then θ is called measure preserving if

$$\mu(\theta^{-1}(A)) = \mu(A)$$

for all $A \in \mathcal{E}$.

Proposition 6.2. For all $f \in L^1(\mu)$, $\mu(f) = \mu(f \circ \theta)$.

Definition 6.3 (invariant function)
A measurable map $f : E \to \mathbb{R}$ is θ-invariant if $f \circ \theta = f$.

Definition 6.4 (invariant set)
A set $A \in \mathcal{E}$ is θ-invariant if $\theta^{-1}(A) = A$. The collection $\mathcal{E}_\theta$ of all θ-invariant sets forms a σ-algebra.

Proposition 6.5. f is θ-invariant if and only if f is $\mathcal{E}_\theta$-measurable.

Definition 6.6 (ergodic)
A measure preserving transformation θ is ergodic if for all $A \in \mathcal{E}_\theta$, $\mu(A) = 0$ or $\mu(A^{\complement}) = 0$.

Proposition 6.7. Suppose θ is ergodic and f is θ-invariant. Then f is a.e. constant.

6.1 Birkhoff and von Neumann ergodic theorems

Throughout this subsection, let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space, with a measure preserving transformation θ. Given a measureable function f, set $S_0 = 0$ and $S_n = f + f \circ \theta + \cdots + f \circ \theta^{n-1}$.

Lemma 6.8 (maximal ergodic)**.** Let $S^* = \sup_{n\geq 0} S_n$. Then

$$\int_{S^*>0} f \mathrm{d}\mu \geq 0$$

Proof. Set $S_n^* = \max_{0\leq m\leq n} S_m$ and $A_n = \{S_n^* > 0\}$. Then for $m = 1, \dots, n$,

$$S_m = f + S_{m-1} \circ \theta \leq f + S_n^* \circ \theta$$

On A_n, we have $S_n^* = \max_{1\leq m\leq n} S_m$, so $S_n^* \leq f + S_n^* \circ \theta$, and on $A_n^\complement$, we have $S_n^* = 0 \leq S_n^* \circ \theta$. Hence we have that

$$\int_E S_n^* \mathrm{d}\mu \leq \int_{A_n} f \mathrm{d}\mu + \int_E S_n^* \circ \theta \mathrm{d}\mu = \int_{A_n} f \mathrm{d}\mu + \int_E S_n^* \mathrm{d}\mu < \infty$$

which forces

$$\int_{A_n} f \mathrm{d}\mu \geq 0$$

As $A_n \nearrow \{S^* > 0\}$, the result follows by dominated convergence theorem. □

Theorem 6.9 (Birkhoff)**.** Suppose $f \in L^1(\mu)$. Then there exists a θ-invariant $\bar{f} \in L^1(\mu)$ such that

$$\frac{S_n}{n} \to f \text{ a.e.}$$

and $\left\|\bar{f}\right\|_1 \leq \|f\|_1$.

Proof. Non-examinable and omitted. □

Theorem 6.10 (von Neumann)**.** Suppose $\mu(E) < \infty$, and $p < \infty$. Then for all $f \in L^p$, $S_n/n \to \bar{f}$ in L^p.

Proof. We have that $\left\|f \circ \theta^m\right\|_p = \left\|f\right\|_p$ for all m, so by the triangle inequality, $\left\|S_n/n\right\|_p \leq \left\|f\right\|_p$. Fix $\varepsilon > 0$, then choose K such that $\left\|f - g\right\| < \varepsilon$, where $g = (-K) \vee f \wedge K$. By Birkhoff's theorem, $S_n(g)/n \to \bar{g}$ a.e. We have that $|S_n(g)/n| \leq K$ for all n, so by bounded convergence, there exists N such that for all $n \geq N$,

$$\left\|S_n(g)/n - \bar{g}\right\|_p < \varepsilon$$

By Fatou's lemma,

$$\begin{aligned}
\left\|\bar{f} - \bar{g}\right\|_p^p &= \int_E \liminf_n |S_n(f-g)/n|^p \mathrm{d}\mu \\
&\leq \liminf_n \int_E |S_n(f-g)/n|^p \mathrm{d}\mu \\
&\leq \left\|f - g\right\|_p^p
\end{aligned}$$

Thus for $n \geq N$,

$$\left\|S_n(f)/n - \bar{f}\right\|_p < 3\varepsilon$$

□

6.2 Shift map

In this section, let $E = \mathbb{R}^{\mathbb{N}}$ with the product topology. Then the Borel σ-algebra is generated by the coordinate projections $\pi_n : E \to \mathbb{R}$, and it is also generated by the π-system

$$\mathcal{C} = \left\{ A = \prod_{n=1}^{\infty} A_n \mid A_n \in \mathcal{B}, A_n = \mathbb{R} \text{ for all but finitely many } n \right\}$$

Let (X_n) be an iid sequence of random variables with distribution m, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then define $X : \Omega \to E$ by

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

which is measureable. Let $\mu = \mathbb{P} \circ X^{-1}$ be the image measure. Then μ is the unique measure such that

$$\mu\left(\prod_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{\infty} m(A_i)$$

Now define the shift map $\theta : E \to E$, $\theta(x) = (x_2, x_3, \dots)$.

Theorem 6.11. The shift map is measure preserving and ergodic.

Proof. For $A \in \mathcal{C}$, we have that

$$\begin{aligned}
\mu(A) &= \mathbb{P}(X_1 \in A_1, \dots, X_N \in A_N) \\
&= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_N \in A_N) \\
&= \mathbb{P}(X_2 \in A_1) \cdots \mathbb{P}(X_{N+1} \in A_N) \\
&= \mu(\theta^{-1}(A))
\end{aligned}$$

Thus by uniqueness, $\mu = \mu \circ \theta^{-1}$, so θ is measure preserving.

Let $\mathcal{T}_n = \sigma(X_{n+1}, \dots)$ and $\mathcal{T} = \bigcap_n \mathcal{T}_n$ be the tail σ-algebra. Then for $A \in \sigma(\mathcal{C})$, $(\theta^n)^{-1}(A) \in \mathcal{T}_n$. Thus, if A is invariant, then $A = (\theta^n)^{-1}(A) \in \mathcal{T}_n$ for all n. Hence by the Kolmogorov zero-one law, $\mu(A) = 0$ or $\mu(A) = 1$. □

7 Limit theorems

Theorem 7.1 (central limit)**.** Let (X_i) be iid random variables with mean 0 and variance 1. Then

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \rightharpoonup Z \sim N(0, 1)$$

That is,

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \to \mathbb{P}(Z \leq x)\right)$$

for all $x \in \mathbb{R}$.

Proof. Set $X = X_1$, then ϕ_X satisfies $\phi_X(0) = 1$, $\phi_X'(u) = i\mathbb{E}(Xe^{iuX})$ and $\phi_X''(u) = -\mathbb{E}(X^2 e^{iuX})$. In particular, $\phi'(0) = 0$ and $\phi''(1) = -1$. Then Taylor's theorem implies that

$$\phi_X(v) = 1 - \frac{v^2}{2} + o(v^2)$$

Thus, if we let ϕ_n be the characteristic function of $(X_1 + \cdots + X_n)/\sqrt{n}$, then by independence $\phi_n(u) = \left(\phi_X(u/\sqrt{n})\right)^n$ by independence.

$$\phi_n(u) = \left(1 - \frac{u^2}{2n} + o(1/n)\right)^n$$

The complex logarithm satisfiex $\log(1+z) = z + o(z)$, so we get that

$$\log(\phi_n(u)) = n\log\big(1 - u^2/(2n) + o(1/n)\big) \to -u^2/2 \quad \text{as} \quad n \to \infty$$

By continuity, $\phi_n(u) \to e^{-u^2/2}$, so the result follows by Levy continuity. □

Theorem 7.2. Suppose $\int_{\mathbb{R}} |x| \mathrm{d}m(x) < \infty$. Set $\nu = \int_{\mathbb{R}} x \mathrm{d}m(x)$. Then

$$\mu\left(x \in \mathbb{R}^{\mathbb{N}} \mid \lim_{n\to\infty}\left(\frac{x_1 + \cdots + x_n}{n}\right) = \nu\right) = 1$$

Proof. Set $f(x) = x_1$, which is in $L^1(\mu)$. So by the ergodic theorems with $\theta = \mathrm{id}$,

$$\mu\left(\frac{x_1 + \cdots + x_n}{n} \to \nu\right) = \mu\left(\frac{S_n}{n} \to \nu\right)$$

By Birkhoff, we have that $S_n/n \to \bar{f}$ a.s. By von Neumann,

$$\bar{f} = \mu(\bar{f}) = \lim_{n\to\infty} \mu\left(\frac{S_n}{n}\right) = \mu(f) = \nu$$

□

Theorem 7.3 (strong law of large numbers)**.** Suppose (X_i) is an iid sequence of integrable random variables. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}X_i$$

almost surely.

Proof. Inject $X : \Omega \to E$ as before, and notice that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}X_i\right) = \mu\left(x \mid \frac{x_1 + \cdots + x_n}{n} \to \nu\right) = 1$$

□

Remark 7.4. By the von Neumann ergodic theorem, the previous two theorems can be strengthened to L^1 convergence.