Part II — Probability and Measure
Theorems with proof

Based on lectures by J. Miller
Notes taken by Dexter Chua

Michaelmas 2016

Analysis II is essential

Measure spaces, σ-algebras, π-systems and uniqueness of extension, statement *and proof* of Carathéodory's extension theorem. Construction of Lebesgue measure on $\mathbb{R}$. The Borel σ-algebra of $\mathbb{R}$. Existence of non-measurable subsets of $\mathbb{R}$. Lebesgue-Stieltjes measures and probability distribution functions. Independence of events, independence of σ-algebras. The Borel–Cantelli lemmas. Kolmogorov's zero-one law. [6]

Measurable functions, random variables, independence of random variables. Construction of the integral, expectation. Convergence in measure and convergence almost everywhere. Fatou's lemma, monotone and dominated convergence, differentiation under the integral sign. Discussion of product measure and statement of Fubini's theorem. [6]

Chebyshev's inequality, tail estimates. Jensen's inequality. Completeness of L^p for $1 \leq p \leq \infty$. The Hölder and Minkowski inequalities, uniform integrability. [4]

L^2 as a Hilbert space. Orthogonal projection, relation with elementary conditional probability. Variance and covariance. Gaussian random variables, the multivariate normal distribution. [2]

The strong law of large numbers, proof for independent random variables with bounded fourth moments. Measure preserving transformations, Bernoulli shifts. Statements *and proofs* of maximal ergodic theorem and Birkhoff's almost everywhere ergodic theorem, proof of the strong law. [4]

The Fourier transform of a finite measure, characteristic functions, uniqueness and inversion. Weak convergence, statement of Lévy's convergence theorem for characteristic functions. The central limit theorem. [2]

Contents

0 Introduction

1 Measures

1.1 Measures

Proposition. A collection $\mathcal{A}$ is a σ-algebra if and only if it is both a π-system and a d-system.

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

Proof. Let $\mathcal{D}$ be the intersection of all d-systems containing $\mathcal{A}$, i.e. the smallest d-system containing $\mathcal{A}$. We show that $\mathcal{D}$ contains $\sigma(\mathcal{A})$. To do so, we will show that $\mathcal{D}$ is a π-system, hence a σ-algebra.

There are two steps to the proof, both of which are straightforward verifications:

(i) We first show that if $B \in \mathcal{D}$ and $A \in \mathcal{A}$, then $B \cap A \in \mathcal{D}$.

(ii) We then show that if $A, B \in \mathcal{D}$, then $A \cap B \in \mathcal{D}$.

Then the result immediately follows from the second part.

We let

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

We note that $\mathcal{D}' \supseteq \mathcal{A}$ because $\mathcal{A}$ is a π-system, and is hence closed under intersections. We check that $\mathcal{D}'$ is a d-system. It is clear that $E \in \mathcal{D}'$. If we have $B_1, B_2 \in \mathcal{D}'$, where $B_1 \subseteq B_2$, then for any $A \in \mathcal{A}$, we have

$$(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A).$$

By definition of $\mathcal{D}'$, we know $B_2 \cap A$ and $B_1 \cap A$ are elements of $\mathcal{D}$. Since $\mathcal{D}$ is a d-system, we know this intersection is in $\mathcal{D}$. So $B_2 \setminus B_1 \in \mathcal{D}'$.

Finally, suppose that (B_n) is an increasing sequence in $\mathcal{D}'$, with $B = \bigcup B_n$. Then for every $A \in \mathcal{A}$, we have that

$$\left(\bigcup B_n\right) \cap A = \bigcup (B_n \cap A) = B \cap A \in \mathcal{D}.$$

Therefore $B \in \mathcal{D}'$.

Therefore $\mathcal{D}'$ is a d-system contained in $\mathcal{D}$, which also contains $\mathcal{A}$. By our choice of $\mathcal{D}$, we know $\mathcal{D}' = \mathcal{D}$.

We now let

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

Since $\mathcal{D}' = \mathcal{D}$, we again have $\mathcal{A} \subseteq \mathcal{D}''$, and the same argument as above implies that $\mathcal{D}''$ is a d-system which is between $\mathcal{A}$ and $\mathcal{D}$. But the only way that can happen is if $\mathcal{D}'' = \mathcal{D}$, and this implies that $\mathcal{D}$ is a π-system. $\square$

Theorem (Caratheodory extension theorem)**.** Let $\mathcal{A}$ be a ring on E, and μ a countably additive set function on $\mathcal{A}$. Then μ extends to a measure on the σ-algebra generated by $\mathcal{A}$.

Proof. (non-examinable) We start by defining what we want our measure to be. For $B \subseteq E$, we set

$$\mu^*(B) = \inf \left\{ \sum_n \mu(A_n) : (A_n) \in \mathcal{A} \text{ and } B \subseteq \bigcup A_n \right\}.$$

If it happens that there is no such sequence, we set this to be ∞. This measure is known as the *outer measure*. It is clear that $\mu^*(\phi) = 0$, and that μ^* is increasing.

We say a set $A \subseteq E$ is μ^*-measurable if

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^C)$$

for all $B \subseteq E$. We let

$$\mathcal{M} = \{\mu^*\text{-measurable sets}\}.$$

We will show the following:

(i) $\mathcal{M}$ is a σ-algebra containing $\mathcal{A}$.

(ii) μ^* is a measure on $\mathcal{M}$ with $\mu^*|_{\mathcal{A}} = \mu$.

Note that it is not true in general that $\mathcal{M} = \sigma(A)$. However, we will always have $M \supseteq \sigma(\mathcal{A})$.

We are going to break this up into five nice bite-size chunks.

Claim. μ^* is countably subadditive.

Suppose $B \subseteq \bigcup_n B_n$. We need to show that $\mu^*(B) \leq \sum_n \mu^*(B_n)$. We can wlog assume that $\mu^*(B_n)$ is finite for all n, or else the inequality is trivial. Let $\varepsilon > 0$. Then by definition of the outer measure, for each n, we can find a sequence $(B_{n,m})_{m=1}^\infty$ in $\mathcal{A}$ with the property that

$$B_n \subseteq \bigcup_m B_{n,m}$$

and

$$\mu^*(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu(B_{n,m}).$$

Then we have

$$B \subseteq \bigcup_n B_n \subseteq \bigcup_{n,m} B_{n,m}.$$

Thus, by definition, we have

$$\mu^*(B) \leq \sum_{n,m} \mu^*(B_{n,m}) \leq \sum_n \left(\mu^*(B_n) + \frac{\varepsilon}{2^n}\right) = \varepsilon + \sum_n \mu^*(B_n).$$

Since ε was arbitrary, we are done.

Claim. μ^* agrees with μ on $\mathcal{A}$.

In the first example sheet, we will show that if $\mathcal{A}$ is a ring and μ is a countably additive set function on μ, then μ is in fact countably subadditive and increasing.

Assuming this, suppose that $A, (A_n)$ are in $\mathcal{A}$ and $A \subseteq \bigcup_n A_n$. Then by subadditivity, we have

$$\mu(A) \leq \sum_n \mu(A \cap A_n) \leq \sum_n \mu(A_n),$$

using that μ is countably subadditivity and increasing. Note that we have to do this in two steps, rather than just applying countable subadditivity, since we did not assume that $\bigcup_n A_n \in \mathcal{A}$. Taking the infimum over all sequences, we have

$$\mu(A) \leq \mu^*(A).$$

Also, we see by definition that $\mu(A) \geq \mu^*(A)$, since A covers A. So we get that $\mu(A) = \mu^*(A)$ for all $A \in \mathcal{A}$.

Claim. $\mathcal{M}$ contains $\mathcal{A}$.

Suppose that $A \in \mathcal{A}$ and $B \subseteq E$. We need to show that

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^C).$$

Since μ^* is countably subadditive, we immediately have $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^C)$. For the other inequality, we first observe that it is trivial if $\mu^*(B)$ is infinite. If it is finite, then by definition, given $\varepsilon > 0$, we can find some (B_n) in $\mathcal{A}$ such that $B \subseteq \bigcup_n B_n$ and

$$\mu^*(B) + \varepsilon \geq \sum_n \mu(B_n).$$

Then we have

$$\begin{aligned} B \cap A &\subseteq \bigcup_n (B_n \cap A)\\ B \cap A^C &\subseteq \bigcup_n (B_n \cap A^C) \end{aligned}$$

We notice that $B_n \cap A^C = B_n \setminus A \in \mathcal{A}$. Thus, by definition of μ^*, we have

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum_n \mu(B_n \cap A) + \sum_n \mu(B_n \cap A^C)\\ &= \sum_n (\mu(B_n \cap A) + \mu(B_n \cap A^C))\\ &= \sum_n \mu(B_n)\\ &\leq \mu^*(B_n) + \varepsilon. \end{aligned}$$

Since ε was arbitrary, the result follows.

Claim. We show that $\mathcal{M}$ is an algebra.

We first show that $E \in \mathcal{M}$. This is true since we obviously have

$$\mu^*(B) = \mu^*(B \cap E) + \mu^*(B \cap E^C)$$

for all $B \subseteq E$.

Next, note that if $A \in \mathcal{M}$, then by definition we have, for all B,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^C).$$

Now note that this definition is symmetric in A and A^C. So we also have $A^C \in M$.

Finally, we have to show that $\mathcal{M}$ is closed under intersection (which is equivalent to being closed under union when we have complements). Suppose $A_1, A_2 \in \mathcal{M}$ and $B \subseteq E$. Then we have

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^C)\\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^C) + \mu^*(B \cap A_1^C)\\
&= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^C \cap A_1)\\
&\quad + \mu^*(B \cap (A_1 \cap A_2)^C \cap A_1^C)\\
&= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^C).
\end{aligned}$$

So we have $A_1 \cap A_2 \in \mathcal{M}$. So $\mathcal{M}$ is an algebra.

Claim. $\mathcal{M}$ is a σ-algebra, and μ^* is a measure on $\mathcal{M}$.

To show that $\mathcal{M}$ is a σ-algebra, we need to show that it is closed under countable unions. We let (A_n) be a disjoint collection of sets in $\mathcal{M}$, then we want to show that $A = \bigcup_n A_n \in \mathcal{M}$ and $\mu^*(A) = \sum_n \mu^*(A_n)$.

Suppose that $B \subseteq E$. Then we have

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^C)$$

Using the fact that $A_2 \in \mathcal{M}$ and $A_1 \cap A_2 = \emptyset$, we have

$$\begin{aligned}
&= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^C \cap A_2^C)\\
&= \cdots\\
&= \sum_{i=1}^n \mu^*(B \cap A_i) + \mu^*(B \cap A_1^C \cap \cdots \cap A_n^C)\\
&\geq \sum_{i=1}^n \mu^*(B \cap A_i) + \mu^*(B \cap A^C).
\end{aligned}$$

Taking the limit as $n \to \infty$, we have

$$\mu^*(B) \geq \sum_{i=1}^\infty \mu^*(B \cap A_i) + \mu^*(B \cap A^C).$$

By the countable-subadditivity of μ^*, we have

$$\mu^*(B \cap A) \leq \sum_{i=1}^\infty \mu^*(B \cap A_i).$$

Thus we obtain

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^C).$$

By countable subadditivity, we also have inequality in the other direction. So equality holds. So $A \in \mathcal{M}$. So $\mathcal{M}$ is a σ-algebra.

To see that μ^* is a measure on $\mathcal{M}$, note that the above implies that

$$\mu^*(B) = \sum_{i=1}^{\infty}(B \cap A_i) + \mu^*(B \cap A^C).$$

Taking $B = A$, this gives

$$\mu^*(A) = \sum_{i=1}^{\infty}(A \cap A_i) + \mu^*(A \cap A^C) = \sum_{i=1}^{\infty} \mu^*(A_i). \qquad \square$$

Theorem. Suppose that μ_1, μ_2 are measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$. If $\mathcal{A}$ is a π-system with $\sigma(A) = \mathcal{E}$, and μ_1 agrees with μ_2 on $\mathcal{A}$, then $\mu_1 = \mu_2$.

Proof. Let

$$\mathcal{D} = \{A \in \mathcal{E} : \mu_1(A) = \mu_2(A)\}$$

We know that $\mathcal{D} \supseteq \mathcal{A}$. By Dynkin's lemma, it suffices to show that $\mathcal{D}$ is a d-system. The things to check are:

(i) $E \in \mathcal{D}$ — this follows by assumption.

(ii) If $A, B \in \mathcal{D}$ with $A \subseteq B$, then $B \setminus A \in \mathcal{D}$. Indeed, we have the equations

$$\mu_1(B) = \mu_1(A) + \mu_1(B \setminus A) < \infty$$
$$\mu_2(B) = \mu_2(A) + \mu_2(B \setminus A) < \infty.$$

Since $\mu_1(B) = \mu_2(B)$ and $\mu_1(A) = \mu_2(A)$, we must have $\mu_1(B \setminus A) = \mu_2(B \setminus A)$.

(iii) Let $(A_n) \in \mathcal{D}$ be an increasing sequence with $\bigcup A_n = A$. Then

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2(A).$$

So $A \in \mathcal{D}$. $\square$

Theorem. There exists a unique Borel measure μ on $\mathbb{R}$ with $\mu([a, b]) = b - a$.

Proof. We first show uniqueness. Suppose $\tilde{\mu}$ is another measure on $\mathcal{B}$ satisfying the above property. We want to apply the previous uniqueness theorem, but our measure is not finite. So we need to carefully get around that problem.

For each $n \in \mathbb{Z}$, we set

$$\mu_n(A) = \mu(A \cap (n, n + 1]))$$
$$\tilde{\mu}_n(A) = \tilde{\mu}(A \cap (n, n + 1]))$$

Then μ_n and $\tilde{\mu}_n$ are finite measures on $\mathcal{B}$ which agree on the π-system of intervals of the form $(a, b]$ with $a, b \in \mathbb{R}$, $a < b$. Therefore we have $\mu_n = \tilde{\mu}_n$ for all $n \in \mathbb{Z}$. Now we have

$$\mu(A) = \sum_{n\in\mathbb{Z}} \mu(A \cap (n, n + 1]) = \sum_{n\in\mathbb{Z}} \mu_n(A) = \sum_{n\in\mathbb{Z}} \tilde{\mu}_n(A) = \tilde{\mu}(A)$$

for all Borel sets A.

To show existence, we want to use the Caratheodory extension theorem. We let $\mathcal{A}$ be the collection of finite, disjoint unions of the form

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n].$$

Then $\mathcal{A}$ is a ring of subsets of R, and $\sigma(\mathcal{A}) = \mathcal{B}$ (details are to be checked on the first example sheet).

We set

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

We note that μ is well-defined, since if

$$A = (a_1, b_1] \cup \cdots \cup (a_n, b_n] = (\tilde{a}_1, \tilde{b}_1] \cup \cdots \cup (\tilde{a}_n, \tilde{b}_n],$$

then

$$\sum_{i=1}^{n} (b_i - a_i) = \sum_{i=1}^{n} (\tilde{b}_i - \tilde{a}_i).$$

Also, if μ is additive, $A, B \in \mathcal{A}$, $A \cap B = \emptyset$ and $A \cup B \in \mathcal{A}$, we obviously have $\mu(A \cup B) = \mu(A) + \mu(B)$. So μ is additive.

Finally, we have to show that μ is in fact countably additive. Let (A_n) be a disjoint sequence in $\mathcal{A}$, and let $A = \bigcup_{i=1}^{\infty} A_n \in \mathcal{A}$. Then we need to show that $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$.

Since μ is additive, we have

$$\begin{aligned}
\mu(A) &= \mu(A_1) + \mu(A \setminus A_1)\\
&= \mu(A_1) + \mu(A_2) + \mu(A \setminus A_1 \cup A_2)\\
&= \sum_{i=1}^{n} \mu(A_i) + \mu\left(A \setminus \bigcup_{i=1}^{n} A_i\right)
\end{aligned}$$

To finish the proof, we show that

$$\mu\left(A \setminus \bigcup_{i=1}^{n} A_i\right) \to 0 \text{ as } n \to \infty.$$

We are going to reduce this to the *finite intersection property* of compact sets in $\mathbb{R}$: if (K_n) is a sequence of compact sets in $\mathbb{R}$ with the property that $\bigcap_{m=1}^{n} K_m \neq \emptyset$ for all n, then $\bigcap_{m=1}^{\infty} K_m \neq \emptyset$.

We first introduce some new notation. We let

$$B_n = A \setminus \bigcup_{m=1}^{n} A_m.$$

We now suppose, for contradiction, that $\mu(B_n) \not\to 0$ as $n \to \infty$. Since the B_n's are decreasing, there must exist $\varepsilon > 0$ such that $\mu(B_n) \geq 2\varepsilon$ for every n.

For each n, we take $C_n \in \mathcal{A}$ with the property that $\overline{C_n} \subseteq B_n$ and $\mu(B_n \setminus C_n) \leq \frac{\varepsilon}{2^n}$. This is possible since each B_n is just a finite union of intervals. Thus we have

$$\begin{aligned}
\mu(B_n) - \mu\left(\bigcap_{m=1}^n C_m\right) &= \mu\left(B_n \setminus \bigcap_{m=1}^n C_m\right)\\
&\leq \mu\left(\bigcup_{m=1}^n (B_m \setminus C_m)\right)\\
&\leq \sum_{m=1}^n \mu(B_m \setminus C_m)\\
&\leq \sum_{m=1}^n \frac{\varepsilon}{2^m}\\
&\leq \varepsilon.
\end{aligned}$$

On the other hand, we also know that $\mu(B_n) \geq 2\varepsilon$.

$$\mu\left(\bigcap_{m=1}^n C_m\right) \geq \varepsilon$$

for all n. We now let that $K_n = \bigcap_{m=1}^n \overline{C_m}$. Then $\mu(K_n) \geq \varepsilon$, and in particular $K_n \neq \emptyset$ for all n.

Thus, the finite intersection property says

$$\emptyset \neq \bigcap_{n=1}^\infty K_n \subseteq \bigcap_{n=1}^\infty B_n = \emptyset.$$

This is a contradiction. So we have $\mu(B_n) \to 0$ as $n \to \infty$. So done. □

Proposition. The Lebesgue measure is *translation invariant*, i.e.

$$\mu(A + x) = \mu(A)$$

for all $A \in \mathcal{B}$ and $x \in \mathbb{R}$, where

$$A + x = \{y + x, y \in A\}.$$

Proof. We use the uniqueness of the Lebesgue measure. We let

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

for $A \in \mathcal{B}$. Then this is a measure on $\mathcal{B}$ satisfying $\mu_x([a, b]) = b - a$. So the uniqueness of the Lebesgue measure shows that $\mu_x = \mu$. □

Proposition. Let $\tilde{\mu}$ be a Borel measure on $\mathbb{R}$ that is translation invariant and $\mu([0, 1]) = 1$. Then $\tilde{\mu}$ is the Lebesgue measure.

Proof. We show that any such measure must satisfy

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

By additivity and translation invariance, we can show that $\mu([p, q]) = q - p$ for all rational $p < q$. By considering $\mu([p, p + 1/n])$ for all n and using the increasing property, we know $\mu(\{p\}) = 0$. So $\mu(([p, q)) = \mu((p, q]) = \mu((p, q)) = q - p$ for all rational p, q.

Finally, by countable additivity, we can extend this to all real intervals. Then the result follows from the uniqueness of the Lebesgue measure. □

1.2 Probability measures

Proposition. Events (A_n) are independent iff the σ-algebras $\sigma(\mathcal{A}_n)$ are independent.

Theorem. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are π-systems in $\mathcal{F}$. If

$$\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2]$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

Proof. This will follow from two applications of the fact that a finite measure is determined by its values on a π-system which generates the entire σ-algebra.

We first fix $A_1 \in \mathcal{A}_1$. We define the measures

$$\mu(A) = \mathbb{P}[A \cap A_1]$$

and

$$\nu(A) = \mathbb{P}[A]\mathbb{P}[A_1]$$

for all $A \in \mathcal{F}$. By assumption, we know μ and ν agree on $\mathcal{A}_2$, and we have that $\mu(\Omega) = \mathbb{P}[A_1] = \nu(\Omega) \leq 1 < \infty$. So μ and ν agree on $\sigma(\mathcal{A}_2)$. So we have

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

for all $A_2 \in \sigma(\mathcal{A}_2)$.

So we have now shown that if $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent, then $\mathcal{A}_1$ and $\sigma(\mathcal{A}_2)$ are independent. By symmetry, the same argument shows that $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent. $\square$

Lemma (Borel–Cantelli lemma)**.** If

$$\sum_n \mathbb{P}[A_n] < \infty,$$

then

$$\mathbb{P}[A_n \text{ i.o.}] = 0.$$

Proof. For each k, we have

$$\begin{aligned} \mathbb{P}[A_n \text{ i.o}] &= \mathbb{P}\left[\bigcap_n \bigcup_{m \geq n} A_m\right] \\ &\leq \mathbb{P}\left[\bigcup_{m \geq k} A_m\right] \\ &\leq \sum_{m=k}^{\infty} \mathbb{P}[A_m] \\ &\to 0 \end{aligned}$$

as $k \to \infty$. So we have $\mathbb{P}[A_n \text{ i.o.}] = 0$. $\square$

Lemma (Borel–Cantelli lemma II)**.** Let (A_n) be independent events. If

$$\sum_n \mathbb{P}[A_n] = \infty,$$

then

$$\mathbb{P}[A_n \text{ i.o.}] = 1.$$

Proof. By example sheet, if (A_n) is independent, then so is (A_n^C). Then we have

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{m=n}^{N} A_m^C\right] &= \prod_{m=n}^{N} \mathbb{P}[A_m^C]\\
&= \prod_{m=n}^{N} (1 - \mathbb{P}[A_m])\\
&\leq \prod_{m=n}^{N} \exp(-\mathbb{P}[A_m])\\
&= \exp\left(-\sum_{m=n}^{N} \mathbb{P}[A_m]\right)\\
&\to 0
\end{aligned}$$

as $N \to \infty$, as we assumed that $\sum_n \mathbb{P}[A_n] = \infty$. So we have

$$\mathbb{P}\left[\bigcap_{m=n}^{\infty} A_m^C\right] = 0.$$

By countable subadditivity, we have

$$\mathbb{P}\left[\bigcup_n \bigcap_{m=n}^{\infty} A_m^C\right] = 0.$$

This in turn implies that

$$\mathbb{P}\left[\bigcap_n \bigcup_{m=n}^{\infty} A_m\right] = 1 - \mathbb{P}\left[\bigcup_n \bigcap_{m=n}^{\infty} A_m^C\right] = 1.$$

So we are done. $\square$

2 Measurable functions and random variables

2.1 Measurable functions

Lemma. Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces, and $\mathcal{G} = \sigma(\mathcal{Q})$ for some $\mathcal{Q}$. If $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{Q}$, then f is measurable.

Proof. We claim that
$$\{A \subseteq G : f^{-1}(A) \in \mathcal{E}\}$$
is a σ-algebra on G. Then the result follows immediately by definition of $\sigma(\mathcal{Q})$.

Indeed, this follows from the fact that f^{-1} preserves everything. More precisely, we have

$$f^{-1}\left(\bigcup_n A_n\right) = \bigcup_n f^{-1}(A_n), \quad f^{-1}(A^C) = (f^{-1}(A))^C, \quad f^{-1}(\emptyset) = \emptyset.$$

So if, say, all $A_n \in \mathcal{A}$, then so is $\bigcup_n A_n$. □

Proposition. Let $f_i : E \to F_i$ be functions. Then $\{f_i\}$ are all measurable iff $(f_i) : E \to \prod F_i$ is measurable, where the function (f_i) is defined by setting the ith component of $(f_i)(x)$ to be $f_i(x)$.

Proof. If the map (f_i) is measurable, then by composition with the projections π_i, we know that each f_i is measurable.

Conversely, if all f_i are measurable, then since the σ-algebra of $\prod F_i$ is generated by sets of the form $\pi_j^{-1}(A) : A \in \mathcal{F}_j$, and the pullback of such sets along (f_i) is exactly $f_j^{-1}(A)$, we know the function (f_i) is measurable. □

Proposition. Let $(E, \mathcal{E})$ be a measurable space. Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions on E. Then the following are measurable:

$$\begin{gathered} f_1 + f_2, \quad f_1 f_2, \quad \max\{f_1, f_2\}, \quad \min\{f_1, f_2\}, \\ \inf_n f_n, \quad \sup_n f_n, \quad \liminf_n f_n, \quad \limsup_n f_n. \end{gathered}$$

The same is true with "real" replaced with "non-negative", provided the new functions are real (i.e. not infinity).

Proof. This is an (easy) exercise on the example sheet. For example, the sum $f_1 + f_2$ can be written as the following composition.

$$E \xrightarrow{(f_1, f_2)} [0, \infty]^2 \xrightarrow{+} [0, \infty].$$

We know the second map is continuous, hence measurable. The first function is also measurable since the f_i are. So the composition is also measurable.

The product follows similarly, but for the infimum and supremum, we need to check explicitly that the corresponding maps $[0, \infty]^{\mathbb{N}} \to [0, \infty]$ is measurable. □

Theorem (Monotone class theorem)**.** Let $(E, \mathcal{E})$ be a measurable space, and $\mathcal{A} \subseteq \mathcal{E}$ be a π-system with $\sigma(\mathcal{A}) = \mathcal{E}$. Let $\mathcal{V}$ be a vector space of functions such that

(i) The constant function $1 = \mathbf{1}_E$ is in $\mathcal{V}$.

(ii) The indicator functions $\mathbf{1}_A \in \mathcal{V}$ for all $A \in \mathcal{A}$

(iii) $\mathcal{V}$ is closed under bounded, monotone limits.

More explicitly, if (f_n) is a bounded non-negative sequence in $\mathcal{V}$, $f_n \nearrow f$ (pointwise) and f is also bounded, then $f \in \mathcal{V}$.

Then $\mathcal{V}$ contains all bounded measurable functions.

Proof. We first deduce that $\mathbf{1}_A \in \mathcal{V}$ for all $A \in \mathcal{E}$.

$$\mathcal{D} = \{A \in \mathcal{E} : \mathbf{1}_A \in \mathcal{V}\}.$$

We want to show that $\mathcal{D} = \mathcal{E}$. To do this, we have to show that $\mathcal{D}$ is a d-system.

(i) Since $\mathbf{1}_E \in \mathcal{V}$, we know $E \in \mathcal{D}$.

(ii) If $\mathbf{1}_A \in V$, then $1 - \mathbf{1}_A = \mathbf{1}_{E \setminus A} \in \mathcal{V}$. So $E \setminus A \in \mathcal{D}$.

(iii) If (A_n) is an increasing sequence in $\mathcal{D}$, then $\mathbf{1}_{A_n} \to \mathbf{1}_{\bigcup A_n}$ monotonically increasingly. So $\mathbf{1}_{\bigcup A_n}$ is in $\mathcal{D}$.

So, by Dynkin's lemma, we know $\mathcal{D} = \mathcal{E}$. So $\mathcal{V}$ contains indicators of all measurable sets. We will now try to obtain any measurable function by approximating.

Suppose that f is bounded and non-negative measurable. We want to show that $f \in \mathcal{V}$. To do this, we approximate it by letting

$$f_n = 2^{-n} \lfloor 2^n f \rfloor = \sum_{k=0}^{\infty} k 2^{-n} \mathbf{1}_{\{k 2^{-n} \leq f < (k+1) 2^{-n}\}}.$$

Note that since f is bounded, this is a finite sum. So it is a finite linear combination of indicators of elements in $\mathcal{E}$. So $f_n \in \mathcal{V}$, and $0 \leq f_n \to f$ monotonically. So $f \in \mathcal{V}$.

More generally, if f is bounded and measurable, then we can write

$$f = (f \vee 0) + (f \wedge 0) \equiv f^+ - f^-.$$

Then f^+ and f^- are bounded and non-negative measurable. So $f \in \mathcal{V}$. $\square$

2.2 Constructing new measures

Lemma. Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, non-decreasing and right continuous. We set

$$g(\pm\infty) = \lim_{x \to \pm\infty} g(x).$$

We set $I = (g(-\infty), g(\infty))$. Since g is non-constant, this is non-empty.

Then there is a non-decreasing, left continuous function $f : I \to \mathbb{R}$ such that for all $x \in I$ and $y \in \mathbb{R}$, we have

$$x \leq g(y) \Leftrightarrow f(x) \leq y.$$

Thus, taking the negation of this, we have

$$x > g(y) \Leftrightarrow f(x) > y.$$

Explicitly, for $x \in I$, we define

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

Proof. We just have to verify that it works. For $x \in I$, consider

$$J_x = \{y \in \mathbb{R} : x \leq g(y)\}.$$

Since g is non-decreasing, if $y \in J_x$ and $y' \geq y$, then $y' \in J_x$. Since g is right-continuous, if $y_n \in J_x$ is such that $y_n \searrow y$, then $y \in J_x$. So we have

$$J_x = [f(x), \infty).$$

Thus, for $f \in \mathbb{R}$, we have

$$x \leq g(y) \Leftrightarrow f(x) \leq y.$$

So we just have to prove the remaining properties of f. Now for $x \leq x'$, we have $J_x \subseteq J_{x'}$. So $f(x) \leq f(x')$. So f is non-decreasing.

Similarly, if $x_n \nearrow x$, then we have $J_x = \bigcap_n J_{x_n}$. So $f(x_n) \to f(x)$. So this is left continuous. □

Theorem. Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, non-decreasing and right continuous. Then there exists a unique Radon measure $\mathrm{d}g$ on $\mathcal{B}$ such that

$$\mathrm{d}g((a, b]) = g(b) - g(a).$$

Moreover, we obtain all non-zero Radon measures on $\mathbb{R}$ in this way.

Proof. Take I and f as in the previous lemma, and let μ be the restriction of the Lebesgue measure to Borel subsets of I. Now f is measurable since it is left continuous. We define $\mathrm{d}g = \mu \circ f^{-1}$. Then we have

$$\begin{aligned}
\mathrm{d}g((a, b]) &= \mu(\{x \in I : a < f(x) \leq b\}) \\
&= \mu(\{x \in I : g(a) < x \leq g(b)\}) \\
&= \mu((g(a), g(b)]) = g(b) - g(a).
\end{aligned}$$

So $\mathrm{d}g$ is a Radon measure with the required property.

There are no other such measures by the argument used for uniqueness of the Lebesgue measure.

To show we get all non-zero Radon measures this way, suppose we have a Radon measure ν on $\mathbb{R}$, we want to produce a g such that $\nu = \mathrm{d}g$. We set

$$g(y) = \begin{cases} -\nu((y, 0]) & y \leq 0 \\ \nu((0, y]) & y > 0 \end{cases}.$$

Then $\nu((a, b]) = g(b) - g(a)$. We see that ν is non-zero, so g is non-constant. It is also easy to see it is non-decreasing and right continuous. So $\nu = \mathrm{d}g$ by continuity. □

2.3 Random variables

Proposition. We have

$$F_X(x) \to \begin{cases} 0 & x \to -\infty \\ 1 & x \to +\infty \end{cases}.$$

Also, $F_X(x)$ is non-decreasing and right-continuous.

Proposition. Let F be any distribution function. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X such that $F_X = F$.

Proof. Take $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0,1), \text{Lebesgue})$. We take $X : \Omega \to \mathbb{R}$ to be

$$X(\omega) = \inf\{x : \omega \leq f(x)\}.$$

Then we have

$$X(\omega) \leq x \iff w \leq F(x).$$

So we have

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[(0, F(x)]] = F(x).$$

Therefore $F_X = F$. □

Proposition. Two real-valued random variables X, Y are independent iff

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x]\mathbb{P}[Y \leq y].$$

More generally, if (X_n) is a sequence of real-valued random variables, then they are independent iff

$$\mathbb{P}[x_1 \leq x_1, \cdots, x_n \leq x_n] = \prod_{j=1}^n \mathbb{P}[X_j \leq x_j]$$

for all n and x_j.

Proof. The $\Rightarrow$ direction is obvious. For the other direction, we simply note that $\{(-\infty, x] : x \in \mathbb{R}\}$ is a generating π-system for the Borel σ-algebra of $\mathbb{R}$. □

Proposition. Let

$$(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0,1), \text{Lebesgue}).$$

be our probability space. Then there exists as sequence R_n of independent Bernoulli$(1/2)$ random variables.

Proof. Suppose we have $\omega \in \Omega = (0, 1)$. Then we write ω as a binary expansion

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

where $\omega_n \in \{0, 1\}$. We make the binary expansion unique by disallowing infinite sequences of zeroes.

We define $R_n(\omega) = \omega_n$. We will show that R_n is measurable. Indeed, we can write

$$R_1(\omega) = \omega_1 = \mathbf{1}_{(1/2,1]}(\omega),$$

where $\mathbf{1}_{(1/2,1]}$ is the indicator function. Since indicator functions of measurable sets are measurable, we know R_1 is measurable. Similarly, we have

$$R_2(\omega) = \mathbf{1}_{(1/4,1/2]}(\omega) + \mathbf{1}_{(3/4,1]}(\omega).$$

So this is also a measurable function. More generally, we can do this for any $R_n(\omega)$: we have

$$R_n(\omega) = \sum_{j=1}^{2^{n-1}} \mathbf{1}_{(2^{-n}(2j-1), 2^{-n}(2j)]}(\omega).$$

So each R_n is a random variable, as each can be expressed as a sum of indicators of measurable sets.

Now let's calculate

$$\mathbb{P}[R_n = 1] = \sum_{j=1}^{2n-1} 2^{-n}((2j) - (2j-1)) = \sum_{j=1}^{2n-1} 2^{-n} = \frac{1}{2}.$$

Then we have

$$\mathbb{P}[R_n = 0] = 1 - \mathbb{P}[R_n = 1] = \frac{1}{2}$$

as well. So $R_n \sim \text{Bernoulli}(1/2)$.

We can straightforwardly check that (R_n) is an independent sequence, since for $n \neq m$, we have

$$\mathbb{P}[R_n = 0 \text{ and } R_m = 0] = \frac{1}{4} = \mathbb{P}[R_n = 0]\mathbb{P}[R_m = 0]. \qquad \square$$

Proposition. Let

$$(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0,1), \text{Lebesgue}).$$

Given any sequence (F_n) of distribution functions, there is a sequence (X_n) of independent random variables with $F_{X_n} = F_n$ for all n.

Proof. Let $m : \mathbb{N}^2 \to \mathbb{N}$ be any bijection, and relabel

$$Y_{k,n} = R_{m(k,n)},$$

where the R_j are as in the previous random variable. We let

$$Y_n = \sum_{k=1}^{\infty} 2^{-k} Y_{k,n}.$$

Then we know that (Y_n) is an independent sequence of random variables, and each is uniform on $(0,1)$. As before, we define

$$G_n(y) = \inf\{x : y \leq F_n(x)\}.$$

We set $X_n = G_n(Y_n)$. Then (X_n) is a sequence of random variables with $F_{X_n} = F_n$. $\square$

2.4 Convergence of measurable functions

Theorem.

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

(ii) For any E, if $f_n \to f$ in measure, then there exists a subsequence (f_{n_k}) such that $f_{n_k} \to f$ a.e.

Proof.

(i) First suppose $\mu(E) < \infty$, and fix $\varepsilon > 0$. Consider

$$\mu(\{x \in E : |f_n(x) - f(x)| \leq \varepsilon\}).$$

We use the result from the first example sheet that for any sequence of events (A_n), we have

$$\liminf \mu(A_n) \geq \mu(\liminf A_n).$$

Applying to the above sequence says

$$\begin{aligned}\liminf \mu(\{x : |f_n(x) - f(x)| \leq \varepsilon\}) &\geq \mu(\{x : |f_m(x) - f(x)| \leq \varepsilon \text{ eventually}\})\\ &\geq \mu(\{x \in E : |f_m(x) - f(x)| \to 0\})\\ &= \mu(E).\end{aligned}$$

As $\mu(E) < \infty$, we have $\mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) \to 0$ as $n \to \infty$.

(ii) Suppose that $f_n \to f$ in measure. We pick a subsequence (n_k) such that

$$\mu\left(\left\{x \in E : |f_{n_k}(x) - f(x)| > \frac{1}{k}\right\}\right) \leq 2^{-k}.$$

Then we have

$$\sum_{k=1}^{\infty} \mu\left(\left\{x \in E : f_{n_k}(x) - f(x)| > \frac{1}{k}\right\}\right) \leq \sum_{k=1}^{\infty} 2^{-k} = 1 < \infty.$$

By the first Borel–Cantelli lemma, we know

$$\mu\left(\left\{x \in E : |f_{n_k}(x) - f(x)| > \frac{1}{k} \text{ i.o.}\right\}\right) = 0.$$

So $f_{n_k} \to f$ a.e. $\square$

Theorem (Skorokhod representation theorem of weak convergence)**.**

(i) If $(X_n), X$ are defined on the same probability space, and $X_n \to X$ in probability. Then $X_n \to X$ in distribution.

(ii) If $X_n \to X$ in distribution, then there exists random variables $(\tilde{X}_n)$ and $\tilde{X}$ defined on a common probability space with $F_{\tilde{X}_n} = F_{X_n}$ and $F_{\tilde{X}} = F_X$ such that $\tilde{X}_n \to \tilde{X}$ a.s.

Proof. Let $S = \{x \in \mathbb{R} : F_X \text{ is continuous}\}$.

(i) Assume that $X_n \to X$ in probability. Fix $x \in S$. We need to show that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$.

We fix $\varepsilon > 0$. Since $x \in S$, this implies that there is some $\delta > 0$ such that

$$\begin{aligned} F_X(x-\delta) &\geq F_X(x) - \frac{\varepsilon}{2} \\ F_X(x+\delta) &\leq F_X(x) + \frac{\varepsilon}{2}. \end{aligned}$$

We fix N large such that $n \geq N$ implies $\mathbb{P}[|X_n - X| \geq \delta] \leq \frac{\varepsilon}{2}$. Then

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}[X_n \leq x] \\ &= \mathbb{P}[(X_n - X) + X \leq x] \end{aligned}$$

We now notice that $\{(X_n - X) + X \leq x\} \subseteq \{X \leq x + \delta\} \cup \{|X_n - X| > \delta\}$. So we have

$$\begin{aligned} &\leq \mathbb{P}[X \leq x + \delta] + \mathbb{P}[|X_n - X| > \delta] \\ &\leq F_X(x+\delta) + \frac{\varepsilon}{2} \\ &\leq F_X(x) + \varepsilon. \end{aligned}$$

We similarly have

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}[X_n \leq x] \\ &\geq \mathbb{P}[X \leq x - \delta] - \mathbb{P}[|X_n - X| > \delta] \\ &\geq F_X(x-\delta) - \frac{\varepsilon}{2} \\ &\geq F_X(x) - \varepsilon. \end{aligned}$$

Combining, we have that $n \geq N$ implying $|F_{x_n}(x) - F_X(x)| \leq \varepsilon$. Since ε was arbitrary, we are done.

(ii) Suppose $X_n \to X$ in distribution. We again let

$$(\Omega, \mathcal{F}, \mathcal{B}) = ((0,1), \mathcal{B}((0,1)), \text{Lebesgue}).$$

We let

$$\begin{aligned} \tilde{X}_n(\omega) &= \inf\{x : \omega \leq F_{X_n}(x)\}, \\ \tilde{X}(\omega) &= \inf\{x : \omega \leq F_X(x)\}. \end{aligned}$$

Recall from before that $\tilde{X}_n$ has the same distribution function as X_n for all n, and $\tilde{X}$ has the same distribution as X. Moreover, we have

$$\begin{aligned} \tilde{X}_n(\omega) \leq x &\Leftrightarrow \omega \leq F_{X_n}(x) \\ x < \tilde{X}_n(\omega) &\Leftrightarrow F_{X_n}(x) < \omega, \end{aligned}$$

and similarly if we replace X_n with X.

We are now going to show that with this particular choice, we have $\tilde{X}_n \to \tilde{X}$ a.s.

Note that $\tilde{X}$ is a non-decreasing function $(0,1) \to \mathbb{R}$. Then by general analysis, $\tilde{X}$ has at most countably many discontinuities. We write

$$\Omega_0 = \{\omega \in (0,1) : \tilde{X} \text{ is continuous at } \omega_0\}.$$

Then $(0,1) \setminus \Omega_0$ is countable, and hence has Lebesgue measure 0. So

$$\mathbb{P}[\Omega_0] = 1.$$

We are now going to show that $\tilde{X}_n(\omega) \to \tilde{X}(\omega)$ for all $\omega \in \Omega_0$.

Note that F_X is a non-decreasing function, and hence the points of discontinuity $\mathbb{R} \setminus S$ is also countable. So S is dense in $\mathbb{R}$. Fix $\omega \in \Omega_0$ and $\varepsilon > 0$. We want to show that $|\tilde{X}_n(\omega) - \tilde{X}(\omega)| \leq \varepsilon$ for all n large enough.

Since S is dense in $\mathbb{R}$, we can find x^-, x^+ in S such that

$$x^- < \tilde{X}(\omega) < x^+$$

and $x^+ - x^- < \varepsilon$. What we *want* to do is to use the characteristic property of $\tilde{X}$ and F_X to say that this implies

$$F_X(x^-) < \omega < F_X(x^+).$$

Then since $F_{X_n} \to F_X$ at the points x^-, x^+, for sufficiently large n, we have

$$F_{X_n}(x^-) < \omega < F_{X_n}(x^+).$$

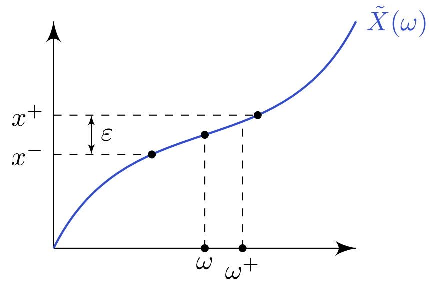
Hence we have

$$x^- < \tilde{X}_n(\omega) < x^+.$$

Then it follows that $|\tilde{X}_n(\omega) - \tilde{X}(\omega)| < \varepsilon$.

However, this doesn't work, since $\tilde{X}(\omega) < x^+$ only implies $\omega \leq F_X(x^+)$, and our argument will break down. So we do a funny thing where we introduce a new variable ω^+.

Since $\tilde{X}$ is continuous at ω, we can find $\omega^+ \in (\omega, 1)$ such that $\tilde{X}(\omega^+) \leq x^+$.



Then we have

$$x^- < \tilde{X}(\omega) \leq \tilde{X}(\omega^+) < x^+.$$

Then we have

$$F_X(x^-) < \omega < \omega^+ \leq F_X(x^+).$$

So for sufficiently large n, we have

$$F_{X_n}(x^-) < \omega < F_{X_n}(x^+).$$

So we have

$$x^- < \tilde{X}_n(\omega) \leq x^+,$$

and we are done. □

2.5 Tail events

Theorem (Kolmogorov 0-1 law)**.** Let (X_n) be a sequence of independent (real-valued) random variables. If $A \in \mathcal{T}$, then $\mathbb{P}[A] = 0$ or 1.

Moreover, if X is a $\mathcal{T}$-measurable random variable, then there exists a constant c such that

$$\mathbb{P}[X = c] = 1.$$

Proof. The proof is very funny the first time we see it. We are going to prove the theorem by checking something that seems very strange. We are going to show that if $A \in \mathcal{T}$, then A is independent of A. It then follows that

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

so $\mathbb{P}[A] = 0$ or 1. In fact, we are going to prove that $\mathcal{T}$ is independent of $\mathcal{T}$.

Let

$$\mathcal{F}_n = \sigma(X_1, \cdots, X_n).$$

This σ-algebra is generated by the π-system of events of the form

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

Similarly, $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \cdots)$ is generated by the π-system of events of the form

$$B = \{X_{n+1} \leq x_{n+1}, \cdots, X_{n+k} \leq x_{n+k}\},$$

where k is any natural number.

Since the X_n are independent, we know for any such A and B, we have

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

Since this is true for all A and B, it follows that $\mathcal{F}_n$ is independent of $\mathcal{T}_n$.

Since $\mathcal{T} = \bigcap_k \mathcal{T}_k \subseteq \mathcal{T}_n$ for each n, we know $\mathcal{F}_n$ is independent of $\mathcal{T}$.

Now $\bigcup_k \mathcal{F}_k$ is a π-system, which generates the σ-algebra $\mathcal{F}_\infty = \sigma(X_1, X_2, \cdots)$. We know that if $A \in \bigcup_n \mathcal{F}_n$, then there has to exist an index n such that $A \in \mathcal{F}_n$. So A is independent of $\mathcal{T}$. So $\mathcal{F}_\infty$ is independent of $\mathcal{T}$.

Finally, note that $\mathcal{T} \subseteq \mathcal{F}_\infty$. So $\mathcal{T}$ is independent of $\mathcal{T}$.

To find the constant, suppose that X is $\mathcal{T}$-measurable. Then

$$\mathbb{P}[X \leq x] \in \{0, 1\}$$

for all $x \in \mathbb{R}$ since $\{X \leq x\} \in \mathcal{T}$.

Now take

$$c = \inf\{x \in \mathbb{R} : \mathbb{P}[X \leq x] = 1\}.$$

Then with this particular choice of c, it is easy to see that $\mathbb{P}[X = c] = 1$. This completes the proof of the theorem. □

3 Integration

3.1 Definition and basic properties

Proposition. A function is simple iff it is measurable, non-negative, and takes on only finitely-many values.

Proposition. Let $f : [0,1] \to \mathbb{R}$ be Riemann integrable. Then it is also Lebesgue integrable, and the two integrals agree.

Theorem (Monotone convergence theorem)**.** Suppose that $(f_n), f$ are non-negative measurable with $f_n \nearrow f$. Then $\mu(f_n) \nearrow \mu(f)$.

Proof. We will split the proof into five steps. We will prove each of the following in turn:

(i) If f_n and f are indicator functions, then the theorem holds.

(ii) If f is an indicator function, then the theorem holds.

(iii) If f is simple, then the theorem holds.

(iv) If f is non-negative measurable, then the theorem holds.

Each part follows rather straightforwardly from the previous one, and the reader is encouraged to try to prove it themself.

We first consider the case where $f_n = \mathbf{1}_{A_n}$ and $f = \mathbf{1}_A$. Then $f_n \nearrow f$ is true iff $A_n \nearrow A$. On the other hand, $\mu(f_n) \nearrow \mu(f)$ iff $\mu(A_n) \nearrow \mu(A)$.

For convenience, we let $A_0 = \emptyset$. We can write

$$\begin{aligned}
\mu(A) &= \mu\left(\bigcup_n A_n \setminus A_{n-1}\right)\\
&= \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n-1})\\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \mu(A_n \setminus A_{n-1})\\
&= \lim_{N\to\infty} \mu(A_N).
\end{aligned}$$

So done.

We next consider the case where $f = \mathbf{1}_A$ for some A. Fix $\varepsilon > 0$, and set

$$A_n = \{f_n > 1 - \varepsilon\} \in \mathcal{E}.$$

Then we know that $A_n \nearrow A$, as $f_n \nearrow f$. Moreover, by definition, we have

$$(1-\varepsilon)\mathbf{1}_{A_n} \leq f_n \leq f = \mathbf{1}_A.$$

As $A_n \nearrow A$, we have that

$$(1-\varepsilon)\mu(f) = (1-\varepsilon)\lim_{n\to\infty} \mu(A_n) \leq \lim_{n\to\infty} \mu(f_n) \leq \mu(f)$$

since $f_n \leq f$. Since ε is arbitrary, we know that

$$\lim_{n\to\infty} \mu(f_n) = \mu(f).$$

Next, we consider the case where f is simple. We write

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

where $a_k > 0$ and A_k are pairwise disjoint. Since $f_n \nearrow f$, we know

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

So we have

$$\mu(f_n) = \sum_{k=1}^{m} \mu(f_n \mathbf{1}_{A_k}) = \sum_{k=1}^{m} a_k \mu(a_k^{-1} f_n \mathbf{1}_{A_k}) \to \sum_{k=1}^{m} a_k \mu(A_k) = \mu(f).$$

Suppose f is non-negative measurable. Suppose $g \leq f$ is a simple function. As $f_n \nearrow f$, we know $f_n \wedge g \nearrow f \wedge g = g$. So by the previous case, we know that

$$\mu(f_n \wedge g) \to \mu(g).$$

We also know that

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

So we have

$$\lim_{n\to\infty} \mu(f_n) \geq \mu(g)$$

for all $g \leq f$. This is possible only if

$$\lim_{n\to\infty} \mu(f_n) \geq \mu(f)$$

by definition of the integral. However, we also know that $\mu(f_n) \leq \mu(f)$ for all n, again by definition of the integral. So we must have equality. So we have

$$\mu(f) = \lim_{n\to\infty} \mu(f_n). \qquad \square$$

Theorem. Let f, g be non-negative measurable, and $\alpha, \beta \geq 0$. We have that

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

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

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

Proof.

(i) Let

$$f_n = 2^{-n}\lfloor 2^n f\rfloor \wedge n$$
$$g_n = 2^{-n}\lfloor 2^n g\rfloor \wedge n.$$

Then f_n, g_n are simple with $f_n \nearrow f$ and $g_n \nearrow g$. Hence $\mu(f_n) \nearrow \mu(f)$ and $\mu(g_n) \nearrow \mu(g)$ and $\mu(\alpha f_n + \beta g_n) \nearrow \mu(\alpha f + \beta g)$, by the monotone convergence theorem. As f_n, g_n are simple, we have that

$$\mu(\alpha f_n + \beta g_n) = \alpha\mu(f_n) + \beta\mu(g_n).$$

Taking the limit as $n \to \infty$, we get

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

(ii) We shall be careful not to use the monotone convergence theorem. We have

$$\begin{aligned}\mu(g) &= \sup\{\mu(h) : h \leq g \text{ simple}\}\\ &\geq \sup\{\mu(h) : h \leq f \text{ simple}\}\\ &= \mu(f).\end{aligned}$$

(iii) Suppose $f \neq 0$ a.e. Let

$$A_n = \left\{x : f(x) > \frac{1}{n}\right\}.$$

Then

$$\{x : f(x) \neq 0\} = \bigcup_n A_n.$$

Since the left hand set has non-negative measure, it follows that there is some A_n with non-negative measure. For that n, we define

$$h = \frac{1}{n}\mathbf{1}_{A_n}.$$

Then $\mu(f) \geq \mu(h) > 0$. So $\mu(f) \neq 0$.

Conversely, suppose $f = 0$ a.e. We let

$$f_n = 2^{-n}\lfloor 2^n f\rfloor \wedge n$$

be a simple function. Then $f_n \nearrow f$ and $f_n = 0$ a.e. So

$$\mu(f) = \lim_{n\to\infty} \mu(f_n) = 0. \qquad \square$$

Theorem. Let f, g be integrable, and $\alpha, \beta \geq 0$. We have that

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

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

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

Proof.

(i) We are going to prove these by applying the previous theorem.

By definition of the integral, we have $\mu(-f) = -\mu(f)$. Also, if $\alpha \geq 0$, then

$$\mu(\alpha f) = \mu(\alpha f^+) - \mu(\alpha f^-) = \alpha\mu(f^+) - \alpha\mu(f^-) = \alpha\mu(f).$$

Combining these two properties, it then follows that if α is a real number, then

$$\mu(\alpha f) = \alpha\mu(f).$$

To finish the proof of (i), we have to show that $\mu(f + g) = \mu(f) + \mu(g)$. We know that this is true for non-negative functions, so we need to employ a little trick to make this a statement about the non-negative version. If we let $h = f + g$, then we can write this as

$$h^+ - h^- = (f^+ - f^-) + (g^+ - g^-).$$

We now rearrange this as

$$h^+ f^- + g^- = f^+ + g^+ + h^-.$$

Now everything is non-negative measurable. So applying μ gives

$$\mu(f^+) + \mu(f^-) + \mu(g^-) = \mu(f^+) + \mu(g^+) + \mu(h^-).$$

Rearranging, we obtain

$$\mu(h^+) - \mu(h^-) = \mu(f^+) - \mu(f^-) + \mu(g^+) - \mu(g^-).$$

This is exactly the same thing as saying

$$\mu(f + g) = \mu(h) = \mu(f) = \mu(g).$$

(ii) If $f \leq g$, then $g - f \geq 0$. So $\mu(g - f) \geq 0$. By (i), we know $\mu(g) - \mu(f) \geq 0$. So $\mu(g) \geq \mu(f)$.

(iii) If $f = 0$ a.e., then $f^+, f^- = 0$ a.e. So $\mu(f^+) = \mu(f^-) = 0$. So $\mu(f) = \mu(f^+) - \mu(f^-) = 0$. □

Proposition. If $\mathcal{A}$ is a π-system with $E \in \mathcal{A}$ and $\sigma(\mathcal{A}) = \mathcal{E}$, and f is an integrable function that

$$\mu(f\mathbf{1}_A) = 0$$

for all $A \in \mathcal{A}$. Then $\mu(f) = 0$ a.e.

Proof. Let

$$\mathcal{D} = \{A \in \mathcal{E} : \mu(f\mathbf{1}_A) = 0\}.$$

It follows immediately from the properties of the integral that $\mathcal{D}$ is a d-system. So $\mathcal{D} = \mathcal{E}$ by Dynkin's lemma. Let

$$A^+ = \{x \in E : f(x) > 0\},$$
$$A^- = \{x \in E : f(x) < 0\}.$$

Then $A^\pm \in \mathcal{E}$, and

$$\mu(f\mathbf{1}_{A^+}) = \mu(f\mathbf{1}_{A^-}) = 0.$$

So $f\mathbf{1}_{A^+}$ and $f\mathbf{1}_{A^-}$ vanish a.e. So f vanishes a.e. □

Proposition. Suppose that (g_n) is a sequence of non-negative measurable functions. Then we have

$$\mu\left(\sum_{n=1}^\infty g_n\right) = \sum_{n=1}^\infty \mu(g_n).$$

Proof. We know

$$\left(\sum_{n=1}^N g_n\right) \nearrow \left(\sum_{n=1}^\infty g_n\right)$$

as $N \to \infty$. So by the monotone convergence theorem, we have

$$\sum_{n=1}^N \mu(g_n) = \mu\left(\sum_{n=1}^N g_n\right) \nearrow \mu\left(\sum_{n=1}^\infty g_n\right).$$

But we also know that

$$\sum_{n=1}^N \mu(g_n) \nearrow \sum_{n=1}^\infty \mu(g_n)$$

by definition. So we are done. □

3.2 Integrals and limits

Theorem (Fatou's lemma)**.** Let (f_n) be a sequence of non-negative measurable functions. Then

$$\mu(\liminf f_n) \leq \liminf \mu(f_n).$$

Proof. We start with the trivial observation that if $k \geq n$, then we always have that

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

By the monotonicity of the integral, we know that

$$\mu\left(\inf_{m \geq n} f_m\right) \leq \mu(f_k).$$

for all $k \geq n$.

So we have

$$\mu\left(\inf_{m \geq n} f_m\right) \leq \inf_{k \geq n} \mu(f_k) \leq \liminf_m \mu(f_m).$$

It remains to show that the left hand side converges to $\mu(\liminf f_m)$. Indeed, we know that

$$\inf_{m \geq n} f_m \nearrow \liminf_m f_m.$$

Then by monotone convergence, we have

$$\mu\left(\inf_{m \geq n} f_m\right) \nearrow \mu\left(\liminf_m f_m\right).$$

So we have

$$\mu\left(\liminf_m f_m\right) \leq \liminf_m \mu(f_m).$$ □

Theorem (Dominated convergence theorem)**.** Let $(f_n), f$ be measurable with $f_n(x) \to f(x)$ for all $x \in E$. Suppose that there is an integrable function g such that

$$|f_n| \leq g$$

for all n, then we have

$$\mu(f_n) \to \mu(f)$$

as $n \to \infty$.

Proof. Note that

$$|f| = \lim_n |f|_n \leq g.$$

So we know that

$$\mu(|f|) \leq \mu(g) < \infty.$$

So we know that f, f_n are integrable.

Now note also that

$$0 \leq g + f_n, \quad 0 \leq g - f_n$$

for all n. We are now going to apply Fatou's lemma twice with these series. We have that

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

Since $\mu(g)$ is finite, we know that

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

We now do the same thing with $g - f_n$. We have

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

Again, since $\mu(g)$ is finite, we know that

$$\mu(f) \geq \limsup_n \mu(f_n).$$

These combine to tell us that

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

So they must be all equal, and thus $\mu(f_n) \to \mu(f)$. □

3.3 New measures from old

Lemma. For $(E, \mathcal{E}, \mu)$ a measure space and $A \in \mathcal{E}$, the restriction to A is a measure space. $\square$

Proposition. Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \mu')$ be measure spaces and $A \in \mathcal{E}$. Let $f : E \to F$ be a measurable function. Then $f|_A$ is $\mathcal{E}_A$-measurable.

Proof. Let $B \in \mathcal{F}$. Then

$$(f|_A)^{-1}(B) = f^{-1}(B) \cap A \in \mathcal{E}_A. \qquad \square$$

Proposition. If f is integrable, then $f|_A$ is μ_A-integrable and $\mu_A(f|_A) = \mu(f\mathbf{1}_A)$. $\square$

Proposition. If g is a non-negative measurable function on G, then

$$\nu(g) = \mu(g \circ f).$$

Proof. Exercise using the monotone class theorem (see example sheet). $\square$

Proposition. The ν defined above is indeed a measure.

Proof.

(i) $\nu(\phi) = \mu(f\mathbf{1}_\emptyset) = \mu(0) = 0$.

(ii) If (A_n) is a disjoint sequence in $\mathcal{E}$, then

$$\nu\left(\bigcup A_n\right) = \mu(f\mathbf{1}_{\bigcup A_n}) = \mu\left(f \sum \mathbf{1}_{A_n}\right) = \sum \mu\left(f\mathbf{1}_{A_n}\right) = \sum \nu(f). \qquad \square$$

3.4 Integration and differentiation

Proposition (Change of variables formula)**.** Let $\phi : [a, b] \to \mathbb{R}$ be continuously differentiable and increasing. Then for any bounded Borel function g, we have

$$\int_{\phi(a)}^{\phi(b)} g(y)\ \mathrm{d}y = \int_a^b g(\phi(x))\phi'(x)\ \mathrm{d}x. \qquad (*)$$

Proof. We let

$$V = \{\text{Borel functions } g \text{ such that } (*) \text{ holds}\}.$$

We will want to use the monotone class theorem to show that this includes all bounded functions.

We already know that

(i) V contains $\mathbf{1}_A$ for all A in the π-system of intervals of the form $[u, v] \subseteq [a, b]$. This is just the fundamental theorem of calculus.

(ii) By linearity of the integral, V is indeed a vector space.

(iii) Finally, let (g_n) be a sequence in V, and $g_n \geq 0$, $g_n \nearrow g$. Then we know that
$$\int_{\phi(a)}^{\phi(b)} g_n(y)\;\mathrm{d}y = \int_a^b g_n(\phi(x))\phi'(x)\;\mathrm{d}x.$$
By the monotone convergence theorem, these converge to
$$\int_{\phi(a)}^{\phi(b)} g(y)\;\mathrm{d}y = \int_a^b g(\phi(x))\phi'(x)\;\mathrm{d}x.$$

Then by the monotone class theorem, V contains all bounded Borel functions. $\square$

Theorem (Differentiation under the integral sign)**.** Let $(E, \mathcal{E}, \mu)$ be a space, and $U \subseteq \mathbb{R}$ be an open set, and $f : U \times E \to \mathbb{R}$. We assume that

(i) For any $t \in U$ fixed, the map $x \mapsto f(t, x)$ is integrable;

(ii) For any $x \in E$ fixed, the map $t \mapsto f(t, x)$ is differentiable;

(iii) There exists an integrable function g such that
$$\left|\frac{\partial f}{\partial t}(t, x)\right| \leq g(x)$$
for all $x \in E$ and $t \in U$.

Then the map
$$x \mapsto \frac{\partial f}{\partial t}(t, x)$$
is integrable for all t, and also the function
$$F(t) = \int_E f(t, x)\mathrm{d}\mu$$
is differentiable, and
$$F'(t) = \int_E \frac{\partial f}{\partial t}(t, x)\;\mathrm{d}\mu.$$

Proof. Measurability of the derivative follows from the fact that it is a limit of measurable functions, and then integrability follows since it is bounded by g.

Suppose (h_n) is a positive sequence with $h_n \to 0$. Then let
$$g_n(x) = \frac{f(t + h_n, x) - f(t, x)}{h_n} - \frac{\partial f}{\partial t}(t, x).$$
Since f is differentiable, we know that $g_n(x) \to 0$ as $n \to \infty$. Moreover, by the mean value theorem, we know that
$$|g_n(x)| \leq 2g(x).$$
On the other hand, by definition of $F(t)$, we have
$$\frac{F(t + h_n) - F(t)}{h_n} - \int_E \frac{\partial f}{\partial t}(t, x)\;\mathrm{d}\mu = \int g_n(x)\;\mathrm{d}x.$$
By dominated convergence, we know the RHS tends to 0. So we know
$$\lim_{n\to\infty} \frac{F(t + h_n) - F(t)}{h_n} \to \int_E \frac{\partial f}{\partial t}(t, x)\;\mathrm{d}\mu.$$
Since h_n was arbitrary, it follows that $F'(t)$ exists and is equal to the integral. $\square$

3.5 Product measures and Fubini's theorem

Lemma. Let $E = E_1 \times E_2$ be a product of σ-algebras. Suppose $f : E \to \mathbb{R}$ is $\mathcal{E}$-measurable function. Then

(i) For each $x_2 \in E_2$, the function $x_1 \mapsto f(x_1, x_2)$ is $\mathcal{E}_1$-measurable.

(ii) If f is bounded or non-negative measurable, then

$$f_2(x_2) = \int_{E_1} f(x_1, x_2)\, \mu_1(\mathrm{d}x_1)$$

is $\mathcal{E}_2$-measurable.

Proof. The first part follows immediately from the fact that for a fixed x_2, the map $\iota_1 : E_1 \to E$ given by $\iota_1(x_1) = (x_1, x_2)$ is measurable, and that the composition of measurable functions is measurable.

For the second part, we use the monotone class theorem. We let V be the set of all measurable functions f such that $x_2 \mapsto \int_{E_1} f(x_1, x_2)\mu_1(\mathrm{d}x_1)$ is $\mathcal{E}_2$-measurable.

(i) It is clear that $\mathbf{1}_E, \mathbf{1}_A \in V$ for all $A \in \mathcal{A}$ (where $\mathcal{A}$ is as in the definition of the product σ-algebra).

(ii) V is a vector space by linearity of the integral.

(iii) Suppose (f_n) is a non-negative sequence in V and $f_n \nearrow f$, then

$$\left(x_2 \mapsto \int_{E_1} f_n(x_1, x_2)\, \mu_1(\mathrm{d}x_1)\right) \nearrow \left(x_2 \mapsto \int_{E_1} f(x_1, x_2)\, \mu(\mathrm{d}x_1)\right)$$

by the monotone convergence theorem. So $f \in V$.

So the monotone class theorem tells us V contains all bounded measurable functions.

Now if f is a general non-negative measurable function, then $f \wedge n$ is bounded and measurable, hence $f \wedge n \in V$. Therefore $f \in V$ by the monotone convergence theorem. $\square$

Theorem. There exists a unique measurable function $\mu = \mu_1 \otimes \mu_2$ on $\mathcal{E}$ such that

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

for all $A_1 \times A_2 \in \mathcal{A}$.

Proof. One might be tempted to just apply the Caratheodory extension theorem, but we have a more direct way of doing it here, by using integrals. We define

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

Here the previous lemma is very important. It tells us that these integrals actually make sense!

We first check that this is a measure:

(i) $\mu(\emptyset) = 0$ is immediate since $1_\emptyset = 0$.

(ii) Suppose (A_n) is a disjoint sequence and $A = \bigcup A_n$. Then we have

$$\begin{aligned}\mu(A) &= \int_{E_1} \left(\int_{E_2} \mathbf{1}_A(x_1, x_2)\, \mu_2(\mathrm{d}x_2)\right) \mu_1(\mathrm{d}x_1)\\ &= \int_{E_1} \left(\int_{E_2} \sum_n \mathbf{1}_{A_n}(x_1, x_2)\, \mu_2(\mathrm{d}x_2)\right) \mu_1(\mathrm{d}x_1)\end{aligned}$$

We now use the fact that integration commutes with the sum of non-negative measurable functions to get

$$\begin{aligned}&= \int_{E_1} \left(\sum_n \left(\int_{E_2} \mathbf{1}_A(x_1, x_2)\, \mu_2(\mathrm{d}x_2)\right)\right) \mu_1(\mathrm{d}x_1)\\ &= \sum_n \int_{E_1} \left(\int_{E_2} \mathbf{1}_{A_n}(x_1, x_2)\, \mu_2(\mathrm{d}x_2)\right) \mu_1(\mathrm{d}x_1)\\ &= \sum_n \mu(A_n).\end{aligned}$$

So we have a working measure, and it clearly satisfies

$$\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2).$$

Uniqueness follows because μ is finite, and is thus characterized by its values on the π-system $\mathcal{A}$ that generates $\mathcal{E}$. □

Theorem (Fubini's theorem)**.**

(i) If f is non-negative measurable, then

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

In particular, we have

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

This is sometimes known as *Tonelli's theorem*.

(ii) If f is integrable, and

$$A = \left\{x_1 \in E : \int_{E_2} |f(x_1, x_2)|\mu_2(\mathrm{d}x_2) < \infty\right\}.$$

then

$$\mu_1(E_1 \setminus A) = 0.$$

If we set

$$f_1(x_1) = \begin{cases} \int_{E_2} f(x_1, x_2)\, \mu_2(\mathrm{d}x_2) & x_1 \in A\\ 0 & x_1 \not\in A\end{cases},$$

then f_1 is a μ_1 integrable function and

$$\mu_1(f_1) = \mu(f).$$

Proof.

(i) Let V be the set of all measurable functions such that $(*)$ holds. Then V is a vector space since integration is linear.

(a) By definition of μ, we know $\mathbf{1}_E$ and $\mathbf{1}_A$ are in V for all $A \in \mathcal{A}$.

(b) The monotone convergence theorem on both sides tell us that V is closed under monotone limits of the form $f_n \nearrow f$, $f_n \geq 0$.

By the monotone class theorem, we know V contains all bounded measurable functions. If f is non-negative measurable, then $(f \wedge n) \in V$, and monotone convergence for $f \wedge n \nearrow f$ gives that $f \in V$.

(ii) Assume that f is μ-integrable. Then

$$x_1 \mapsto \int_{E_2} |f(x_1, x_2)|\, \mu(\mathrm{d}x_2)$$

is $\mathcal{E}_1$-measurable, and, by (i), is μ_1-integrable. So A_1, being the inverse image of ∞ under that map, lies in $\mathcal{E}_1$. Moreover, $\mu_1(E_1 \setminus A_1) = 0$ because integrable functions can only be infinite on sets of measure 0.

We set

$$\begin{aligned} f_1^+(x_1) &= \int_{E_2} f^+(x_1, x_2)\, \mu_2(\mathrm{d}x_2) \\ f_1^-(x_1) &= \int_{E_2} f^-(x_1, x_2)\, \mu_2(\mathrm{d}x_2). \end{aligned}$$

Then we have

$$f_1 = (f_1^+ - f_1^-)\mathbf{1}_{A_1}.$$

So the result follows since

$$\mu(f) = \mu(f^+) - \mu(f^-) = \mu(f_1^+) - \mu_1(f_1^-) = \mu_1(f_1).$$

by (i). ◻

Proposition. Let $X_1, \cdots, X_n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(E_1, \mathcal{E}_1), \cdots, (E_n, \mathcal{E}_n)$ respectively. We define

$$E = E_1 \times \cdots \times E_n, \quad \mathcal{E} = \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n.$$

Then $X = (X_1, \cdots, X_n)$ is $\mathcal{E}$-measurable and the following are equivalent:

(i) $X_1, \cdots, X_n$ are independent.

(ii) $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$.

(iii) For any $f_1, \cdots, f_n$ bounded and measurable, we have

$$\mathbb{E}\left[\prod_{k=1}^n f_k(X_k)\right] = \prod_{k=1}^n \mathbb{E}[f_k(X_k)].$$

Proof.

- (i) $\Rightarrow$ (ii): Let $\nu = \mu_{X_1} \times \cdots \otimes \mu_{X_n}$. We want to show that $\nu = \mu_X$. To do so, we just have to check that they agree on a π-system generating the entire σ-algebra. We let

  $$\mathcal{A} = \{A_1 \times \cdots \times A_n : A_1 \in \mathcal{E}_1, \cdots, A_k \in \mathcal{E}_k\}.$$

  Then $\mathcal{A}$ is a generating π-system of $\mathcal{E}$. Moreover, if $A = A_1 \times \cdots \times A_n \in \mathcal{A}$, then we have

  $$\begin{aligned}\mu_X(A) &= \mathbb{P}[X \in A]\\ &= \mathbb{P}[X_1 \in A_1, \cdots, X_n \in A_n]\end{aligned}$$

  By independence, we have

  $$\begin{aligned}&= \prod_{k=1}^n \mathbb{P}[X_k \in A_k]\\ &= \nu(A).\end{aligned}$$

  So we know that $\mu_X = \nu = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$ on $\mathcal{E}$.

- (ii) $\Rightarrow$ (iii): By assumption, we can evaluate the expectation

  $$\begin{aligned}\mathbb{E}\left[\prod_{k=1}^n f_k(X_k)\right] &= \int_E \prod_{k=1}^n f_k(x_k)\mu(\mathrm{d}x_k)\\ &= \prod_{k=1}^n \int_{E_k} f(x_k)\mu_k(\mathrm{d}x_k)\\ &= \prod_{k=1}^n \mathbb{E}[f_k(X_k)].\end{aligned}$$

  Here in the middle we have used Fubini's theorem.

- (iii) $\Rightarrow$ (i): Take $f_k = \mathbf{1}_{A_k}$ for $A_k \in \mathcal{E}_k$. Then we have

  $$\begin{aligned}\mathbb{P}[X_1 \in A_1, \cdots, X_n \in A_n] &= \mathbb{E}\left[\prod_{k=1}^n \mathbf{1}_{A_k}(X_k)\right]\\ &= \prod_{k=1}^n \mathbb{E}[\mathbf{1}_{A_k}(X_k)]\\ &= \prod_{k=1}^n \mathbb{P}[X_k \in A_k]\end{aligned}$$

  So $X_1, \cdots, X_n$ are independent. □

4 Inequalities and L^p spaces

4.1 Four inequalities

Proposition (Chebyshev's/Markov's inequality)**.** Let f be non-negative measurable and $\lambda > 0$. Then

$$\mu(\{f \geq \lambda\}) \leq \frac{1}{\lambda}\mu(f).$$

Proof. We write

$$f \geq f\mathbf{1}_{f\geq\lambda} \geq \lambda\mathbf{1}_{f\geq\lambda}.$$

Taking μ gives the desired answer. □

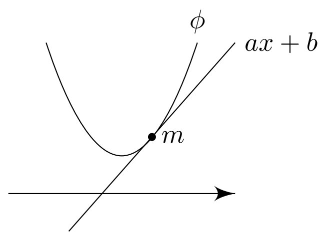
Proposition (Jensen's inequality)**.** Let X be an integrable random variable with values in I. If $c : I \to \mathbb{R}$ is convex, then we have

$$\mathbb{E}[c(X)] \geq c(E[X]).$$

Lemma. If $c : I \to \mathbb{R}$ is a convex function and m is in the interior of I, then there exists real numbers a, b such that

$$c(x) \geq ax + b$$

for all $x \in I$, with equality at $x = m$.



Proof. If c is smooth, then we know $c'' \geq 0$, and thus c' is non-decreasing. We are going to show an analogous statement that does not mention the word "derivative". Consider $x < m < y$ with $x, y, m \in I$. We want to show that

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

To show this, we turn off our brains and do the only thing we can do. We can write

$$m = tx + (1 - t)y$$

for some t. Then convexity tells us

$$c(m) \leq tc(x) + (1 - t)c(y).$$

Writing $c(m) = tc(m) + (1 - t)c(m)$, this tells us

$$t(c(m) - c(x)) \leq (1 - t)(c(y) - c(m)).$$

To conclude, we simply have to compute the actual value of t and plug it in. We have

$$t = \frac{y-m}{y-x}, \quad 1-t = \frac{m-x}{y-x}.$$

So we obtain

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

Cancelling the $y-x$ and dividing by the factors gives the desired result.

Now since x and y are arbitrary, we know there is some $a \in \mathbb{R}$ such that

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

for all $x < m < y$. If we rearrange, then we obtain

$$c(t) \geq a(t-m) + c(m)$$

for all $t \in I$. □

Proof of Jensen's inequality. To apply the previous result, we need to pick a right m. We take

$$m = \mathbb{E}[X].$$

To apply this, we need to know that m is in the interior of I. So we assume that X is *not* a.s. constant (that case is boring). By the lemma, we can find some $a, b \in \mathbb{R}$ such that

$$c(X) \geq aX + b.$$

We want to take the expectation of the LHS, but we have to make sure the $\mathbb{E}[c(X)]$ is a sensible thing to talk about. To make sure it makes sense, we show that $\mathbb{E}[c(X)^-] = \mathbb{E}[(-c(X)) \vee 0]$ is finite.

We simply bound

$$[c(X)]^- = [-c(X)] \vee 0 \leq |a||X| + |b|.$$

So we have

$$\mathbb{E}[c(X)^-] \leq |a|\mathbb{E}|X| + |b| < \infty$$

since X is integrable. So $\mathbb{E}[c(X)]$ makes sense.

We then just take

$$\mathbb{E}[c(X)] \geq \mathbb{E}[aX+b] = a\mathbb{E}[X] + b = am + b = c(m) = c(\mathbb{E}[X]).$$

So done. □

Proposition (Hölder's inequality)**.** Let $p, q \in (1, \infty)$ be conjugate. Then for f, g measurable, we have

$$\mu(|fg|) = \|fg\|_1 \leq \|f\|_p \|g\|_q.$$

When $p = q = 2$, then this is the Cauchy-Schwarz inequality.

Proof. We assume that $\|f\|_p > 0$ and $\|f\|_p < \infty$. Otherwise, there is nothing to prove. By scaling, we may assume that $\|f\|_p = 1$. We make up a probability measure by

$$\mathbb{P}[A] = \int |f|^p \mathbf{1}_A \, \mathrm{d}\mu.$$

Since we know

$$\|f\|_p = \left(\int |f|^p \, \mathrm{d}\mu\right)^{1/p} = 1,$$

we know $\mathbb{P}[\,\cdot\,]$ is a probability measure. Then we have

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

Now use the fact that $(\mathbb{E}|X|)^q \leq \mathbb{E}[|X|^q]$ since $x \mapsto x^q$ is convex for $q > 1$. Then we obtain

$$\leq \left(\mathbb{E}\left[\frac{|g|^q}{|f|^{(p-1)q}}\mathbf{1}_{\{|f|>0\}}\right]\right)^{1/q}.$$

The key realization now is that $\frac{1}{q} + \frac{1}{p} = 1$ means that $q(p-1) = p$. So this becomes

$$\mathbb{E}\left[\frac{|g|^q}{|f|^p}\mathbf{1}_{\{|f|>0\}}\right]^{1/q} = \mu(|g|^q)^{1/q} = \|g\|_q.$$

Using the fact that $\|f\|_p = 1$, we obtain the desired result. □

Alternative proof. We wlog $0 < \|f\|_p, \|g\|_q < \infty$, or else there is nothing to prove. By scaling, we wlog $\|f\|_p = \|g\|_q = 1$. Then we have to show that

$$\int |f||g| \, \mathrm{d}\mu \leq 1.$$

To do so, we notice if $\frac{1}{p} + \frac{1}{q} = 1$, then the concavity of log tells us for any $a, b > 0$, we have

$$\frac{1}{p}\log a + \frac{1}{q}\log b \leq \log\left(\frac{a}{p} + \frac{b}{q}\right).$$

Replacing a with a^p; b with b^p and then taking exponentials tells us

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

While we assumed $a, b > 0$ when deriving, we observe that it is also valid when some of them are zero. So we have

$$\int |f||g| \, \mathrm{d}\mu \leq \int \left(\frac{|f|^p}{p} + \frac{|g|^q}{q}\right) \, \mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1.$$ □

Lemma. Let $a, b \geq 0$ and $p \geq 1$. Then

$$(a+b)^p \leq 2^p(a^p + b^p).$$

Proof. We wlog $a \leq b$. Then

$$(a+b)^p \leq (2b)^p \leq 2^p b^p \leq 2^p(a^p + b^p). \qquad \square$$

Theorem (Minkowski inequality)**.** Let $p \in [1, \infty]$ and f, g measurable. Then

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

Proof. We do the boring cases first. If $p = 1$, then

$$\|f+g\|_1 = \int |f+g| \leq \int (|f| + |g|) = \int |f| + \int |g| = \|f\|_1 + \|g\|_1.$$

The proof of the case of $p = \infty$ is similar.

Now note that if $\|f + g\|_p = 0$, then the result is trivial. On the other hand, if $\|f + g\|_p = \infty$, then since we have

$$|f+g|^p \leq (|f| + |g|)^p \leq 2^p(|f|^p + |g|^p),$$

we know the right hand side is infinite as well. So this case is also done.

Let's now do the interesting case. We compute

$$\begin{aligned}
\mu(|f+g|^p) &= \mu(|f+g||f+g|^{p-1})\\
&\leq \mu(|f||f+g|^{p-1}) + \mu(|g||f+g|^{p-1})\\
&\leq \|f\|_p \||f+g|^{p-1}\|_q + \|g\|_p \||f+g|^{p-1}\|_q\\
&= (\|f\|_p + \|g\|_p)\||f+g|^{p-1}\|_q\\
&= (\|f\|_p + \|g\|_p)\mu(|f+g|^{(p-1)q})^{1-1/p}\\
&= (\|f\|_p + \|g\|_p)\mu(|f+g|^p)^{1-1/p}.
\end{aligned}$$

So we know

$$\mu(|f+g|^p) \leq (\|f\|_p + \|g\|_p)\mu(|f+g|^p)^{1-1/p}.$$

Then dividing both sides by $(\mu(|f+g|^p)^{1-1/p}$ tells us

$$\mu(|f+g|^p)^{1/p} = \|f+g\|_p \leq \|f\|_p + \|g\|_p. \qquad \square$$

4.2 L^p spaces

Theorem. Let $1 \leq p \leq \infty$. Then $\mathcal{L}^p$ is a Banach space. In other words, if (f_n) is a sequence in L^p, with the property that $\|f_n - f_m\|_p \to 0$ as $n, m \to \infty$, then there is some $f \in L^p$ such that $\|f_n - f\|_p \to 0$ as $n \to \infty$.

Proof. We will only give the proof for $p < \infty$. The $p = \infty$ case is left as an exercise for the reader.

Suppose that (f_n) is a sequence in L^p with $\|f_n - f_m\|_p \to 0$ as $n, m \to \infty$. Take a subsequence (f_{n_k}) of (f_n) with

$$\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$$

for all $k \in \mathbb{N}$. We then find that

$$\left\| \sum_{k=1}^M |f_{n_{k+1}} - f_{n_k}| \right\|_p \leq \sum_{k=1}^M \|f_{n_{k+1}} - f_{n_k}\|_p \leq 1.$$

We know that

$$\sum_{k=1}^{M} |f_{n_{k+1}} - f_{n_k}| \nearrow \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \text{ as } M \to \infty.$$

So applying the monotone convergence theorem, we know that

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

In particular,

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

So $f_{n_k}(x)$ converges a.e., since the real line is complete. So we set

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

By an exercise on the first example sheet, this function is indeed measurable. Then we have

$$\begin{aligned} \|f_n - f\|_p^p &= \mu(|f_n - f|^p) \\ &= \mu\left(\liminf_{k\to\infty} |f_n - f_{n_k}|^p\right) \\ &\leq \liminf_{k\to\infty} \mu(|f_n - f_{n_k}|^p), \end{aligned}$$

which tends to 0 as $n \to \infty$ since the sequence is Cauchy. So f is indeed the limit.

Finally, we have to check that $f \in L^p$. We have

$$\begin{aligned} \mu(|f|^p) &= \mu(|f - f_n + f_n|^p) \\ &\leq \mu((|f - f_n| + |f_n|)^p) \\ &\leq \mu(2^p(|f - f_n|^p + |f_n|^p)) \\ &= 2^p(\mu(|f - f_n|^p) + \mu(|f_n|^p)^2) \end{aligned}$$

We know the first term tends to 0, and in particular is finite for n large enough, and the second term is also finite. So done. □

4.3 Orthogonal projection in $\mathcal{L}^2$

Theorem. Let V be a closed subspace of L^2. Then each $f \in L^2$ has an *orthogonal decomposition*

$$f = u + v,$$

where $v \in V$ and $u \in V^\perp$. Moreover,

$$\|f - v\|_2 \leq \|f - g\|_2$$

for all $g \in V$ with equality iff $g \sim v$.

Lemma (Pythagoras identity).

$$\|f+g\|^2 = \|f\|^2 + \|g\|^2 + 2\langle f, g\rangle.$$

Lemma (Parallelogram law).

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2).$$

Proof of orthogonal decomposition. Given $f \in L^2$, we take a sequence (g_n) in V such that

$$\|f - g_n\|_2 \to d(f, V) = \inf_g \|f - g\|_2.$$

We now want to show that the infimum is attained. To do so, we show that g_n is a Cauchy sequence, and by the completeness of L^2, it will have a limit.

If we apply the parallelogram law with $u = f - g_n$ and $v = f - g_m$, then we know

$$\|u+v\|_2^2 + \|u-v\|_2^2 = 2(\|u\|_2^2 + \|v\|_2^2).$$

Using our particular choice of u and v, we obtain

$$\left\|2\left(f - \frac{g_n+g_m}{2}\right)\right\|_2^2 + \|g_n - g_m\|_2^2 = 2(\|f-g_n\|_2^2 + \|f-g_m\|_2^2).$$

So we have

$$\|g_n - g_m\|_2^2 = 2(\|f-g_n\|_2^2 + \|f-g_m\|_2^2) - 4\left\|f - \frac{g_n+g_m}{2}\right\|_2^2.$$

The first two terms on the right hand side tend to $d(f,V)^2$, and the last term is bounded below in magnitude by $4d(f,V)$. So as $n, m \to \infty$, we must have $\|g_n - g_m\|_2 \to 0$. By completeness of $\mathcal{L}^2$, there exists a $g \in L^2$ such that $g_n \to g$.

Now since V is assumed to be closed, we can find a $v \in V$ such that $g = v$ a.e. Then we know

$$\|f - v\|_2 = \lim_{n\to\infty} \|f - g_n\|_2 = d(f, V).$$

So v attains the infimum. To show that this gives us an orthogonal decomposition, we want to show that

$$u = f - v \in V^\perp.$$

Suppose $h \in V$. We need to show that $\langle u, h\rangle = 0$. We need to do another funny trick. Suppose $t \in \mathbb{R}$. Then we have

$$\begin{aligned} d(f,V)^2 &\leq \|f - (v + th)\|_2^2 \\ &= \|f - v\|^2 + t^2\|h\|_2^2 - 2t\langle f - v, h\rangle. \end{aligned}$$

We think of this as a quadratic in t, which is minimized when

$$t = \frac{\langle f - v, h\rangle}{\|h\|_2^2}.$$

But we know this quadratic is minimized when $t = 0$. So $\langle f - v, h\rangle = 0$. $\square$

Proposition. The conditional expectation of X given $\mathcal{G}$ is the projection of X onto the subspace $L^2(\mathcal{G}, \mathbb{P})$ of $\mathcal{G}$-measurable L^2 random variables in the ambient space $L^2(\mathbb{P})$.

Proof. Let Y be the conditional expectation. It suffices to show that $\mathbb{E}[(X-W)^2]$ is minimized for $W = Y$ among $\mathcal{G}$-measurable random variables. Suppose that W is a $\mathcal{G}$-measurable random variable. Since

$$\mathcal{G} = \sigma(G_n : n \in \mathbb{N}),$$

it follows that

$$W = \sum_{n=1}^{\infty} a_n \mathbf{1}_{G_n}.$$

where $a_n \in \mathbb{R}$. Then

$$\begin{aligned}
\mathbb{E}[(X-W)^2] &= \mathbb{E}\left[\left(\sum_{n=1}^{\infty}(X - a_n)\mathbf{1}_{G_n}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{n}(X^2 + a_n^2 - 2a_n X)\mathbf{1}_{G_n}\right] \\
&= \mathbb{E}\left[\sum_{n}(X^2 + a_n^2 - 2a_n \mathbb{E}[X \mid G_n])\mathbf{1}_{G_n}\right]
\end{aligned}$$

We now optimize the quadratic

$$X^2 + a_n^2 - 2a_n\mathbb{E}[X \mid G_n]$$

over a_n. We see that this is minimized for

$$a_n = \mathbb{E}[X \mid G_n].$$

Note that this does not depend on what X is in the quadratic, since it is in the constant term.

Therefore we know that $\mathbb{E}[X \mid G_n]$ is minimized for $W = Y$. □

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

Theorem (Bounded convegence theorem)**.** Suppose $X, (X_n)$ are random variables. Assume that there exists a (non-random) constant $C > 0$ such that $|X_n| \leq C$. If $X_n \to X$ in probability, then $X_n \to X$ in L^1.

Proof. We first show that $|X| \leq C$ a.e. Let $\varepsilon > 0$. We then have

$$\begin{aligned}
\mathbb{P}[|X| > C + \varepsilon] &\leq \mathbb{P}[|X - X_n| + |X_n| > C + \varepsilon] \\
&\leq \mathbb{P}[|X - X_n| > \varepsilon] + \mathbb{P}[|X_n| > C]
\end{aligned}$$

We know the second term vanishes, while the first term $\to 0$ as $n \to \infty$. So we know

$$\mathbb{P}[|X| > C + \varepsilon] = 0$$

for all ε. Since ε was arbitrary, we know $|X| \leq C$ a.s.

Now fix an $\varepsilon > 0$. Then

$$\begin{aligned}\mathbb{E}[|X_n - X| &= \mathbb{E}\left[|X_n - X|(\mathbf{1}_{|X_n - X| \leq \varepsilon} + \mathbf{1}_{|X_n - X| > \varepsilon})\right] \\ &\leq \varepsilon + 2C\,\mathbb{P}\left[|X_n - X| > \varepsilon\right].\end{aligned}$$

Since $X_n \to X$ in probability, for N sufficiently large, the second term is $\leq \varepsilon$. So $\mathbb{E}[|X_n - X|] \leq 2\varepsilon$, and we have convergence in L^1. □

Proposition. Finite unions of uniformly integrable sets are uniformly integrable.

Proposition. Let $\mathcal{X}$ be an L^p-bounded family for some $p > 1$. Then $\mathcal{X}$ is uniformly integrable.

Proof. We let

$$C = \sup\{\|X\|_p : X \in \mathcal{X}\} < \infty.$$

Suppose that $X \in \mathcal{X}$ and $A \in \mathcal{F}$. We then have

$$\mathbb{E}[|X|\mathbf{1}_A] =\leq \mathbb{E}[|X|^p]^{1/p}\mathbb{P}[A]^{1/q} \leq C\mathbb{P}[A]^{1/q}.$$

by Hölder's inequality, where p, q are conjugates. This is now a uniform bound depending only on $\mathbb{P}[A]$. So done. □

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

$$\sup\{\mathbb{E}[|X|\mathbf{1}_{|X|>k}] : X \in \mathcal{X}\} \to 0$$

as $k \to \infty$.

Proof.

(⇒) Suppose that χ is uniformly integrable. For any k, and $X \in \mathcal{X}$ by Chebyshev inequality, we have

$$\mathbb{P}[|X| \geq k] \leq \frac{\mathbb{E}[X]}{k}.$$

Given $\varepsilon > 0$, we pick δ such that $\mathbb{P}[|X|\mathbf{1}_A] < \varepsilon$ for all A with $\mu(A) < \delta$. Then pick k sufficiently large such that $k\delta < \sup\{\mathbb{E}[X] : X \in \mathcal{X}\}$. Then $\mathbb{P}[|X| \geq k] < \delta$, and hence $\mathbb{E}[|X|\mathbf{1}_{|X|>k}] < \varepsilon$ for all $X \in \mathcal{X}$.

(⇐) Suppose that the condition in the lemma holds. We first show that $\mathcal{X}$ is L^1-bounded. We have

$$\mathbb{E}[|X|] = \mathbb{E}[|X|(\mathbf{1}_{|X|\leq k} + \mathbf{1}_{|X|>k})] \leq k + \mathbb{E}[|X|\mathbf{1}_{|X|>k}] < \infty$$

by picking a large enough k.

Next note that for any measurable A and $X \in \mathcal{X}$, we have

$$\mathbb{E}[|X|\mathbf{1}_A] = \mathbb{E}[|X|\mathbf{1}_A(\mathbf{1}_{|X|>k} + \mathbf{1}_{|X|\leq k})] \leq \mathbb{E}[|X|\mathbf{1}_{|X|>k}] + k\mathbb{P}[A].$$

Thus, for any $\varepsilon > 0$, we can pick k sufficiently large such that the first term is $< \frac{\varepsilon}{2}$ for all $X \in \mathcal{X}$ by assumption. Then when $\mathbb{P}[A] < \frac{\varepsilon}{2k}$, we have $\mathbb{E}|X|\mathbf{1}_A] \leq \varepsilon$. □

Corollary. Let $\mathcal{X} = \{X\}$, where $X \in L^1(\mathbb{P})$. Then $\mathcal{X}$ is uniformly integrable.

Hence, a finite collection of L^1 functions is uniformly integrable.

Proof. Note that

$$\mathbb{E}[|X|] = \sum_{k=0}^{\infty} \mathbb{E}[|X|\mathbf{1}_{X\in[k,k+1)}].$$

Since the sum is finite, we must have

$$\mathbb{E}[|X|\mathbf{1}_{|X|\geq K}] = \sum_{k=K}^{\infty} \mathbb{E}[|X|\mathbf{1}_{X\in[k,k+1)}] \to 0.$$

□

Theorem. Let $X, (X_n)$ be random variables. Then the following are equivalent:

(i) $X_n, X \in L^1$ for all n and $X_n \to X$ in L^1.

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

Proof. We first assume that X_n, X are L^1 and $X_n \to X$ in L^1. We want to show that $\{X_n\}$ is uniformly integrable and $X_n \to X$ in probability.

We first show that $X_n \to X$ in probability. This is just going to come from the Chebyshev inequality. For $\varepsilon > 0$. Then we have

$$\mathbb{P}[|X - X_n| > \varepsilon] \leq \frac{\mathbb{E}[|X - X_n|]}{\varepsilon} \to 0$$

as $n \to \infty$.

Next we show that $\{X_n\}$ is uniformly integrable. Fix $\varepsilon > 0$. Take N such that $n \geq N$ implies $\mathbb{E}[|X - X_n|] \leq \frac{\varepsilon}{2}$. Since *finite* families of L^1 random variables are uniformly integrable, we can pick $\delta > 0$ such that $A \in \mathcal{F}$ and $\mathbb{P}[A] < \delta$ implies

$$\mathbb{E}[X\mathbf{1}_A], \mathbb{E}[|X_n|\mathbf{1}_A] \leq \frac{\varepsilon}{2}$$

for $n = 1, \cdots, N$.

Now when $n > N$ and $A \in \mathcal{F}$ with $\mathbb{P}[A] \leq \delta$, then we have

$$\begin{aligned}
\mathbb{E}[|X_n|\mathbf{1}_A] &\leq \mathbb{E}[|X - X_n|\mathbf{1}_A] + \mathbb{E}[|X|\mathbf{1}_A]\\
&\leq \mathbb{E}[|X - X_n|] + \frac{\varepsilon}{2}\\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\\
&= \varepsilon.
\end{aligned}$$

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

Assume that $\{X_n\}$ is uniformly integrable and $X_n \to X$ in probability.

The first step is to show that $X \in L^1$. We want to use Fatou's lemma, but to do so, we want almost sure convergence, not just convergence in probability.

Recall that we have previously shown that there is a subsequence (X_{n_k}) of (X_n) such that $X_{n_k} \to X$ a.s. Then we have

$$\mathbb{E}[|X|] = \mathbb{E}\left[\liminf_{k\to\infty} |X_{n_k}|\right] \leq \liminf_{k\to\infty} \mathbb{E}[|X_{n_k}|] < \infty$$

since uniformly integrable families are L^1 bounded. So $\mathbb{E}[|X|] < \infty$, hence $X \in L^1$.

Next we want to show that $X_n \to X$ in L^1. Take $\varepsilon > 0$. Then there exists $K \in (0, \infty)$ such that

$$\mathbb{E}\left[|X|\mathbf{1}_{\{|X|>K\}}\right], \mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|>K\}}\right] \leq \frac{\varepsilon}{3}.$$

To set things up so that we can use the bounded convergence theorem, we have to invent new random variables

$$X_n^K = (X_n \vee -K) \wedge K, \quad X^K = (X \vee -K) \wedge K.$$

Since $X_n \to X$ in probability, it follows that $X_n^K \to X^K$ in probability.

Now bounded convergence tells us that there is some N such that $n \geq N$ implies

$$\mathbb{E}[|X_n^K - X^K|] \leq \frac{\varepsilon}{3}.$$

Combining, we have for $n \geq N$ that

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n^K - X^K|] + \mathbb{E}[|X|\mathbf{1}_{\{|X|\geq K\}}] + \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|\geq K\}}] \leq \varepsilon.$$

So we know that $X_n \to X$ in L^1. $\square$

5 Fourier transform

5.1 The Fourier transform

Proposition.

$$\|\hat{f}\|_\infty \leq \|f\|_1, \quad \|\hat{\mu}\|_\infty \leq \mu(\mathbb{R}^d).$$

Proposition. The functions $\hat{f}, \hat{\mu}$ are continuous.

Proof. If $u_n \to u$, then

$$f(x)e^{i(u_n, x)} \to f(x)e^{i(u,x)}.$$

Also, we know that

$$|f(x)e^{i(u_n,x)}| = |f(x)|.$$

So we can apply dominated convergence theorem with $|f|$ as the bound. $\square$

5.2 Convolutions

Proposition. For any $f \in L^p$ and ν a probability measure, we have

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

Proposition.

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

Proof. We have

$$\begin{aligned}
\widehat{f * \nu}(u) &= \int \left(\int f(x - y)\nu(\mathrm{d}y)\right) e^{i(u,x)}\;\mathrm{d}x\\
&= \iint f(x-y)e^{i(u,x)}\;\mathrm{d}x\;\nu(\mathrm{d}y)\\
&= \int \left(\int f(x-y)e^{i(u,x-y)}\;\mathrm{d}(x-y)\right) e^{i(u,y)}\;\mu(\mathrm{d}y)\\
&= \int \left(\int f(x)e^{i(u,x)}\;\mathrm{d}(x)\right) e^{i(u,y)}\;\mu(\mathrm{d}y)\\
&= \int \hat{f}(u)e^{i(u,x)}\;\mu(\mathrm{d}y)\\
&= \hat{f}(u)\int e^{i(u,x)}\;\mu(\mathrm{d}y)\\
&= \hat{f}(u)\hat{\nu}(u).
\end{aligned}$$

$\square$

Proposition. Let μ, ν be probability measures, and X, Y be independent variables with laws μ, ν respectively. Then

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

Proof. We have

$$\widehat{\mu * \nu}(u) = \mathbb{E}[e^{i(u, X + Y)}] = \mathbb{E}[e^{i(u,X)}]\mathbb{E}[e^{i(u,Y)}] = \hat{\mu}(u)\hat{\nu}(u).$$

$\square$

5.3 Fourier inversion formula

Theorem (Fourier inversion formula)**.** Let $f, \hat{f} \in L^1$. Then

$$f(x) = \frac{1}{(2\pi)^d} \int \hat{f}(u) e^{-i(u,x)}\;\mathrm{d}u \text{ a.e.}$$

Proposition. Let $Z \sim N(0, 1)$. Then

$$\phi_Z(a) = e^{-u^2/2}.$$

Proof. We have

$$\begin{aligned} \phi_Z(u) &= \mathbb{E}[e^{iuZ}] \\ &= \frac{1}{\sqrt{2\pi}} \int e^{iux} e^{-x^2/2}\;\mathrm{d}x. \end{aligned}$$

We now notice that the function is bounded, so we can differentiate under the integral sign, and obtain

$$\begin{aligned} \phi_Z'(u) &= \mathbb{E}[iZe^{iuZ}] \\ &= \frac{1}{\sqrt{2\pi}} \int ixe^{iux} e^{-x^2/2}\;\mathrm{d}x \\ &= -u\phi_Z(u), \end{aligned}$$

where the last equality is obtained by integrating by parts. So we know that $\phi_Z(u)$ solves

$$\phi_Z'(u) = -u\phi_Z(u).$$

This is easy to solve, since we can just integrate this. We find that

$$\log \phi_Z(u) = -\frac{1}{2}u^2 + C.$$

So we have

$$\phi_Z(u) = Ae^{-u^2/2}.$$

We know that $A = 1$, since $\phi_Z(0) = 1$. So we have

$$\phi_Z(u) = e^{-u^2/2}. \qquad \square$$

Proposition. Let $Z = (Z_1, \cdots, Z_d)$ with $Z_j \sim N(0, 1)$ independent. Then $\sqrt{t}Z$ has density

$$g_t(x) = \frac{1}{(2\pi t)^{d/2}} e^{-|x|^2/(2t)}.$$

with

$$\hat{g}_t(u) = e^{-|u|^2t/2}.$$

Proof. We have

$$\begin{aligned}\hat{g}_t(u) &= \mathbb{E}[e^{i(u,\sqrt{t}Z)}]\\ &= \prod_{j=1}^d \mathbb{E}[e^{i(u_j,\sqrt{t}Z_j)}]\\ &= \prod_{j=1}^d \phi_Z(\sqrt{t}u_j)\\ &= \prod_{j=1}^d e^{-tu_j^2/2}\\ &= e^{-|u|^2t/2}.\end{aligned}$$

□

Lemma. The Fourier inversion formula holds for the Gaussian density function.

Proposition.

$$\|f * g_t\|_1 \leq \|f\|_1.$$

Proposition.

$$\|f * g_t\|_\infty \leq (2\pi t)^{-d/2}\|f\|_1.$$

Proposition.

$$\|\widehat{f * g_t}\|_1 = \|\hat{f}\hat{g}_t\|_1 \leq (2\pi)^{d/2}t^{-d/2}\|\hat{f}\|_1,$$

and

$$\|\widehat{f * g_t}\|_\infty \leq \|\hat{f}\|_\infty.$$

Lemma. The Fourier inversion formula holds for Gaussian convolutions.

Proof. We have

$$\begin{aligned}f * g_t(x) &= \int f(x-y)g_t(y)\;\mathrm{d}y\\ &= \int f(x-y)\left(\frac{1}{(2\pi)^d}\int \hat{g}_t(u)e^{-i(u,y)}\;\mathrm{d}u\right)\;\mathrm{d}y\\ &= \left(\frac{1}{2\pi}\right)^d \iint f(x-y)\hat{g}_t(u)e^{-i(u,y)}\;\mathrm{d}u\;\mathrm{d}y\\ &= \left(\frac{1}{2\pi}\right)^d \int \left(\int f(x-y)e^{-i(u,x-y)}\;\mathrm{d}y\right)\hat{g}_t(u)e^{-i(u,x)}\;\mathrm{d}u\\ &= \left(\frac{1}{2\pi}\right)^d \int \hat{f}(u)\hat{g}_t(u)e^{-i(u,x)}\;\mathrm{d}u\\ &= \left(\frac{1}{2\pi}\right)^d \int \widehat{f * g_t}(u)e^{-i(u,x)}\;\mathrm{d}u\end{aligned}$$

So done. □

Theorem (Fourier inversion formula)**.** Let $f \in L^1$ and

$$f_t(x) = (2\pi)^{-d}\int \hat{f}(u)e^{-|u|^2t/2}e^{-i(u,x)}\;\mathrm{d}u = (2\pi)^{-d}\int \widehat{f * g_t}(u)e^{-i(u,x)}\;\mathrm{d}u.$$

Then $\|f_t - f\|_1 \to 0$, as $t \to 0$, and the Fourier inversion holds whenever $f, \hat{f} \in L^1$.

Lemma. Suppose that $f \in L^p$ with $p \in [1, \infty)$. Then $\|f * g_t - f\|_p \to 0$ as $t \to 0$.

Proof. We fix $\varepsilon > 0$. By a question on the example sheet, we can find h which is continuous and with compact support such that $\|f - h\|_p \leq \frac{\varepsilon}{3}$. So we have

$$\|f * g_t - h * g_t\|_p = \|(f - h) * g_t\|_p \leq \|f - h\|_p \leq \frac{\varepsilon}{3}.$$

So it suffices for us to work with a continuous function h with compact support. We let

$$e(y) = \int |h(x - y) - h(x)|^p \, \mathrm{d}x.$$

We first show that e is a bounded function:

$$\begin{aligned} e(y) &\leq \int 2^p (|h(x - y)|^p + |h(x)|^p) \, \mathrm{d}x \\ &= 2^{p+1} \|h\|_p^p. \end{aligned}$$

Also, since h is continuous and bounded, the dominated convergence theorem tells us that $e(y) \to 0$ as $y \to 0$.

Moreover, using the fact that $\int g_t(y) \, \mathrm{d}y = 1$, we have

$$\|h * g_t - h\|_p^p = \int \left| \int (h(x - y) - h(x)) g_t(y) \, \mathrm{d}y \right|^p \, \mathrm{d}x$$

Since $g_t(y) \, \mathrm{d}y$ is a probability measure, by Jensen's inequality, we can bound this by

$$\begin{aligned} &\leq \iint |h(x - y) - h(x)|^p g_t(y) \, \mathrm{d}y \, \mathrm{d}x \\ &= \int \left(\int |h(x - y) - h(x)|^p \, \mathrm{d}x \right) g_t(y) \, \mathrm{d}y \\ &= \int e(y) g_t(y) \, \mathrm{d}y \\ &= \int e(\sqrt{t} y) g_1(y) \, \mathrm{d}y, \end{aligned}$$

where we used the definition of g and substitution. We know that this tends to 0 as $t \to 0$ by the bounded convergence theorem, since we know that e is bounded.

Finally, we have

$$\begin{aligned} \|f * g_t - f\|_p &\leq \|f * g_t - h * g_t\|_p + \|h * g_t - h\|_p + \|h - f\|_p \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \|h * g_t - h\|_p \\ &= \frac{2\varepsilon}{3} + \|h * g_t - h\|_p. \end{aligned}$$

Since we know that $\|h * g_t - h\|_p \to 0$ as $t \to 0$, we know that for all sufficiently small t, the function is bounded above by ε. So we are done. $\square$

Proof of Fourier inversion theorem. The first part is just a special case of the previous lemma. Indeed, recall that

$$\widehat{f * g_t}(u) = \hat{f}(u) e^{-|u|^2 t/2}.$$

Since Gaussian convolutions satisfy Fourier inversion formula, we know that

$$f_t = f * g_t.$$

So the previous lemma says exactly that $\|f_t - f\|_1 \to 0$.

Suppose now that $\hat{f} \in L^1$ as well. Then looking at the integrand of

$$f_t(x) = (2\pi)^{-d} \int \hat{f}(u) e^{-|u|^2 t/2} e^{-i(u,x)} \;\mathrm{d}u,$$

we know that

$$\left|\hat{f}(u) e^{-|u|^2 t/2} e^{-i(u,x)}\right| \leq |\hat{f}|.$$

Then by the dominated convergence theorem with dominating function $|\hat{f}|$, we know that this converges to

$$f_t(x) \to (2\pi)^{-d} \int \hat{f}(u) e^{-i(u,x)} \;\mathrm{d}u \text{ as } t \to 0.$$

By the first part, we know that $\|f_t - f\|_1 \to 0$ as $t \to 0$. So we can find a sequence (t_n) with $t_n > 0$, $t_n \to 0$ so that $f_{t_n} \to f$ a.e. Combining these, we know that

$$f(x) = \int \hat{f}(u) e^{-i(u,x)} \;\mathrm{d}u \text{ a.e.}$$

So done. □

5.4 Fourier transform in $\mathcal{L}^2$

Theorem (Plancherel identity). For any function $f \in L^1 \cap L^2$, the *Plancherel identity* holds:

$$\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2.$$

Proof. We first work with the special case where $f, \hat{f} \in L^1$, since the Fourier inversion formula holds for f. We then have

$$\begin{aligned}
\|f\|_2^2 &= \int f(x) \overline{f(x)} \;\mathrm{d}x \\
&= \frac{1}{(2\pi)^d} \int \left(\int \hat{f}(u) e^{-i(u,x)} \;\mathrm{d}u\right) \overline{f(x)} \;\mathrm{d}x \\
&= \frac{1}{(2\pi)^d} \int \hat{f}(u) \left(\overline{f(x)} e^{-i(u,x)} \;\mathrm{d}x\right) \;\mathrm{d}u \\
&= \frac{1}{(2\pi)^d} \int \hat{f}(u) \overline{\left(f(x) e^{i(u,x)} \;\mathrm{d}x\right)} \;\mathrm{d}u \\
&= \frac{1}{(2\pi)^d} \int \hat{f}(u) \overline{\hat{f}(u)} \;\mathrm{d}u \\
&= \frac{1}{(2\pi)^d} \|\hat{f}(u)\|_2^2.
\end{aligned}$$

So the Plancherel identity holds for f.

To prove it for the general case, we use this result and an approximation argument. Suppose that $f \in L^1 \cap L^2$, and let $f_t = f * g_t$. Then by our earlier lemma, we know that

$$\|f_t\|_2 \to \|f\|_2 \text{ as } t \to 0.$$

Now note that

$$\hat{f}_t(u) = \hat{f}(u)\hat{g}_t(u) = \hat{f}(u)e^{-|u|^2 t/2}.$$

The important thing is that $e^{-|u|^2 t/2} \nearrow 1$ as $t \to 0$. Therefore, we know

$$\|\hat{f}_t\|_2^2 = \int |\hat{f}(u)|^2 e^{-|u|^2 t} \, \mathrm{d}u \to \int |\hat{f}(u)|^2 \, \mathrm{d}u = \|\hat{f}\|_2^2$$

as $t \to 0$, by monotone convergence.

Since $f_t, \hat{f}_t \in L^1$, we know that the Plancherel identity holds, i.e.

$$\|\hat{f}_t\|_2 = (2\pi)^{d/2}\|f_t\|_2.$$

Taking the limit as $t \to 0$, the result follows. □

Theorem. There exists a unique Hilbert space automorphism $F : \mathcal{L}^2 \to \mathcal{L}^2$ such that

$$F([f]) = [(2\pi)^{-d/2}\hat{f}]$$

whenever $f \in L^1 \cap L^2$.

Here $[f]$ denotes the equivalence class of f in $\mathcal{L}^2$, and we say $F : \mathcal{L}^2 \to \mathcal{L}^2$ is a Hilbert space automorphism if it is a linear bijection that preserves the inner product.

Proof. We define $F_0 : \mathcal{L}^1 \cap \mathcal{L}^2 \to \mathcal{L}^2$ by

$$F_0([f]) = [(2\pi)^{-d/2}\hat{f}].$$

By the Plancherel identity, we know F_0 preserves the L^2 norm, i.e.

$$\|F_0([f])\|_2 = \|[f]\|_2.$$

Also, we know that $\mathcal{L}^1 \cap \mathcal{L}^2$ is dense in $\mathcal{L}^2$, since even the continuous functions with compact support are dense. So we know F_0 extends uniquely to an isometry $F : \mathcal{L}^2 \to \mathcal{L}^2$.

Since it preserves distance, it is in particular injective. So it remains to show that the map is surjective. By Fourier inversion, the subspace

$$V = \{[f] \in \mathcal{L}^2 : f, \hat{f} \in L^1\}$$

is sent to itself by the map F. Also if $f \in V$, then $F^4[f] = [f]$ (note that applying it twice does not suffice, because we actually have $F^2[f](x) = [f](-x)$). So V is contained in the image F, and also V is dense in $\mathcal{L}^2$, again because it contains all Gaussian convolutions (we have $\hat{f}_t = \hat{f}\hat{g}_t$, and $\hat{f}$ is bounded and $\hat{g}_t$ is decaying exponentially). So we know that F is surjective. □

5.5 Properties of characteristic functions

Theorem. The characteristic function ϕ_X of a distribution μ_X of a random variable X determines μ_X. In other words, if X and $\tilde{X}$ are random variables and $\phi_X = \phi_{\tilde{X}}$, then $\mu_X = \mu_{\tilde{X}}$

Proof sketch. Use the Fourier inversion to show that ϕ_X determines $\mu_X(g) = \mathbb{E}[g(X)]$ for any bounded, continuous g. □

Theorem. If ϕ_X is integrable, then μ_X has a bounded, continuous density function

$$f_X(x) = (2\pi)^{-d} \int \phi_X(u) e^{-i(u,x)}\; \mathrm{d}u.$$

Proof sketch. Let $Z \sim N(0,1)$ be independent of X. Then $X + \sqrt{t}Z$ has a bounded continuous density function which, by Fourier inversion, is

$$f_t(x) = (2\pi)^{-d} \int \phi_X(u) e^{-|u|^2 t/2} e^{-i(u,x)}\; \mathrm{d}u.$$

Sending $t \to 0$ and using the dominated convergence theorem with dominating function $|\phi_X|$. □

Theorem. Let $X, (X_n)$ be random variables with values in $\mathbb{R}^d$. If $\phi_{X_n}(u) \to \phi_X(u)$ for each $u \in \mathbb{R}^d$, then $\mu_{X_n} \to \mu_X$ weakly.

Proof sketch. By the example sheet, it suffices to show that $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for all compactly supported $g \in C^\infty$. We then use Fourier inversion and convergence of characteristic functions to check that

$$\mathbb{E}[g(X_n + \sqrt{t}Z)] \to \mathbb{E}[g(X + \sqrt{t}Z)]$$

for all $t > 0$ for $Z \sim N(0,1)$ independent of $X, (X_n)$. Then we check that $\mathbb{E}[g(X_n + \sqrt{t}Z)]$ is close to $\mathbb{E}[g(X_n)]$ for $t > 0$ small, and similarly for X. □

5.6 Gaussian random variables

Proposition. Let $X \sim N(\mu, \sigma^2)$. Then

$$\mathbb{E}[X] = \mu, \quad \mathrm{var}(X) = \sigma^2.$$

Also, for any $a, b \in \mathbb{R}$, we have

$$aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Lastly, we have

$$\phi_X(u) = e^{-i\mu u - u^2\sigma^2/2}.$$

Proof. All but the last of them follow from direct calculation, and can be found in IA Probability.

For the last part, if $X \sim N(\mu, \sigma^2)$, then we can write

$$X = \sigma Z + \mu,$$

where $Z \sim N(0,1)$. Recall that we have previously found that the characteristic function of a $N(0,1)$ function is

$$\phi_Z(u) = e^{-|u|^2/2}.$$

So we have

$$\begin{aligned}\phi_X(u) &= \mathbb{E}[e^{iu(\sigma Z + \mu)}]\\ &= e^{iu\mu}\mathbb{E}[e^{iu\sigma \mathbb{Z}}]\\ &= e^{iu\mu}\phi_Z(iu\sigma)\\ &= e^{iu\mu - u^2\sigma^2/2}.\end{aligned}$$

□

Theorem. Let X be Gaussian on $\mathbb{R}^n$, and le tA be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then

(i) $AX + b$ is Gaussian on $\mathbb{R}^m$.

(ii) $X \in L^2$ and its law μ_X is determined by $\mu = \mathbb{E}[X]$ and $V = \mathrm{var}(X)$, the covariance matrix.

(iii) We have
$$\phi_X(u) = e^{i(u,\mu) - (u, Vu)/2}.$$

(iv) If V is invertible, then X has a density of
$$f_X(x) = (2\pi)^{-n/2}(\det V)^{-1/2}\exp\left(-\frac{1}{2}(x-\mu, V^{-1}(x-\mu))\right).$$

(v) If $X = (X_1, X_2)$ where $X_i \in \mathbb{R}^{n_i}$, then $\mathrm{cov}(X_1, X_2) = 0$ iff X_1 and X_2 are independent.

Proof.

(i) If $u \in \mathbb{R}^m$, then we have
$$(AX + b, u) = (AX, u) + (b, u) = (X, A^T u) + (b, u).$$
Since $(X, A^T u)$ is Gaussian and (b, u) is constant, it follows that $(AX + b, u)$ is Gaussian.

(ii) We know in particular that each component of X is a Gaussian random variable, which are in L^2. So $X \in L^2$. We will prove the second part of (ii) with (iii)

(iii) If $\mu = \mathbb{E}[X]$ and $V = \mathrm{var}(X)$, then if $u \in \mathbb{R}^n$, then we have
$$\mathbb{E}[(u, X)] = (u, \mu), \quad \mathrm{var}((u, X)) = (u, Vu).$$
So we know
$$(u, X) \sim N((u, \mu), (u, Vu)).$$
So it follows that
$$\phi_X(u) = \mathbb{E}[e^{i(u,X)}] = e^{i(u,\mu) - (u,Vu)/2}.$$
So μ and V determine the characteristic function of X, which in turn determines the law of X.

(iv) We start off with a boring Gaussian vector $Y = (Y_1, \cdots, Y_n)$, where the $Y_i \sim N(0, 1)$ are independent. Then the density of Y is

$$f_Y(y) = (2\pi)^{-n/2} e^{-|y|^2/2}.$$

We are now going to construct X from Y. We define

$$\tilde{X} = V^{1/2} Y + \mu.$$

This makes sense because V is always non-negative definite. Then $\tilde{X}$ is Gaussian with $\mathbb{E}[\tilde{X}] = \mu$ and $\operatorname{var}(\tilde{X}) = V$. Therefore X has the same distribution as $\tilde{X}$. Since V is assumed to be invertible, we can compute the density of $\tilde{X}$ using the change of variables formula.

(v) It is clear that if X_1 and X_2 are independent, then $\operatorname{cov}(X_1, X_2) = 0$.

Conversely, let $X = (X_1, X_2)$, where $\operatorname{cov}(X_1, X_2) = 0$. Then we have

$$V = \operatorname{var}(X) = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix}.$$

Then for $u = (u_1, u_2)$, we have

$$(u, Vu) = (u_1 V_{11} u_1) + (u_2, V_{22} u_2),$$

where $V_{11} = \operatorname{var}(X_1)$ and $V_{22} \operatorname{var}(X_2)$. Then we have

$$\begin{aligned} \phi_X(u) &= e^{i\mu u - (u, Vu)/2} \\ &= e^{i\mu_1 u_1 - (u_1, V_{11} u_1)/2} e^{i\mu_2 u_2 - (u_2, V_{22} u_2)/2} \\ &= \phi_{X_1}(u_1) \phi_{X_2}(u_2). \end{aligned}$$

So it follows that X_1 and X_2 are independent. □

6 Ergodic theory

Proposition. If f is integrable and Θ is measure-preserving. Then $f \circ \Theta$ is integrable and
$$\int f \circ \Theta \mathrm{d}\mu = \int_E f \mathrm{d}\mu.$$

Proposition. If Θ is ergodic and f is invariant, then there exists a constant c such that $f = c$ a.e.

Theorem. The shift map Θ is an ergodic, measure preserving transformation.

Proof. It is an exercise to show that Θ is measurable and measure preserving.

To show that Θ is ergodic, recall the definition of the tail σ-algebra
$$\mathcal{T}_n = \sigma(X_m : m \geq n+1), \quad \mathcal{T} = \bigcap_n \mathcal{T}_n.$$
Suppose that $A \in \prod_{n \in \mathbb{N}} A_n \in \mathcal{A}$. Then
$$\Theta^{-n}(A) = \{X_{n+k} \in A_k \text{ for all } k\} \in \mathcal{T}_n.$$
Since $\mathcal{T}_n$ is a σ-algebra, we and $\Theta^{-n}(A) \in \mathcal{T}_N$ for all $A \in \mathcal{A}$ and $\sigma(\mathcal{A}) = \mathcal{E}$, we know $\Theta^{-n}(A) \in \mathcal{T}_N$ for all $A \in \mathcal{E}$.

So if $A \in \mathcal{E}_\Theta$, i.e. $A = \Theta^{-1}(A)$, then $A \in \mathcal{T}_N$ for all N. So $A \in \mathcal{T}$.

From the Kolmogorov 0-1 law, we know either $\mu[A] = 1$ or $\mu[A] = 0$. So done. □

6.1 Ergodic theorems

Lemma (Maximal ergodic lemma)**.** Let f be integrable, and
$$S^* = \sup_{n \geq 0} S_n(f) \geq 0,$$
where $S_0(f) = 0$ by convention. Then
$$\int_{\{S^* > 0\}} f \, \mathrm{d}\mu \geq 0.$$

Proof. We let
$$S_n^* = \max_{0 \leq m \leq n} S_m$$
and
$$A_n = \{S_n^* > 0\}.$$
Now if $1 \leq m \leq n$, then we know
$$S_m = f + S_{m-1} \circ \Theta \leq f + S_n^* \circ \Theta.$$
Now on A_n, we have
$$S_n^* = \max_{1 \leq m \leq n} S_m,$$
since $S_0 = 0$. So we have
$$S_n^* \leq f + S_n^* \circ \Theta.$$

On A_n^C, we have
$$S_n^* = 0 \leq S_n^* \circ \Theta.$$
So we know
$$\begin{aligned}\int_E S_n^* \,\mathrm{d}\mu &= \int_{A_n} S_n^* \,\mathrm{d}\mu + \int_{A_n^C} S_n^* \,\mathrm{d}\mu\\ &\leq \int_{A_n} f \,\mathrm{d}\mu + \int_{A_n} S_n^* \circ \Theta \,\mathrm{d}\mu + \int_{A_n^C} S_n^* \circ \Theta \,\mathrm{d}\mu\\ &= \int_{A_n} f \,\mathrm{d}\mu + \int_E S_n^* \circ \Theta \,\mathrm{d}\mu\\ &= \int_{A_n} f \,\mathrm{d}\mu + \int_E S_n^* \,\mathrm{d}\mu\end{aligned}$$
So we know
$$\int_{A_n} f \,\mathrm{d}\mu \geq 0.$$
Taking the limit as $n \to \infty$ gives the result by dominated convergence with dominating function f. □

Theorem (Birkhoff's ergodic theorem)**.** Let $(E, \mathcal{E}, \mu)$ be σ-finite and f be integrable. There exists an invariant function $\bar{f}$ such that
$$\mu(|\bar{f}|) \leq \mu(|f|),$$
and
$$\frac{S_n(f)}{n} \to \bar{f} \text{ a.e.}$$
If Θ is ergodic, then $\bar{f}$ is a constant.

Theorem (von Neumann's ergodic theorem)**.** Let $(E, \mathcal{E}, \mu)$ be a *finite* measure space. Let $p \in [1, \infty)$ and assume that $f \in L^p$. Then there is some function $\bar{f} \in L^p$ such that
$$\frac{S_n(f)}{n} \to \bar{f} \text{ in } L^p.$$

Proof of Birkhoff's ergodic theorem. We first note that
$$\limsup_n \frac{S_n}{n}, \quad \limsup_n \frac{S_n}{n}$$
are invariant functions, Indeed, we know
$$\begin{aligned}S_n \circ \Theta &= f \circ \Theta + f \circ \Theta^2 + \cdots + f \circ \Theta^n\\ &= S_{n+1} - f\end{aligned}$$
So we have
$$\limsup_{n \to \infty} \frac{S_n \circ \Theta}{n} = \limsup_{n \to \infty} \frac{S_n}{n} + \frac{f}{n} \to \limsup_{n \to \infty} \frac{S_n}{n}.$$
Exactly the same reasoning tells us the lim inf is also invariant.

What we now need to show is that the set of points on which lim sup and lim inf do not agree have measure zero. We set $a < b$. We let

$$D = D(a, b) = \left\{x \in E : \liminf_{n\to\infty} \frac{S_n(x)}{n} < a < b < \limsup_{n\to\infty} \frac{S_n(x)}{n}\right\}.$$

Now if $\limsup \frac{S_n(x)}{n} \neq \liminf \frac{S_n(x)}{n}$, then there is some $a, b \in \mathbb{Q}$ such that $x \in D(a, b)$. So by countable subadditivity, it suffices to show that $\mu(D(a, b)) = 0$ for all a, b.

We now fix a, b, and just write D. Since $\limsup \frac{S_n}{n}$ and $\liminf \frac{S_n}{n}$ are both invariant, we have that D is invariant. By restricting to D, we can assume that $D = E$.

Suppose that $B \in \mathcal{E}$ and $\mu(G) < \infty$. We let

$$g = f - b\mathbf{1}_B.$$

Then g is integrable because f is integrable and $\mu(B) < \infty$. Moreover, we have

$$S_n(g) = S_n(f - b\mathbf{1}_B) \geq S_n(f) - nb.$$

Since we know that $\limsup_n \frac{S_n(f)}{n} > b$ by definition, we can find an n such that $S_n(g) > 0$. So we know that

$$S^*(g)(x) = \sup_n S_n(g)(x) > 0$$

for all $x \in D$. By the maximal ergodic lemma, we know

$$0 \leq \int_D g \, \mathrm{d}\mu = \int_D f - b\mathbf{1}_B \, \mathrm{d}\mu = \int_D f \, \mathrm{d}\mu - b\mu(B).$$

If we rearrange this, we know

$$b\mu(B) \leq \int_D f \, \mathrm{d}\mu.$$

for all measurable sets $B \in \mathcal{E}$ with finite measure. Since our space is σ-finite, we can find $B_n \nearrow D$ such $\mu(B_n) < \infty$ for all n. So taking the limit above tells

$$b\mu(D) \leq \int_D f \, \mathrm{d}\mu. \tag{†}$$

Now we can apply the same argument with $(-a)$ in place of b and $(-f)$ in place of f to get

$$(-a)\mu(D) \leq -\int_D f \, \mathrm{d}\mu. \tag{‡}$$

Now note that since $b > a$, we know that at least one of $b > 0$ and $a < 0$ has to be true. In the first case, (†) tells us that $\mu(D)$ is finite, since f is integrable. Then combining with (‡), we see that

$$b\mu(D) \leq \int_D f \, \mathrm{d}\mu \leq a\mu(D).$$

But $a < b$. So we must have $\mu(D) = 0$. The second case follows similarly (or follows immediately by flipping the sign of f).

We are almost done. We can now define

$$\bar{f}(x) = \begin{cases} \lim S_n(f)/n & \text{the limit exists} \\ 0 & \text{otherwise} \end{cases}$$

Then by the above calculations, we have

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

Also, we know $\bar{f}$ is invariant, because $\lim S_n(f)/n$ is invariant, and so is the set where the limit exists.

Finally, we need to show that

$$\mu(\bar{f}) \leq \mu(|f|).$$

This is since

$$\mu(|f \circ \Theta^n|) = \mu(|f|)$$

as Θ^n preserves the metric. So we have that

$$\mu(|S_n|) \leq n\mu(|f|) < \infty.$$

So by Fatou's lemma, we have

$$\begin{aligned} \mu(|\bar{f}|) &\leq \mu\left(\liminf_n \left|\frac{S_n}{n}\right|\right) \\ &\leq \liminf_n \mu\left(\frac{S_n}{n}\right) \\ &\leq \mu(|f|) \end{aligned}$$

□

Proof of von Neumann ergodic theorem. It is an exercise on the example sheet to show that

$$\|f \circ \Theta\|_p^p = \int |f \circ \Theta|^p \,\mathrm{d}\mu = \int |f|^p \,\mathrm{d}\mu = \|f\|_p^p.$$

So we have

$$\left\|\frac{S_n}{n}\right\|_p = \frac{1}{n}\|f + f \circ \Theta + \cdots + f \circ \Theta^{n-1}\| \leq \|f\|_p$$

by Minkowski's inequality.

So let $\varepsilon > 0$, and take $M \in (0, \infty)$ so that if

$$g = (f \vee (-M)) \wedge M,$$

then

$$\|f - g\|_p < \frac{\varepsilon}{3}.$$

By Birkhoff's theorem, we know

$$\frac{S_n(g)}{n} \to \bar{g}$$

a.e.

Also, we know

$$\left|\frac{S_n(g)}{n}\right| \leq M$$

for all n. So by bounded convergence theorem, we know

$$\left\|\frac{S_n(g)}{n} - \bar{g}\right\|_p \to 0$$

as $n \to \infty$. So we can find N such that $n \geq N$ implies

$$\left\|\frac{S_n(g)}{n} - \bar{g}\right\|_p < \frac{\varepsilon}{3}.$$

Then we have

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

So if $n \geq N$, then we know

$$\left\|\frac{S_n(f)}{n} - \bar{f}\right\|_p \leq \left\|\frac{S_n(f-g)}{n}\right\|_p + \left\|\frac{S_n(g)}{n} - \bar{f}\right\|_p + \|\bar{g} - \bar{f}\|_p \leq \varepsilon.$$

So done. □

7 Big theorems

7.1 The strong law of large numbers

Theorem (Strong law of large numbers assuming finite fourth moments)**.** Let (X_n) be a sequence of independent random variables such that there exists $\mu \in \mathbb{R}$ and $M > 0$ such that
$$\mathbb{E}[X_n] = \mu, \quad \mathbb{E}[X_n^4] \leq M$$
for all n. With $S_n = X_1 + \cdots + X_n$, we have that
$$\frac{S_n}{n} \to \mu \text{ a.s. as } n \to \infty.$$

Proof. We reduce to the case that $\mu = 0$ by setting
$$Y_n = X_n - \mu.$$
We then have
$$\mathbb{E}[Y_n] = 0, \quad \mathbb{E}[Y_n^4] \leq 2^4(\mathbb{E}[\mu^4 + X_n^4]) \leq 2^4(\mu^4 + M).$$
So it suffices to show that the theorem holds with Y_n in place of X_n. So we can assume that $\mu = 0$.

By independence, we know that for $i \neq j$, we have
$$\mathbb{E}[X_i X_j^3] = \mathbb{E}[X_i]\mathbb{E}[X_j^3] = 0.$$
Similarly, for all i, j, k, ℓ distinct, we have
$$\mathbb{E}[X_i X_j X_k^2] = \mathbb{E}[X_i X_j X_k X_\ell] = 0.$$
Hence we know that
$$\mathbb{E}[S_n^4] = \mathbb{E}\left[\sum_{k=1}^n X_k^4 + 6 \sum_{1 \leq i < j \leq n} X_i^2 X_j^2\right].$$
We know the first term is bounded by nM, and we also know that for $i \neq j$, we have
$$\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] \leq \sqrt{\mathbb{E}[X_i^4]\mathbb{E}[X_j^4]} \leq M$$
by Jensen's inequality. So we know
$$\mathbb{E}\left[6 \sum_{1 \leq i < j \leq n} X_i^2 X_j^2\right] \leq 3n(n-1)M.$$
Putting everything together, we have
$$\mathbb{E}[S_n^4] \leq nM + 3n(n-1)M \leq 3n^2 M.$$
So we know
$$\mathbb{E}\left[(S_n/n)^4\right] \leq \frac{3M}{n^2}.$$

So we know
$$\mathbb{E}\left[\sum_{n=1}^\infty \left(\frac{S_n}{n}\right)^4\right] \leq \sum_{n=1}^\infty \frac{3M}{n^2} < \infty.$$
So we know that
$$\sum_{n=1}^\infty \left(\frac{S_n}{n}\right)^4 < \infty \text{ a.s.}$$
So we know that $(S_n/n)^4 \to 0$ a.s., i.e. $S_n/n \to 0$ a.s. □

Theorem (Strong law of large numbers)**.** Let (Y_n) be an iid sequence of integrable random variables with mean ν. With $S_n = Y_1 + \cdots + Y_n$, we have
$$\frac{S_n}{n} \to \nu \text{ a.s.}$$

Proof. Let m be the law of Y_1 and let $\mathbf{Y} = (Y_1, Y_2, Y_3, \cdots)$. We can view Y as a function
$$Y : \Omega \to \mathbb{R}^\mathbb{N} = E.$$
We let $(E, \mathcal{E}, \mu)$ be the canonical space associated with the distribution m so that
$$\mu = \mathbb{P} \circ Y^{-1}.$$
We let $f : E \to \mathbb{R}$ be given by
$$f(x_1, x_2, \cdots) = X_1(x_1, \cdots, x_n) = x_1.$$
Then X_1 has law given by m, and in particular is integrable. Also, the shift map $\Theta : E \to E$ given by
$$\Theta(x_1, x_2, \cdots) = (x_2, x_3, \cdots)$$
is measure-preserving and ergodic. Thus, with
$$S_n(f) = f + f \circ \Theta + \cdots + f \circ \Theta^{n-1} = X_1 + \cdots + X_n,$$
we have that
$$\frac{S_n(f)}{n} \to \bar{f} \text{ a.e.}$$
by Birkhoff's ergodic theorem. We also have convergence in L^1 by von Neumann ergodic theorem.

Here $\bar{f}$ is $\mathcal{E}_\Theta$-measurable, and Θ is ergodic, so we know that $\bar{f} = c$ a.e. for some constant c. Moreover, we have
$$c = \mu(\bar{f}) = \lim_{n\to\infty} \mu(S_n(f)/n) = \nu.$$

So done. □

7.2 Central limit theorem

Theorem. Let (X_n) be a sequence of iid random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_1^2] = 1$. Then if we set

$$S_n = X_1 + \cdots + X_n,$$

then for all $x \in \mathbb{R}$, we have

$$\mathbb{P}\left[\frac{S_n}{\sqrt{n}} \leq x\right] \to \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,\mathrm{d}y = \mathbb{P}[N(0,1) \leq x]$$

as $n \to \infty$.

Proof. Let $\phi(u) = \mathbb{E}[e^{iuX_1}]$. Since $\mathbb{E}[X_1^2] = 1 < \infty$, we can differentiate under the expectation twice to obtain

$$\phi(u) = \mathbb{E}[e^{iuX_1}], \quad \phi'(u) = \mathbb{E}[iX_1 e^{iuX_1}], \quad \phi''(u) = \mathbb{E}[-X_1^2 e^{iuX_1}].$$

Evaluating at 0, we have

$$\phi(0) = 1, \quad \phi'(0) = 0, \quad \phi''(0) = -1.$$

So if we Taylor expand ϕ at 0, we have

$$\phi(u) = 1 - \frac{u^2}{2} + o(u^2).$$

We consider the characteristic function of $S_n/\sqrt{n}$

$$\begin{aligned}
\phi_n(u) &= \mathbb{E}[e^{iuS_n/\sqrt{n}}]\\
&= \prod_{i=1}^{n} \mathbb{E}[e^{iuX_j/\sqrt{n}}]\\
&= \phi(u/\sqrt{n})^n\\
&= \left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right)^n.
\end{aligned}$$

We now take the logarithm to obtain

$$\begin{aligned}
\log \phi_n(u) &= n \log\left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right)\\
&= -\frac{u^2}{2} + o(1)\\
&\to -\frac{u^2}{2}
\end{aligned}$$

So we know that

$$\phi_n(u) \to e^{-u^2/2},$$

which is the characteristic function of a $N(0,1)$ random variable.

So we have convergence in characteristic function, hence weak convergence, hence convergence in distribution. □