1 Probability Measure and Integration Theory in a Nutshell

1.1 Measurable Space and Measurable Functions

Definition 1.1. A measurable space is a tuple $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ a σ-*algebra* on Ω, that is a collection of subset of Ω such that

(i) $\emptyset \in \mathcal{F}$;[1]

(ii) $A^c \in \mathcal{F}$ whenever $A \in \mathcal{F}$;

(iii) $\cup A_n \in \mathcal{F}$ for every sequence $(A_n) \subseteq \mathcal{F}$.

In probability theory

- Ω is a set modelling different *states* of the world about which there is an uncertainty concerning the realization of it. It is called the *state space*. For instance:
  - Coin flipping. Let $\Omega = \{H, T\}$ where H and T denotes the states "Head occurs" and "Tail occurs" as the outcome of throwing a coin.
  - Temperature tomorrow. Let $\Omega = \mathbb{R}$ where $x \in \mathbb{R}$ is the temperature at 8:00 am tomorrow.
  - Financial decision. Let $\Omega = [0, 100]^2$ where for $(x, y) \in \Omega$, x and y represents the interest rate in % that the central banks of USA and EU, respectively, will fix next month.
- $\mathcal{F}$ is the collection of *events*, an event being a collection of states that might happen. For instance for the previous examples
  - $A = \{H\}$ is the event that head will occur;
  - $A = [13, 19]$ is the event that the temperature tomorrow at 8:00am will lie within 13 degrees and 19 degrees;
  - $A = [0.25, 0.75] \times [0.9, 1.8] \cup \{1\} \times [1.7, 2.1]$ is the event that next month the USA will fix an interest rate between 0.25% and 0.75% and the EU an interest rate between 0.9% and 1.8% OR the USA will fix an interest rate of 1% and the EU an interest rate between 1.7% and 2.1%.

Remark 1.2. The following points follows from the definition of a σ-algebra:

- $X \in \mathcal{F}$. Indeed, $\emptyset \in \mathcal{F}$ by *(i)* therefore $X = \emptyset^c \in \mathcal{F}$ by condition *(ii)*.
- $\cap A_n \in \mathcal{F}$ for every sequence $(A_n) \subseteq \mathcal{F}$. Indeed, by *(iii)* it follows that $A_n^c \in \mathcal{F}$ for every n, that is, the sequence $(A_n^c) \subseteq \mathcal{F}$. Hence $\cup A_n^c \in \mathcal{F}$. Using *(ii)*, it follows that $(\cup A_n^c)^c \in \mathcal{F}$. However $(\cup A_n^c)^c = \cap (A_n^c)^c = \cap A_n$. ♦

Lemma 1.3. *Let $(\mathcal{F}_i)$ be an arbitrary non-empty collections of σ-algebras on Ω. It holds*

$$\mathcal{F} := \cap \mathcal{F}_i = \{A \subseteq \Omega \colon A \in \mathcal{F}_i \text{ for all } i\}$$

is a σ-algebra on Ω.
Given a collection $\mathcal{C}$ of subsets of Ω, there exists a smallest σ-algebra that contains $\mathcal{C}$ which is denoted by $\sigma(\mathcal{C})$.

Proof. Homework sheet. □

[1] Note that this assumption follows directly when $\mathcal{F}$ is supposed to be non-empty. Why?

Definition 1.4 (Dynkin system). Let $(\Omega, \mathcal{F})$ be a measurable space. A collection of subsets $\mathcal{C}$ of Ω is called a

(i) λ- or Dynkin-system if

- $\Omega \in \mathcal{C}$;
- $B^c \in \mathcal{C}$ whenever $B \in \mathcal{C}$;
- $\cup A_n \in \mathcal{F}$ for every sequence of pairwise disjoints events $(A_n) \subseteq \mathcal{C}$.

(ii) π-system if

- $A \cap B \in \mathcal{F}$ whenever $A, B \in \mathcal{F}$.

Theorem 1.5. *Let Ω be a state space, $\mathcal{P}$ be a π-system, and $\mathcal{C}$ be a λ-system that contains $\mathcal{P}$. Then $\sigma(\mathcal{P}) \subseteq \mathcal{C}$.*

Proof. We first show that if $\mathcal{C}$ is closed under finite intersection, then it is a σ-algebra. Let $(A_n) \subseteq \mathcal{C}$ and define $B_n = A_n \setminus (\cup_{k<n} A_k)$, $n > 1$ and $B_1 = A_1$. As for we supposed that $\mathcal{C}$ is closed under finite intersection, it follows that $(B_n) \subseteq \mathcal{C}$. From $\cup B_n = \cup A_n$ and (B_n) pairwise disjoint, it follows from the λ-system assumption on $\mathcal{C}$ that $\cup A_n \in \mathcal{C}$.
Now, it clearly holds $\lambda(\mathcal{P}) \subseteq \lambda(\mathcal{C}) = \mathcal{C}$ where $\lambda(\cdot)$ denotes the smallest λ-system generated by $(\cdot)$. From what previously hold, we just have to show that $\lambda(\mathcal{P})$ is closed under finite intersection, as so $\lambda(\mathcal{P})$ would then be a σ-algebra containing $\mathcal{P}$ and so $\sigma(\mathcal{P}) \subseteq \lambda(\mathcal{P}) \subseteq \mathcal{C}$. Let then $D \in \lambda(\mathcal{P})$ and define $\mathcal{D}_D = \{A \subseteq \Omega \colon A \cap D \in \lambda(\mathcal{P})\}$. We show that $\mathcal{D}_D$ is a λ-system. Clearly $X \in \mathcal{D}_D$. If $A \in \mathcal{D}_D$, it follows that

$$A^c \cap D = (A^c \cup D^c) \cap D = (A \cap D)^c \cap D = ((A \cap D) \cup D^c)^c .$$

By assumption, $A \cap D \in \lambda(\mathcal{P})$. Since $\lambda(\mathcal{P})$ is stable under complementation and countable intersection of disjoints elements, it follows that $A^c \cap D \in \lambda(\mathcal{P})$ and therefore $A^c \in \mathcal{D}_D$. Let now (A_n) be a sequence in $\mathcal{D}_D$ of pairwise disjoints elements. From the stability of $\lambda(\mathcal{P})$ under pairwise disjoint elements and the fact that $(\cup A_n) \cap D = \cup(A_n \cap D)$, it follows that $\cup A_n \in \mathcal{D}_D$. Hence, $\mathcal{D}_D$ is a λ-system too.
Since $\mathcal{P}$ is stable under finite intersection it follows that $\mathcal{P} \subseteq \mathcal{D}_B$ for every $B \in \mathcal{P}$. Hence, $\lambda(\mathcal{P}) \subseteq \mathcal{D}_B$ for every $B \in \mathcal{P}$. In particular, for every $A \in \lambda(\mathcal{P})$ and $B \in \mathcal{P}$ it holds $A \cap B \in \lambda(\mathcal{P}) \subseteq \mathcal{D}_B$. Per definition, this also means that $B \in \mathcal{D}_A$ for every $B \in \mathcal{P}$ and $A \in \lambda(\mathcal{P})$ showing that $\mathcal{P} \sqsubseteq \mathcal{D}_A$ for every $A \in \lambda(\mathcal{P})$. Hence $\lambda(\mathcal{P}) \sqsubseteq \mathcal{D}_A$ for every $A \in \lambda(\mathcal{P})$. Thus, for $A, B \in \lambda(\mathcal{P})$ it holds $B \in \mathcal{D}_A$ which per definition means $A \cap B \in \lambda(\mathcal{P})$ showing that $\lambda(\mathcal{P})$ is closed under finite intersection and therefore, by the first step of the proof, a σ-algebra. $\square$

As we see, σ-algebra are structures of set, similar to another very important structure of set which is a topology.

Definition 1.6. A topological space is a tuple $(\Omega, \mathcal{T})$ where $\mathcal{T}$ is a collection of subsets of a set Ω such that

(i) $\emptyset, X$ are in $\mathcal{T}$;

(ii) $O_1 \cap O_2 \in \mathcal{T}$ whenever $O_1, O_2 \in \mathcal{T}$;

(iii) $\cup O_i \in \mathcal{T}$ for any arbitrary family (O_i) of elements in $\mathcal{T}$.

Elements of τ are called *open sets*. The complement of any open set is called a *closed set*.

A topological space is stable under arbitrary union, finite intersection but not complementation. As σ-algebras, topologies are stable under arbitrary intersections, and therefore we can define the smallest topology $\mathcal{T}(\mathcal{B})$ generated by a collection $\mathcal{B}$ of subsets of Ω. Just as dynkin systems, or semi-rings and ring as we will see later, some smaller structures often describe topologies, namely, topological bases.

Definition 1.7. A topological base on a set Ω is a collection $\mathcal{B}$ of subsets of Ω such that

(i) $\cup\{O\colon O \in \mathcal{B}\} = \Omega$;

(ii) for every $O_1, O_2 \in \mathcal{B}$ such that $O_1 \cap O_2 \neq \emptyset$, there exists an non-empty $O_3 \in \mathcal{B}$ such that $O_3 \subseteq O_1 \cap O_2$.

It is easy to check that the open sets of the topology $\mathcal{T}(\mathcal{B})$ generated by a topological base are unions – not necessarily countable – of elements of $\mathcal{B}$.

Definition 1.8. Let $(\Omega, \mathcal{T})$ be a topological space. The σ-algebra generated by the open sets of Ω is called the Borel σ-algebra, and denoted by $\mathcal{B}(\mathcal{T})$.

Example 1.9. In $\mathbb{R}^d$, the balls $B_r(q) = \{x \in \mathbb{R}^2\colon \|x - q\| < r\}$ where $r \in \mathbb{Q}_{++} = \{r \in \mathbb{Q}\colon r > 0\}$ and $q \in \mathbb{Q}^d$ constitute a topological basis.[2] The resulting topology is the usual euclidean topology on $\mathbb{R}^d$. A particularity of this topology is that the topological base is countable.[3] It follows that any open set in $\mathbb{R}^d$ can be written as a countable union of open balls.
In general if you have a metric space (Ω, d) which is *separable*, that is, there exists a countable dense subset $(x_n) \subseteq \Omega$, then the countable collection of open balls $B_r(x_n) = \{y \in \Omega\colon d(x_n, y) < r\}$ for $r \in \mathbb{Q}_{++}$ and n is a countable topological base of Ω.
Such spaces play a central role in probability theory since the Borel σ-algebra coincide with the σ-algebra generated by this countable family of balls. $\Diamond$

Exercice 1.10. Let $\Omega = \mathbb{R}$, and $\mathcal{F} = \mathcal{B}(\mathbb{R})$ the Borel σ-algebra of $\mathbb{R}$, that is the σ-algebra generated by the collection $\mathcal{C} = \{O\colon O \text{ open set in } \mathbb{R}\}$. Show that $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{A})$ whenever

$\mathcal{A} = \{F\colon F \text{ closed subset of } \mathbb{R}\}$	
$\mathcal{A} = \{]a,b[\colon a \leq b \text{ with } a,b \in \mathbb{R}\}$	$\mathcal{A} = \{[a,b]\colon a \leq b \text{ with } a,b \in \mathbb{R}\}$
$\mathcal{A} = \{]a,b]\colon a \leq b \text{ with } a,b \in \mathbb{R}\}$	$\mathcal{A} = \{[a,b[\colon a \leq b \text{ with } a,b \in \mathbb{R}\}$
$\mathcal{A} = \{]-\infty,b]\colon b \in \mathbb{R}\}$	$\mathcal{A} = \{]-\infty,b[\colon b \in \mathbb{R}\}$
$\mathcal{A} = \{[a,\infty[\colon a \in \mathbb{R}\}$	$\mathcal{A} = \{]a,\infty[\colon a \in \mathbb{R}\}$
$\mathcal{A} = \{]a,b[\colon a \leq b \text{ with } a,b \in \mathbb{Q}\}$	$\mathcal{A} = \{[a,b]\colon a \leq b \text{ with } a,b \in \mathbb{Q}\}$
$\mathcal{A} = \{]a,b]\colon a \leq b \text{ with } a,b \in \mathbb{Q}\}$	$\mathcal{A} = \{[a,b[\colon a \leq b \text{ with } a,b \in \mathbb{Q}\}$
$\mathcal{A} = \{]-\infty,b]\colon b \in \mathbb{Q}\}$	$\mathcal{A} = \{]-\infty,b[\colon b \in \mathbb{Q}\}$
$\mathcal{A} = \{[a,\infty[\colon a \in \mathbb{Q}\}$	$\mathcal{A} = \{]a,\infty[\colon a \in \mathbb{Q}\}$

$\Diamond$

Exercice 1.11. Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable spaces and $X : \Omega \to S$ a function. Show that

- $\{X^{-1}(B)\colon B \in \mathcal{S}\}$ is a σ-algebra, that is denoted by $\sigma(X)$.
- give a counter example that $\{X(A)\colon A \in \mathcal{F}\}$ is not a σ-algebra.

Hint: think about the properties of direct images and pre-images with respect to operations on sets. $\Diamond$

Definition 1.12. Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable spaces. A measurable function[4] is a function $X : \Omega \to S$ such that

$$X^{-1}(B) = \{\omega\colon X(\omega) \in B\} \in \mathcal{F}, \quad \text{for every } B \in \mathcal{S}.$$

If $S = \mathbb{R}$ and $\mathcal{S} = \mathcal{B}$, we call X a *random variable*.

[2] Check this using the triangular inequality and the density of $\mathbb{Q}^d$ in $\mathbb{R}^d$.
[3] Such topologies generated by a countable base are called second countable topologies.
[4] If necessary, we say $\mathcal{F}$-$\mathcal{S}$-measurable if the context is not clear with respect to which we are measurable.

Remark 1.13. In probability theory, we often abuse notations whenever it is clear what is the image space, that is we use the shorthand notations for random variables

- $\{X \in B\} := \{X^{-1}(B)\}$.
- $\{X = x\} := \{X^{-1}(\{x\})\}$;
- $\{X \leq x\} := \{X^{-1}(]-\infty, x])\}$ ♦

Remark 1.14. If Ω is a state space without a predefined σ-algebra, and $X : \Omega \to S$ is a function, then

$$\sigma(X) := \sigma\left(\left\{X^{-1}(B) \colon B \in \mathcal{S}\right\}\right)$$

is the smallest σ-algebra for which X is a measurable function. In other words, in the framework of the Definition 1.12 of measurable function, it holds $\sigma(X) \subseteq \mathcal{F}$.
More generally, let (X_i) be a family of functions $X_i : \Omega \to S$, then $\sigma(X_i \colon i) = \sigma(\{X_i^{-1}(B) \colon B \in \mathcal{S}, i\})$ is the smallest σ-algebra such that each X_i is measurable. ♦

Lemma 1.15. *The composition of measurable functions is measurable.*

Proof. Left as an exercise. □

Proposition 1.16. *Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable spaces. If $\mathcal{C}$ is a collection of subsets of S such that $\mathcal{S} = \sigma(\mathcal{C})$, then for $X : \Omega \to S$, the following assertions are equivalent*

(i) *X is measurable;*

(ii) *$\{X \in B\} \in \mathcal{F}$ for every $B \in \mathcal{C}$.*

Proof. Clearly, *(i)* implies *(ii)*. Reciprocally, let $\mathcal{D} := \{B \in \mathcal{S} \colon X^{-1}(B) \in \mathcal{F}\}$. By assumption, $\mathcal{C} \subseteq \mathcal{D}$, henceforth by assumption, $\mathcal{S} = \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{D}) \subseteq \mathcal{S}$. It follows that X is measurable. □

Combined with Exercise 1.10, it follows that for every function $X : \Omega \to \mathbb{R}$ to be a random variable, it suffices to check that $\{X \leq x\} \in \mathcal{F}$ for every $x \in \mathbb{R}$.
The concept of measurability is the measurable pendant to continuity for functions between topological spaces.

Definition 1.17. Let $(\Omega, \mathcal{T}), (\Omega', \mathcal{T}')$ be two topological spaces. A function $X : \Omega \to \Omega'$ is called *continuous* if $X^{-1}(O)$ is open for every open set O' in Ω'.
In the case where $\Omega' = \mathbb{R}$ or $\Omega' = [-\infty, \infty]$, we say that a function is

- lower semi-continuous if $\{X \leq t\}$ is closed for every $t \in \mathbb{R}$.
- upper semi-continuous if $\{X \geq t\}$ is closed for every $t \in \mathbb{R}$.

If Ω is a metric space, the following are equivalent

- X is continuous, lower semi-continuous or upper semi-continuous, respectively
- $\lim X(\omega_n) \to X(\omega)$, $\liminf X(\omega_n) \geq X(\omega)$, or $\limsup X(\omega_n) \leq X(\omega)$ for everu $\omega_n \to \omega$, respectively.

This has to do with the fact that a set $F \subseteq \Omega$ in a metric space is closed if and only if $\omega_n \to \omega$ implies $\omega \in F$ whenever $(\omega_n) \subseteq F$.

Remark 1.18. As for measurable functions, you can define topologies generated by family of functions, analogue to Remark 1.14 as the smallest topology that makes functions continuous. Also, the composition of continuous functions is continuous. ♦

Corollary 1.19. *Let $X : \Omega \to \mathbb{R}$ be a function where Ω is a topological space endowed with the Borel σ-algebra. Under the following assumptions, X is a random variable*

- *X is a continuous function;*
- *X is an upper semi-continuous function;*[5]
- *X is a lower semi-continuous function;*[6]

Proof. As for the continuity, we make use of the fact that the Borel σ-algebra on the real line is generated by the closed sets $]-\infty, t]$ for $t \in \mathbb{R}$. From the definition of continuity, $\{X \leq t\}$ is closed and therefore measurable for every $t \in \mathbb{R}$. It follows by Proposition 1.16 that X is measurable. The same argumentation holds for lower semi-continuous functions. For the upper semi-continuous one, we use the intervals $[t, \infty[$ for $t \in \mathbb{R}$. □

Definition 1.20. Let $(\Omega_i, \mathcal{F}_i)$ be a non-empty family of measurable spaces. The *product* σ-algebra, denoted by $\otimes \mathcal{F}_i$ on the product state space $\Omega = \prod \Omega_i$, is defined as the σ-algebra generated by the family of projections

$$\begin{aligned} \pi_i : \Omega = \prod \Omega_i &\longrightarrow \Omega_i \\ \omega = (\omega_i) &\longmapsto \omega_i \end{aligned}$$

As an exercise, show that in 2 dimensions, it holds $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A_1 \times A_2 \colon A_i \in \mathcal{F}_i, i = 1, 2\})$.
Under the notations of Definition 1.20, a *product cylinder set* is a set $A \subseteq \Omega$ of the form – assuming that the index set is directed[7] –

$$A = \prod_{i < i_1} \Omega_i \times A_{i_1} \times \prod_{i_1 < i < i_2} \Omega_i \times A_{i_2} \ldots \times \prod_{i_{n-1} < i < i_n} \Omega_i \times A_{i_n} \times \prod_{i_n < i} \Omega_i$$

where $A_{i_k} \in \mathcal{F}_{i_k}$ for $k = 1, \ldots, n$.
As an exercise, show that the family of product cylinder generates the product σ-algebra.

Example 1.21. Consider now our example of coin tossing. Suppose that we are not only observing one coin toss but infinitely – countably – many such as for instance every minutes. Setting -1 for a tail and 1 for a head, we can formalize our state space as follows:

$$\Omega = \prod_n \{-1, 1\} = \{-1, 1\}^{\mathbb{N}} = \{\omega = (\omega_n) \colon \omega_n = \pm 1 \text{ for every } n\}$$

This state space can also be seen as the set of binary sequences for instance in computer science. On each $\Omega_n = \{-1, 1\}$ we consider the σ-algebra $\mathcal{F}_n = \{\emptyset, \{-1\}, \{1\}, \{-1, 1\}\}$. We endow this state space of the never ending realization of a coin toss with the product σ-algebra, that is, according to what has been stated previously, generated by the product cylinders that in this special case take the form:

$$C = \{\omega \text{ binary sequences such that } \omega_{n_k} = b_k, k = 1, \ldots n\}$$

for a given set of values $b_k \in \{-1, 1\}$, $k = 1, \ldots, n$. ◊

[5] That is $\{\omega \colon X(\omega) \geq x\}$ is closed for every $x \in \mathbb{R}$.
[6] That is $\{\omega \colon X(\omega) \leq x\}$ is closed for every $x \in \mathbb{R}$.
[7] Which is always possible from the general theory of boolean algebra.

We now focus uniquely on random variables, since for the following theorem we will use a lot the structure of $\mathbb{R}$, in particular its complete order that generates the topology.[8] From now on, we are given a measurable space $(\Omega, \mathcal{F})$ and denote by $\mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{F})$ the set of random variables on $(\Omega, \mathcal{F})$.

Proposition 1.22. *Let X, Y be random variables as well as (X_n) be a sequence of random variables. It holds*

- $aX + bY$ *is a random variable for every* $a, b \in \mathbb{R}$*;*
- XY *is a random variable;*
- $\max(X, Y)$ *and* $\min(X, Y)$ *are random variables;*
- $\sup X_n$ *and* $\inf X_n$ *are extended real valued random variables;*[9]
- $\liminf X_n := \inf_n \sup_{k \geq n} X_k$ *and* $\limsup X_n := \inf_n \sup_{k \geq n} X_k$ *are extended real valued random variables;*
- $A := \{\lim X_n \textit{ exists}\} := \{\omega\colon \lim X_n(\omega) \textit{ exists}\} = \{\liminf X_n = \limsup X_n\}$ *is measurable.*

Proof. See homework sheet. □

1.2 Probability Measures

Definition 1.23. A probability measure P on the measurable space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, \infty]$ such that

- $P[\emptyset] = 0$ and $P[\Omega] = 1$;
- $P[\cup A_n] = \sum P[A_n]$ for every sequence of pairwise disjoint[10] events $(A_n) \subseteq \mathcal{F}$.

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

In probability theory, a probability measure returns a quantification of the uncertainty that an event occurs. For instance

- $P[\{H\}] = 1/2$ is the probability that head occurs in the case of a "fair" coin.
- $P[[13, 19]] = \int_{13}^{19} e^{(x-17)^2/2} dx/(2\sqrt{\pi})$ if your model of the temperature distribution is normally distributed around 17 degrees with a standard deviation of 1.

Lemma 1.24. *Let P be a measure on a measurable space $(\Omega, \mathcal{F})$. For every $A, B \in \mathcal{F}$ and sequence $(A_n) \subseteq \mathcal{F}$, it holds*

- $P[B] = P[A] + P[B \setminus A] \geq P[A]$ *whenever* $A \subseteq B$*;*
- $P[A^c] = 1 - P[A]$*;*
- $P[A \cup B] + P[A \cap B] = P[A] + P[B]$*;*
- $P[\cup A_n] \leq \sum P[A_n]$
- $\lim_n P[\cup_{k \leq n} A_k] = P[\cup A_n]$

[8] Think why for each of the following assertions, the structure of $\mathbb{R}$ is so important.

[9] With respect to the Borel σ-algebra on $[-\infty.\infty]$ generated by the metric $d(x, y) = |\arctan(x) - \arctan(y)|$ that coincide with the euclidean topology on $\mathbb{R}$.

[10] That is $A_n \cap A_m = \emptyset$ for every $m \neq n$.

- $\lim_n P[\cap_{k\leq n} A_k] = P[\cup A_n]$

Proof. The last three ones follows from Lemma 1.30. Clearly, B is the disjoint union of A and $B \setminus A$. Using σ-additivity, the first assertion follows. The second one follows with $B = \Omega$ and $P[\Omega] = 1$. The third one follows from $A \cup B$ being the disjoint union of A and $B \setminus (A \cap B)$ and $P[B \setminus (A \cap B)] = P[B] - P[A \cap B]$. □

Note that a probability measure only take value in $[0, 1]$ due the monotony property and the $P[\Omega] = 1$. If we drop the assumption that $P[\Omega] = 1$, then P is a measure – usually denoted with the Greek letters μ, ν, ...

- If given a measure μ we assume that $\mu(\Omega) < \infty$ then we say that μ is a *finite measure*. However this is almost like a probability measure since if μ is non zero, defining $P = \mu/\mu(\Omega)$ gives a probability measure.
- If given a measure μ we assume that there exists an increasing sequence of measurable sets $A_1 \subseteq A_2 \subseteq \dots$ with $\lim A_n := \cup A_n$ such that $\mu(A_n) < \infty$, then we say that μ is a *σ-finite measure*. This is for instance the case of the Lebesgue measure λ on $\mathbb{R}^d$.
- If for every $A \in \mathcal{F}$ with $P[A] > 0$, there exists $B \subseteq A$ with $0 < P[B] < P[A]$, we say that P is an *atom free probability measure*.
- A set[11] $N \subseteq \Omega$ is called a *zero-set*, a *set of null measure*, a *negligible set* if there exists $A \in \mathcal{F}$ such that $P[A] = 0$ and $N \subseteq A$.
- The σ-algebra $\mathcal{F}^P = \sigma(\mathcal{F}, \mathcal{N})$ where $\mathcal{N}$ denotes the collection of all negligible sets is called the *completion* under P of $\mathcal{F}$.[12]
- A probability measure Q on $\mathcal{F}$ is called *absolutely continuous* with respect to P, denoted by $Q \ll P$, if $P[A] = 0$ implies $Q[A] = 0$ for every $A \in \mathcal{F}$. We say that Q is *equivalent* to P if $Q \ll P$ and $P \ll Q$.

Remark 1.25. In probability theory, we often adopt the following short handwritings

- $P[X \in B] := P[X^{-1}(B)]$.
- $P[X = x] := P[X^{-1}(\{x\})]$;
- $P[X \leq x] := P[X^{-1}(]-\infty, x])]$ ♦

Example 1.26 (Examples of Probability Measures). Let $(\Omega, \mathcal{F})$ be a measurable space.

1) **Probablity on countable sets**. Suppose that Ω is a countable set – a fortiori finite. Then each probability measure P on $\mathcal{F} = \mathcal{P}(\Omega) = 2^\Omega$ is of the form

$$P[A] = \sum_{\omega \in A} p(\omega)$$

for some function $p : \Omega \to [0, 1]$ with $\sum p(\omega) = 1$.[13] An important example of which is when $\Omega = \{1, \dots, N\}$ for $N \in \mathbb{N}$ if we take $p(n) = 1/N$ where $n \in \{1, \dots, N\}$ which is called the *uniform probability distribution* on Ω.

[11] Not necessarily measurable

[12] Be careful that the completed σ-algebra depends on P.

[13] Why?

2) **Dirac measure**. Suppose that $\{\omega_0\} \in \mathcal{F}$ for some $\omega_0 \in \Omega$. The Dirac measure at ω_0 is defined as the set value function

$$\delta_{\omega_0}(A) = \begin{cases} 1 & \text{if } \omega_0 \in A \\ 0 & \text{otherwise} \end{cases}$$

Other names for the Dirac measure are, point measure at ω_0.

3) **Counting measure**. Define

$$\mu(A) = \begin{cases} \#A & \text{if } A \text{ is finite} \\ \infty & \text{otherwise} \end{cases}.$$

It is easy to check that μ is an additive measure which is σ-stable if and only if A is finite. It is a probability measure if $\#\Omega = 1$.

4) **Normal Distribution**. For $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}$ we define

$$P[A] = \frac{1}{\sigma\sqrt{2\pi}} \int_A e^{-\frac{(x-\mu)^2}{2\sigma^2}} \lambda(dx)$$

where λ is the Lebesgue measure on $\mathbb{R}$. This is the seminal normal distribution, and it is over classical that this defines a probability measure. For instance in our example of the temperature for tomorrow morning we can assume that at this time of the year in Shanghai, these are normally distributed around 24 with a variance of 1. $\diamondsuit$

In Example 1.21, we introduced the state space of tossing infinitely a coin. Supposing that the coin is fair, we know that tossing and getting head is $1/2$. We could extend with combinatoric arguments what is the probability of a finite sequence of coin tosses, for instance of having tail head and then head in three tosses. The main question is whether it is possible to find a probability measure that is defined for any sequence of coin tossing but coincide for any finite sequence to what we intuitively understand for finitely many coin toss. The answer is in the so called Caratheordory measure extension that we won't prove here, but can be found in any measure text book.

Definition 1.27. A collection $\mathcal{R}$ of subsets of Ω is called a

- *semi-ring* if
  - *(i)* $\emptyset \in \mathcal{R}$
  - *(ii)* $A \cap B \in \mathcal{R}$ whenever $A, B \in \mathcal{R}$;
  - *(iii)* if $A, B \in \mathcal{R}$, there exists $C_1, \ldots, C_n \in \mathcal{R}$ pairwise disjoints such that $A \setminus B = \cup_{k \leq n} C_k$.
- *ring* if
  - *(i)* $\emptyset \in \mathcal{R}$
  - *(ii)* $A \cup B \in \mathcal{R}$ whenever $A, B \in \mathcal{R}$;
  - *(iii)* $A \setminus B \in \mathcal{R}$ whenever $A, B \in \mathcal{R}$;

From the identity $A \cap B = A \setminus (A \setminus B)$, it follows that a ring is closed under intersections and therefore a ring is a semi-ring.
The definition of a semi-ring might be quite artificial but it is actually useful together with Caratheory's extension theorem. Indeed, when you practically want to define a measure "per hand", it is often if not impossible to define it on such a complex collection as a σ-algebra and ensure that it has the good properties. Therefore, you often search for a simple collection of sets where the definition makes sense, and the following theorem ensures that you can find a measure that corresponds to the one you defined on the smallest subset.

Example 1.28. The collection of cylinders on $\Omega = \{-1, 1\}^{\mathbb{N}}$ is a semi-ring that generates the product σ-algebra. The collection $\{[a, b[\colon a < b, a, b \in \mathbb{R}\}$ that generates the Borel σ-algebra of the real line is a semi-ring but not a ring! $\Diamond$

Theorem 1.29 (Caratheordory Extension Theorem). *Let Ω be a non empty-set and $\mathcal{R}$ be a semi-ring. Let $P : \mathcal{R} \to [0, \infty]$ be a function such that*

- $P[\emptyset] = 0$*;*
- $P[\cup A_n] = \sum P[A_n]$ *if $(A_n) \subseteq \mathcal{R}$ is a sequence of pairwise disjoints elements such that $\cup A_n \in \mathcal{A}$.*

Such a function is called a content *and can be extended to a measure P on $\mathcal{F} = \sigma(A)$.*

You will construct in homework an important measure on the infinite sequence of coin tossing. The construction of Lebesgue measure follows the same strategy but is more difficult. A key lemma for this is the following one:

Lemma 1.30. *Let $\mathcal{R}$ be a ring and $P : \mathcal{C} \to [0, \infty]$ a finite content. Then the following are equivalent*

(i) σ-additivity: $P[\cup A_n] = \sum P[A_n]$ *for every pairwise disjoint sequence $(A_n) \subseteq \mathcal{R}$ such that $\cup A_n \in \mathcal{R}$;*

(ii) Lower semi-continuity: $\sup_n P[A_n] = P[\cup A_n]$ *for every increasing sequence $(A_n) \subseteq \mathcal{R}$ such that $\cup A_n \in \mathcal{R}$;*

(iii) Upper semi-continuity: $\inf_n P[A_n] = P[\cap A_n]$ *for every decreasing sequence $(A_n) \subseteq \mathcal{R}$ such that $\cap A_n \in \mathcal{R}$*

(iv) Continuous at $\emptyset$: $\inf_n P[A_n] = 0$ *for every sequence $(A_n) \subseteq \mathcal{R}$ such that $\cap A_n = \emptyset$*

Let $\mathcal{R}$ be a semi-ring and $P : \mathcal{R} \to [0, \infty]$ a function. Then, the following are equivalent

- *P is a measure;*
- *P is a σ-sub-additive content, that is for every sequence $(A_n) \subseteq \mathcal{R}$, such that $\cup A_n \in \mathcal{R}$ it holds $P[\cup A_n] \leq \sum P[A_n]$.*

Proof. Let us show that (i) implies (ii). Let (A_n) be an increasing sequence such that $A = \cup A_n \in \mathcal{R}$. Defining $B_n = A_n \setminus \cup_{k<n} A_k = A_n \setminus B_{n-1}$ for $n > 1$ and $B_1 = A_1$ provides a disjoint sequence of elements in $\mathcal{R}$. Indeed per induction starting with $B_1 = A_1 \in \mathcal{R}$, suppose that $B_{n-1} \in \mathcal{R}$ is follows by the definition of a ring – recall a ring is closed under union intersection and difference – and $A_n \in \mathcal{R}$ that $B_n = A_n \setminus B_{n-1} \in \mathcal{R}$. Since $A_n = \cup_{k \leq n} B_k$ and $A = \cup B_n$, it follows from σ-additivity that

$$P[A] = \sum P[B_n] = \sup \sum_{k \leq n} P[B_k] = \sup P[\cup_{k \leq n} B_k] = \sup P[A_n]$$

To show that (ii) implies (i) is analogue. Let $(A_n) \subseteq \mathcal{R}$ be a pairwise disjoint sequence of sets such that $A = \cup A_n \in \mathcal{R}$. Defining $B_n = \cup_{k \leq n} A_k$ provides an increasing sequence of element in $\mathcal{R}$ and $\cup B_n = A \in \mathcal{R}$. Hence

$$P[A] = \sup P[B_n] = \sup \sum_{k \leq n} P[A_k] = \sum P[A_k]$$

Let us show that (ii) implies (iii). Let $(A_n) \subseteq \mathcal{R}$ be a decreasing sequence such that $A = \cap A_n \in \mathcal{R}$. It follows that $B_n = A_1 \setminus A_n$ define an increasing sequence such that $B = \cup B_n = A_1 \setminus \cap A_n = A_1 \setminus A \in \mathcal{R}$. Lower semi-continuity and additivity implies[14]

$$P[A_1] - \inf P[A_n] = \sup(P[A_1] - P[A_n]) = \sup P[A_1 \setminus A_n] = P[\cup A_1 \setminus A_n] = P[A_1 \setminus A] = P[A_1] - P[A]$$

Let us show that (iii) implies (ii). Let $(A_n) \subseteq \mathcal{R}$ be an increasing sequence such that $A = \cup A_n \in \mathcal{R}$, then $B_n = A \setminus A_n$ is a decreasing sequence in $\mathcal{R}$ such that $\cap B_n = A \setminus \cup A_n \in \mathcal{R}$. The same argumentation as above yields the assertion.
The fact that (iii) implies (iv) is immediate, so let us show that (iv) implies (iii). It is left as an exercise by noting that a decreasing family $(A_n) \subseteq \mathcal{R}$ such that $A = \cap_n A_n \in \mathcal{R}$ defines a decreasing family $B_n = A_n \setminus A$ of elements in $\mathcal{R}$ which intersection is the empty-set.
Let us show the second part of the Lemma. If P is a measure, it follows that is is a monotone and a sub-additive content. From the previous equivalence it is also lower semi-continuous. Let now (A_n) be a sequence in $\mathcal{R}$ such that $\cup A_n \in \mathcal{R}$. Define $B_n = \cup_{k \leq n} A_k$ which is an increasing sequence in $\mathcal{R}$ such that $\cup B_n = \cup A_n \in \mathcal{R}$. It follows that

$$P[\cup A_n] = P[\cup B_n] = \sup P[B_n] = \sup P[\cup_{k \leq n} A_k] \leq \sup \sum_{k \leq n} P[A_k] = \sum P[A_n]$$

showing the σ-sub-additivity.
Reciprocally, let P be a σ-subadditive content on $\mathcal{R}$. An easy exercise is to show that it extends to a content $\bar{P}$ on the ring generated by $\mathcal{R}$, which is monotone. Let (A_n) be a disjoint family of events in $\mathcal{R}$ such that $\cup A_n \in \mathcal{R}$. It follows that

$$\sum_{k \leq n} P[A_k] = \sum_{k \leq n} \bar{P}[A_k] = \bar{P}[\cup_{k \leq n} A_k] \leq \bar{P}[\cup A_n] = \bar{P}[A] = P[A]$$

hence

$$\sum P[A_n] \leq P[\cup_n A_n]$$

reciprocally σ-subadditivity yields the reverse equality, showing σ-additivity. $\square$

Example 1.31. Let $(\Omega, \mathcal{F})$ be the measurable space of infinite coin tossing, that is, $\Omega = \{-1, 1\}^{\mathbb{N}}$ and $\mathcal{F}$ the product σ-algebra on it generated by the finite dimensional cylinder. We suppose that the coin is fair, that is, the probability of getting 1 staying for head is $1/2$. Hence for a cylinder of the form

$$\begin{aligned} C &= \{\omega \text{ binary sequences such that } \omega_{n_k} = b_k, k = 1, \ldots n\} \\ &= \{-1,1\} \times \ldots \times \{b_1\} \times \{-1,1\} \ldots \{-1,1\} \times \{b_n\} \times \{-1,1\} \times \ldots \end{aligned}$$

a probability P on $\mathcal{F}$ should have the property

$$P[C] = 2^{-n}$$

where n is the number of times where we observe a coin toss. To extend this measure, we show in exercise that the collection of cylinders is a semi-ring (why?). We then show in exercise that we just have to check that P is a σ-sub-additivit measure according to the previous lemma. To do so, some topological arguments are needed. $\Diamond$

The main question though is, as in the previous example, if such a measure is unique. This follows however from Dynkin Theorem 1.5.

[14] Show that for a content on a ring, it holds $P[A \setminus B] = P[A] = P[B]$ whenever $A, B \in \mathcal{R}$.

Proposition 1.32. *Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{P}$ a π-system on Ω that generates $\mathcal{F}$. Suppose that two measures P and Q on $\mathcal{F}$ coincide on $\mathcal{P}$, then $P = Q$.*

Proof. Let $\mathcal{C}$ be the collection of measurable sets on which P and Q coincide. By assumption $\mathcal{P} \subseteq \mathcal{C}$. Further, it can be easily checked – do it!! – that $\mathcal{C}$ is a λ-system. Therefore, applying Theorem 1.5, it follows that $\mathcal{F} = \sigma(\mathcal{P}) \subseteq \mathcal{C} \subseteq \mathcal{F}$ showing that $P = Q$. □

Since any semi-ring is in particular a π-system, the uniqueness of P constructed in Example 1.21 follows.

1.3 Integration

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable $X \in \mathcal{L}^0$ is said to be *simple* or a step function, if

$$X = \sum_{k \leq n} \alpha_k 1_{A_k}$$

for a $A_1, \ldots, A_n \in \mathcal{F}$ disjoint and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Note that this representation is not unique (why?). We denote by $\mathcal{L}^{0,step}$ the collection of these step functions which is a linear subspace of $\mathcal{L}^0$ and define the *expectation* of X with respect to P as

$$E[X] := \sum_{k \leq n} \alpha_k P[A_k]$$

Exercice 1.33. Show that the definition of the expectation is a well defined operator on $\mathcal{L}^{0,step}$ (why?).◊

Proposition 1.34. *On $\mathcal{L}^{0,step}$, the following properties hold*

- Montonie: $E[X] \leq E[Y]$ *whenever* $X \leq Y$.
- Linearity: E *is a linear operator on* $\mathcal{L}^{0,step}$;

Proof. Left as an exercise. □

Definition 1.35. For $X \in \mathcal{L}^0_+ := \{X \in \mathcal{L}^0 \colon X \geq 0\}$ we define

$$E[X] := \sup \left\{ E[Y] \colon Y \leq X, Y \in \mathcal{L}^{0,step} \right\}$$

A random variable $X \in \mathcal{L}^0$ is said to be *integrable* if $E[X^+]$ and $E[X^-]$ take both finite values. The collection of integrable random variable is denoted by $\mathcal{L}^1$. The expectation of elements in $\mathcal{L}^1$ is defined as

$$E[X] = E[X^+] - E[X^-]$$

Exercice 1.36. Show that E is well defined on $\mathcal{L}^0_+$ as well as on $\mathcal{L}^1$. ◊

Theorem 1.37. *Let (X_n) be an increasing sequence of positive random variables then*

$$\sup E[X_n] = \lim E[X_n] = E[\sup X_n] = E[X]$$

where $X = \sup X_n$ random variable extended real valued random variable.

Proof. By monotonicity, we clearly have $E[X_n] \leq E[X]$ for every n, therefore $\sup E[X_n] \leq E[X]$. Reciprocally, suppose that $E[X] < \infty$ and pick $\varepsilon > 0$ and $Y \in \mathcal{L}^{0,step}$ such that $Y \leq X$ and $E[X] - \varepsilon \leq E[Y]$ (why is it possible?). For $0 < c < 1$ define the sets $A_n = \{X_n \geq cY\}$. Since Y^n is increasing, it follows that A_n is an increasing sequence. Furthermore, since $cY \leq Y \leq X$ and $cY < X$ on $\{X > 0\}$, it follows that $\cup A_n = \Omega$. By non-negativity of X_n and monotonicity, it follows that

$$cE[1_{A_n} Y] \leq E[1_{A_n} X_n] \leq E[X_n]$$

and so

$$c \sup E[1_{A_n} Y] \leq c \sup E[X_n]$$

Since $Y = \sum_{l \leq k} \alpha_l 1_{B_l}$ for $\alpha_1, \ldots, \alpha_k \in \mathbb{R}_+$ and $B_1, \ldots, B_k \in \mathcal{F}$, it follows that

$$E\left[1_{A_n} Y\right] = \sum_{l \leq k} \alpha_l P[A_n \cap B_l].$$

However, since P is a probability measure, and A_n is increasing to Ω, it follows that $P[A_n \cap B_l] \nearrow P[\Omega \cap B_l] = P[B_l]$, and so

$$\sup E[1_{A_n} Y] = \sum_{l \leq k} \alpha_l \sup P[A_n \cap B_l] = \sum \alpha_l P[B_l] = E[Y]$$

Consequently

$$E[X] \geq \lim E[X_n] = \sup E[X_n] \geq cE[Y] \geq= cE[X] - c\varepsilon$$

which by letting c converging to 1 and ε to 0 yields the result. □

Proposition 1.38. *For each $X \in \mathcal{L}^0_+$, there exists an increasing sequence $(X_n) \subseteq \mathcal{L}^{0,step}_+$ such that $X_n(\omega) \nearrow X(\omega)$ and uniformly on each set $\{X \leq M\}$ where $M \in \mathbb{R}$.*

Proof. Let $A^n_k = \{(k-1)/2^n \leq X < k/2^n\}$ for $k = 1, \ldots, n2^n$ and every n. Define

$$X_n := \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{A^n_k} + n 1_{X > n}.$$

From the definition follows that $X_n \leq X$ for every n and $X(\omega) - 2^{-n} \leq X_n(\omega)$ for every $\omega \in \{X \leq n\}$ which up to the monotonicity which is left as an exercise ends the proof. □

Proposition 1.39. *The space $\mathcal{L}^1$ is a linear space and E is a monotone linear functional on it using the previous approximation theorem.*

Exercice 1.40. Show that $\mathcal{L}^1$ is a linear space and that E is linear and monotone on it. ♢

We finish this section with two of the most important assertions of integration theory.

Theorem 1.41. *Let (X_n) be a sequence in $\mathcal{L}^0$.*

Fatou's Lemma: *Suppose that $X_n \geq Y$ for some $Y \in \mathcal{L}^1$. Then it holds*

$$E\left[\liminf X_n\right] \leq \liminf E\left[X_n\right].$$

Dominated Convergence Theorem: *Suppose that $|X_n| \leq Y$ and $X_n \to X$, then it holds*

$$E\left[X\right] = \lim E\left[X_n\right]$$

Proof. Up to the variable change $X_n - Y$, we can assume that X_n is positive. Let $Y_n = \inf_{k\geq n} X_n$ which is an increasing sequence of positive random variable that converges to $\liminf X_n = \sup_n \inf_{k\geq n} X_k$. Notice also that $Y_n \leq X_k$ for every $k \geq n$ and therefore by monotonicity of the expectation $E[Y_n] \leq \inf_{k\geq n} E[X_k]$. We conclude Fatou's lemma with the monotone convergence theorem as follows

$$E\left[\liminf X_n\right] = \lim E\left[Y_n\right] = \sup E\left[Y_n\right] \leq \sup_n \inf_{k\geq n} E[X_k] = \liminf E[X_n]$$

A simple sign change shows that Fatou's lemma holds in the other direction, that is, if $X_n \leq Y$ for some $Y \in \mathcal{L}^1$, then it holds

$$\limsup E\left[X_n\right] \leq E\left[\limsup X_n\right]$$

Now the dominated convergence theorem assumptions yields that $-Y \leq X_n \leq Y$ for some $Y \in \mathcal{L}^1$. Hence, since $X_n \to X$, it follows that

$$\limsup E\left[X_n\right] \leq E\left[\limsup X_n\right] = E\left[X\right] = E\left[\liminf X_n\right] \leq \liminf E\left[X_n\right]$$

However, $\liminf E\left[X_n\right] \leq \limsup E[X_n]$ showing that $E[X_n]$ converges and

$$E[X] = \liminf E[X_n] = \limsup E[X_n] = \lim E[X_n].$$

which ends the proof. $\square$

One important property of the Lebesgue integral is that it is independant of the null sets on which functions may differ.

Proposition 1.42. *Let $X, Y \in \mathcal{L}^1_+$. Suppose that $X \geq Y$ P-almost surely, that is $P[X \geq Y] = 1$, then it follows that $E[X] \geq E[Y]$.*

In particular, if $X = Y$ P-almost surely, then it holds $E[X] = E[Y]$.

Proof. Suppose that $X \geq Y$ P-almost surely and defines $A = \{X < Y\}$ which is a negligeable set. It follows that $(X - Y)1_{A^c} \in \mathcal{L}^0_+$, and so $E[(X - Y)1_{A^c}] = E[X1_{A^c}] - E[Y1_{A^c}] \geq 0$ by monotonicity. On the other hand, $(Y - X)1_A \in \mathcal{L}^0_+$, and let $Z^n = \sum \alpha_k 1_{B^n_k}$ be an increasing sequence of step random variables that converges to $(Y - X)1_A$. Since $(Y - X)1_A = 0$ on A^c, it follows that $B^n_k =\subseteq A$ for every k, n and therefore $P[B^n_k] \leq P[A] = 0$ for every k, n. We deduce that $E[Z^n] = 0$ for every n and by Lebesgue monotone convergence, it follows that $E[(Y - X)1_A] = 0$. We conclude by noticing that $(X - Y) = (X - Y)1_{A^c} - (Y - X)1_A$. $\square$

This proposition allows to change in the monotone convergence theorem, Fatou's lemma as well as dominated convergence to replace convergence of random variable and inequalities by P-almost sure convergence and P-almost sure inequalities.
On $\mathcal{L}^1$ we can define the operator $X \mapsto \|X\|_1 = E[|X|]$. Verify that

- $X = 0$ implies $\|X\|_1 = 0$;
- $\|X + Y\|_1 \leq \|X\|_1 + \|Y\|_1$;
- $\|\lambda X\|_1 = |\lambda| \|X\|_1$

In other words, $\|\cdot\|$ is "almost" a norm if in the first point we had equivalence and not only implication. However, as the previous proposition shows, it actually holds

- $\|X\|_1 = 0$ if and only if $X = 0$ P-almost surely.

We therefore proceed as in Algebra. Define the equivalence relation[15] on $\mathcal{L}^0$ as $X \sim Y$ if and only if $X = Y$ P-almost surely (verify that this is indeed an equivalence relation). We can therefore define the quotient of equivalence classes $L^0 = \mathcal{L}^0/\sim$. We can work there just as in $\mathcal{L}^0$ in the P-almost sure sense, that is $X = Y$ means $X = Y$ P-almost surely, even if X is actually just a representant of its equivalence class. Inequality is also compatible with the equivalence relation and therefore $X \geq Y$ means $X \geq Y$ P-almost surely.

Every operation that is blind with respect to null measure sets can be carry over to L^0. This is the case of the expectation. Similarly, we can define L^1 as the set of equivalence classes of integrable random variable that coincide P-almost surely. Also, the operator $\|\cdot\|_1$ carries over and is on L^1 a true norm, making $(L^1, \|\cdot\|)$ a normed space.

We can further define for $1 \leq p \leq \infty$ the following operators on L^0,

$$\|X\|_p = \begin{cases} E\left[|X|^p\right]^{1/p} & \text{if } p < \infty \\ \inf\left\{m \colon P\left[|X| \leq m\right] = 1\right\} & \text{if } p = \infty \end{cases}$$

that give rise to the spaces

$$L^p := \left\{X \in L^0 \colon \|X\|_p < \infty\right\}$$

Theorem 1.43 (Jensen's inequality). *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$. It holds*

$$\varphi\left(E\left[X\right]\right) \leq E\left[\varphi(X)\right].$$

Proof. Let $x_0 = E[X]$. Since φ is a convex real valued function, the existence of sub-derivative for convex functions implies the existence of $a, b \in \mathbb{R}$ such that

$$\varphi(x) \geq ax + b, \text{ for all } x \in \mathbb{R} \quad \text{and} \quad \varphi(x_0) = ax_0 + b$$

Hence

$$E\left[\varphi(X)\right] \geq aE[X] + b = ax_0 + b = \varphi\left(E[X]\right)$$

which ends the proof. □

Exercice 1.44. Using Jensen's inequality, prove that $(\prod a_i)^{1/n} \leq 1/n \sum a_i$ where $a_1, \ldots, a_n > 0$. ◊

Theorem 1.45 (Hölder and Minkowsky Inequalities). *Let $p, q \in [1, \infty]$ be such that $1/p + 1/q = 1$. For every $X \in L^p$ and $Y \in L^q$, the Hölder inequality reads as follows:*

$$\|XY\|_1 = E\left[|XY|\right] \leq E\left[|X|^p\right]^{1/p} E\left[|Y|^q\right]^{1/q} = \|X\|_p \|Y\|_q.$$

For every $X, Y \in L^p$, the Minkowsky reads as follows:

$$\|X + Y\|_p = E\left[|X + Y|^p\right]^{1/p} \leq E\left[|X|^p\right]^{1/p} + E\left[|Y|^p\right]^{1/q} = \|X\|_p + \|Y\|_p.$$

Proof. As for the Hölder inequality, the case where $p = 1$ and $q = \infty$, the inequality follows from $|XY| \leq |X| \|Y\|_\infty$. Suppose therefore that p, q are conjugate with values in $]1, \infty[$. Without loss of generality, we may assume that X and Y a positive.

$$E[XY] = E[Y^q] \int XY^{1-q} \frac{Y^q dP}{E[Y^q]}$$

[15] An equivalence relation $\sim$ is a binary relation which is symmetric, that is $x \sim y$ if and only if $y \sim x$, reflexive, that is $x \sim x$ and transitive, that is $x \sim y$ and $y \sim z$ implies $x \sim z$.

Since $x \mapsto x^p$ is convex and $dQ := \frac{Y^q dP}{E[Y^q]}$ is a probability measure density, Jensen's inequality together with the fact that $p(1-q)+q=0$ and $1-1/p=1/q$ yields

$$E[XY] = E[Y^q]\int XY^{1-q}dQ \leq E[Y^q]\left[\int X^p Y^{p(1-q)}\frac{Y^q}{E[Y^q]}dQ\right]^{1/p} = E[X^p]^{1/p}E[Y^q]^{1/q}$$

As for the Minkowski inequality, in the case where $p=1$, it follows from $|x+y| \leq |x|+|y|$. The case where $p=\infty$ is also easy. Suppose therefore that $1<p<\infty$. First notice that by convexity it holds $|x+y|^p \leq 1/2\,|2x|^p + 1/2\,|2y|^p = 2^{p-1}\left(|x|+|y|\right)$. For information, this inequality ensures that L^p is a vector space. Now using the triangular inequality and Jensen's inequality in the end we get

$$\begin{aligned}
\|X+Y\|_p^p &= E\left[|X+Y|^p\right]\\
&\leq E\left[|X|\,|X+Y|^{p-1}\right] + E\left[|Y|\,|X+Y|^{p-1}\right]\\
&\leq \left(E\left[|X|^p\right]^{1/p} + E\left[|Y|^p\right]^{1/p}\right)E\left[|X+Y|^{(p-1)(p/(p-1))}\right]^{(p-1)/p}\\
&= \left(\|X\|_p + \|Y\|_p\right)E\left[|X+Y|^p\right]^{1-1/p} = \left(\|X\|_p + \|Y\|_p\right)\|X+Y\|_p^{p-1}
\end{aligned}$$

if $\|X+Y\|_p = 0$ the inequality is trivial, otherwise divide both sides by $\|X+Y\|^{p-1}$ $\square$

It follows in particular that L^p is a vector space and that $\|\|_p$ is a norm on L^p. We say that $X_n \to X$ in L^p for $(X_n), X$ in L^p if $\|X_n - X\|_p \to 0$.

Proposition 1.46. *Let (X_n) be a Cauchy sequence in $(L^p, \|\cdot\|_p)$ for $1 \leq p \leq \infty$. It follows that $X_n \to X$ in L^p for some $X \in L^p$.*

This proposition states that $(L^p, \|\cdot\|_p)$ is a Banach space.

Proof. We do the proof for $p<\infty$. Let (X_n) be a Cauchy sequence. By Cauchy property, we can take a subsequence (Y_n) of (X_n) such that $|Y_{n+1}-Y_n| \leq 2^{-n}$ and define $Z_n = |Y_1| + \sum_{k\leq n-1}|Y_{k+1}-Y_k|$ which is n increasing sequence of positive random variables converging to $Z = \sup Z_n$. Hence, the monoton convergence theorem shows that $E[Z^p] = \lim E[Z_n^p]$. By Minkowsky inequality it holds

$$E\left[Z_n^p\right] = \|Z_n\|_p^p \leq \left(\|Y_1\|_p + \sum_{k\leq n-1}\|Y_{k+1}-Y_k\|_p\right)^p \leq \left(\|Y_1\|_p + 1\right)^p$$

The left hand-side being independent of n, it follows by passing to the limit that $Z \in L^p$ and therefore $Z < \infty$ P-almost surely. On the other hand, since the absolute serie, $\sum |Z_{k+1}-Z_k|$ converges, it follows that $Y_n = Y-1+\sum_{k\leq n-1} Y_{k+1}-Y_k$ converges P-almost surely to some Y. Hence, $Y=\lim Y_n$ is in L^p since $|Y| = \lim |Y_n| \leq Z \in L^p$. We make use of dominated convergence on (Y_n) since $Y_n^p \to Y^p$ P-almost surely and $|Y_n|^p \leq Z^p \in L^p$, which implies that $E[|Y_n - Y|^p] \to 0$. It shows that a subsequence (Y_n) of (X_n) converges in L^p to some Y. As an exercise, using the Cauchy property, show that $X_n \to Y$ in L^p. $\square$

Definition 1.47. Let (X_n) be a sequence of random variables and X a random variable. We say that

- $X_n \to X$ P-almost surely if $P[\limsup X_n = \liminf X_n] = 1$;
- $X_n \to X$ in probability if $\lim P[|X_n - X| > \varepsilon] = 0$ for every $\varepsilon > 0$;
- $X_n \to X$ in L^p if $\|X_n - X\|_p \to 0$.

Proposition 1.48. *Let (X_n) be a sequence of random variables and X a random variable. The following assertions hold:*

(i) $X_n \to X$ *P-almost surely implies* $X_n \to X$ *in probability;*

(ii) $X_n \to X$ *in probability implies that* $Y_n \to X$ *P-almost surely for some subsequence* (Y_n) *of* (X_n)*;*

(iii) $X_n \to X$ *in* L^p *implies that* $Y_n \to X$ *P-almost surely for some subsequence* (Y_n) *of* (X_n)*.*

(iv) $X_n \to X$ *in probability and* $|X_n| \leq Y$ *for some* $Y \in L^1$ *implies* $X_n \to X$ *in* L^1*;*

Proof. Homework sheet. □

Proposition 1.49 (Chebyshev/Markov inequality). *Let X be a random variable, $\varepsilon > 0$. For every $0 < p < \infty$, the Chebyshev inqueality reads*

$$P\left[|X| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^p} E\left[|X|^p\right].$$

In the case where $p = 1$, the inequality is due to Markov.

Proof. Define $A_t = \{|X| \geq t\}$ and $g(x) = x^p$ which is an increasing function, so that consequently yields $0 \leq g(\varepsilon)1_{A_\varepsilon} \leq g(|X|)1_{A_\varepsilon}$. Thus, $0 \leq g(\varepsilon)P[A_\varepsilon] = E[g(\varepsilon)1_{A_\varepsilon}] \leq E[g(|X|)1_{A_\varepsilon}] \leq E[g(|X|)]$ which ends the proof.[16] □

1.4 Radon-Nikodym, Conditional Expectation

In this section we will make use of a central theorem of Functional analysis applied in the special case of Hilbert spaces.

Theorem 1.50. *Let H be an Hilbert space, and $T : H \to \mathbb{R}$ be a continuous linear functional. Then there exists $y \in H$ such that $T(x) = \langle y, x\rangle$ for every $x \in H$.*

This theorem will allow us to treat the following central theorem of measure theory in a rather simple way.

Theorem 1.51 (Radon-Nikodym Theorem). *Let $(\Omega, \mathcal{F})$ be measurable space and μ, ν two finite measures on $\mathcal{F}$ such that $\nu \ll \mu$. Then there exists an $f \in L^1(\mu)$ μ-almost surely unique and positive such that*

$$\nu\left(A\right) = \int_A f d\mu, \quad \textit{for every } A \in \mathcal{F}$$

The random variable f is called the *Radon-Nikodym derivative* of ν with respect to μ and denoted by $d\mu/d\nu$.

Proof. The proof is based on the argumentation of John von Neumann. Define $\sigma = \mu + \nu$. Since $\nu \ll \mu$, it follows that σ is equivalent to μ. Furthermore, it holds $L^2(\sigma) \subseteq L^2(\mu) \subseteq L^1(\mu)$. Define the linear functional $T : L^2(\sigma) \to \mathbb{R}$, $u \mapsto \int u d\mu$ which is well defined since μ is a finite measure equivalent to σ. Furthermore, using Jensen's inequality, it holds

$$\left|\int u d\mu\right| \leq \int |u|\, d(\mu + \nu) = \sigma(\Omega) \int |u| \frac{d\sigma}{\sigma(\Omega)} \leq \sigma(\Omega) \left(\int u^2 \frac{d\sigma}{\sigma(\Omega)}\right)^{1/2} = \sqrt{\sigma(\Omega)}\, \|u\|_{L^2(\sigma)}$$

[16]Note that the theorem holds by replacing the function $g(x) = x^p$ by any increasing function on $\mathbb{R}_+$.

showing that T is an $L^2(\sigma)$-continuous linear functional. Applying Riesz representation theorem, there exists $g \in L^2(\sigma)$ such that

$$T(u) = \int ug d\sigma, \quad u \in L^2(\sigma).$$

Taking $u = 1_A$ where first $A = \{g \leq 0\}$ and then $A = \{g > 1\}$ show that $0 < g \leq 1$ μ and σ almost surely.[17] Now, $1/g$ is measurable, positive μ and σ almost surely and it holds

$$\int_A \frac{d\mu}{g} = \int_A d\sigma = \sigma(A).$$

Taking $A = \Omega$, it follows from the finiteness of σ that $1/g \in L^1(\mu)$. Defining $f = 1/g - 1$ which is a positive measurable function in $L^1(\eta)$, it follows that

$$\nu(A) = \sigma(A) - \nu(A) = \int_A \frac{d\mu}{g} - \mu(A) = \int f d\mu$$

for every $A \in \mathcal{F}$ which ends the proof of the existence. Uniqueness is left as an exercise. $\square$

The Radon-Nikodym Theorem allows us to prove easily the existence of conditional expectations. Throughout this script, we adopt the notation

$$E[X; A] := E[X1_A]$$

Theorem 1.52. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. For every integrable random variable X, there exists P-almost surely a unique $\mathcal{G}$-measurable and integrable random variable Y such that*

$$E[X; A] = E[Y; A], \quad \textit{for every } A \in \mathcal{G}$$

Denoting $E[X|\mathcal{F}] := Y$, it holds – the following random variables are all in L^1:

(i) $E[|E[X|\mathcal{F}]|] \leq E[|X|]$*;*

(ii) $X \mapsto E[X\mathcal{F}]$ *is linear;*

(iii) $E[X|\mathcal{F}] \geq 0$ *P-almost surely whenever $0 \leq X$ P-almost surely;*

(iv) $E[X_n|\mathcal{F}] \nearrow E[X|\mathcal{F}]$ *whenever* $0 \leq X_n \nearrow X$*;*

(v) $E[YX|\mathcal{F}] = YE[X|\mathcal{F}]$ *whenever Y is $\mathcal{G}$-measurable;*

(vi) $E[XE[Y|\mathcal{F}]] = E[E[X|\mathcal{F}]Y] = E[E[X|\mathcal{F}]E[Y|\mathcal{F}]]$*;*

(vii) $E[E[X|\mathcal{G}_2]|\mathcal{G}_1]$ *whenever the σ-algebras are such that $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$.*

This unique random variable is called the *$\mathcal{G}$-conditional expectation* of X, and is denoted by $E[X|\mathcal{G}]$.

Proof. For X in L^1, it defines two finite measures on $\mathcal{G}$ given by

$$\nu^{\pm}(A) = E\left[X^{\pm}; A\right], \quad A \in \mathcal{G}$$

which are by definition both absolutely continuous with respect to P.[18] It follows from Radon-Nikodym Theorem 1.51 that there exists two P-almost surely unique positive random variables $Z^{\pm} \in L^1(\mathcal{G})$ such that

$$\nu^{\pm}(A) = E[Z^{\pm}; A]$$

[17] Check it as an exercise!

[18] Verify that these are indeed measures!

Defining $E[X|\mathcal{G}] = Z^+ - Z^- \in L^1(G)$ as the conditional expectation end the proof of the existence and uniqueness.
The properties *(i)*–*(vii)* are left as an exercise, where the monotone or dominated convergence of Lebesgue as to be used for some. □

Exercice 1.53. Under the assumptions of the Theorem 1.43, show that for a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, if $\varphi(X)$ is integrable, then it holds

$$\varphi\left(E[X|\mathcal{G}]\right) \leq E\left[\varphi(X)|\mathcal{G}\right] \qquad \Diamond$$

For your interest, here is the proof of the existence of conditional expectation using Hilbert projections.

Proof. Suppose first that $X \in L^2(\mathcal{F})$. Note that $L^2(\mathcal{F})$ is an Hilbert space for the norm $\|\cdot\|_2$ and $L^2(\mathcal{G})$ is a closed linear subspace of $L^2(\mathcal{F})$. Hence, by Hilbert's projection theorem, there exists a unique $Y \in L^2(\mathcal{G})$ such that $X - Y$ is orthogonal to $L^2(\mathcal{G})$. Since $1_A \in L^2(\mathcal{G})$ for every $A \in \mathcal{G}$ it follows that

$$E[(X-Y)1_A] = \langle X - Y, 1_A\rangle = 0, \quad A \in \mathcal{G}$$

showing the main assertion. The properties *(ii)*–*(vii)* are easy to verify in L^2 from the definition and therefore left as an exercise.
We show property *(i)*. For $X \in L^2$, let $A = \{E[X|\mathcal{F}] \geq 0\}$ which is an event in $\mathcal{G}$, it follows that

$$E\left[|E[X|\mathcal{F}]|\right] = E\left[E[X|\mathcal{F}]; A\right] - E\left[E[X|\mathcal{F}]; A^c\right] = E[X; A] - E[X; A^c] \leq E[|X|]$$

Hence

$$\sup\left\{E\left[|E[X|\mathcal{G}]|\right] : X \in L^2, \|X\|_1 = E[|X|] \leq 1\right\} \leq 1 \qquad \square$$

showing that the linear functional $E[\cdot|\mathcal{F}]$ on L^2 is L^1-continuous. Since L^2 is dense in L^1 which is complete, it follows that this linear extension extends uniquely to a continuous one on L^1, and the properties *(i)*–*(vii)* extends as well to L^1 which ends the proof.