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) $\mathcal{F}$ is closed under complementation. That is, $A^c \in \mathcal{F}$ whenever $A \in \mathcal{F}$;

(iii) $\mathcal{F}$ is closed under countable union. That is, $\cup A_n \in \mathcal{F}$ for every sequence (A_n) of elements in $\mathcal{F}$.

In probability theory, the components of a measurable space have the following meaning.

• Ω is a set modelling different *states* of the world about which there is uncertainty concerning its realization. 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 $x \in \Omega$ represents the temperature at 8:00 am tomorrow.
- Financial decision. Let $\Omega = [-1, 10]^2$ where for $(x, y) \in \Omega$, x and y represents the interest rate 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. Following the previous examples

- $A = \{H\}$ is the event that head will occur;
- $A = [13, 19]$ is the event that tomorrow at 8:00am, the temperature will lie between 13 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 fix an interest rate between 0.25% and 0.75% while the EU fix one between 0.9% and 1.8%, OR the USA fix an interest rate of 1% while the EU fix one between 1.7% and 2.1%.

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

- $\Omega \in \mathcal{F}$. Indeed, $\emptyset \in \mathcal{F}$ by *(i)* therefore $\Omega = \emptyset^c \in \mathcal{F}$ by condition *(ii)*.
- $\mathcal{F}$ is closed under countable intersection. That is, $\cap A_n \in \mathcal{F}$ for every sequence (A_n) of elements in $\mathcal{F}$. Indeed, by *(iii)* it follows that $A_n^c \in \mathcal{F}$ for every n, that is (A_n^c) is a sequence of elements in $\mathcal{F}$. 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 $\sigma(\mathcal{C})$.

[1] Note that this assumption follows from *(ii)* and *(iii)* when $\mathcal{F}$ is supposed to be non-empty. Why?

Proof. Let us show that $\mathcal{F}$ is a σ-algebra. Since $\mathcal{F}_i$ is a σ-algebra for every i, it follows from condition *(i)* that $\emptyset \in \mathcal{F}_i$ for every i and therefore $\emptyset \in \mathcal{F}$. Also, for $A \in \mathcal{F}$, that is, $A \in \mathcal{F}_i$ for every i, condition *(ii)* yields $A^c \in \mathcal{F}_i$ for every i, which, by definition, means $A^c \in \mathcal{F}$. Finally, for a sequence (A_n) of elements in $\mathcal{F}$, it follows that $A_n \in \mathcal{F}_i$ for every i and every n. Hence, $\cup A_n \in \mathcal{F}_i$ for every i by condition *(iii)*. Thus, $\cup A_n \in \mathcal{F}$. As for the second assertion, note that the power set $2^\Omega := \{A \colon A \subseteq \Omega\}$ of Ω is itself a σ-algebra that contains any collection of subsets of Ω, in particular $\mathcal{C}$. Hence, the intersection over all σ-algebra that contains $\mathcal{C}$ is non-empty and therefore, by what has been just proved, $\sigma(\mathcal{C})$ is a σ-algebra. The fact that it is the smaller one in terms of inclusion follows from the definition. □

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 $A, B \in \mathcal{F}$.

Remark 1.5. Just as in Lemma 1.3, arbitrary intersection of π-system or λ-system are themselves π-system or λ-system, respectively. Given a collection $\mathcal{C}$ of subset of Ω, we denote by $\pi(\mathcal{C})$ and $\lambda(\mathcal{C})$ the smallest π-system and λ-system containing $\mathcal{C}$, respectively. Clearly, any σ-algebra is a π- as well as a λ-system. ♦

Theorem 1.6. *Let Ω be a state space and $\mathcal{P}$ be a π-system. Then, the λ-system generated by $\mathcal{P}$ is a σ-algebra, that is $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$.*

Proof. We first show that if $\mathcal{C}$ is a λ-system closed under finite intersection, then it is a σ-algebra. By definition of a λ-system, we just have to check the stability under arbitrary countable union. To this end, let (A_n) be a sequence of elements in $\mathcal{C}$ and define $B_n = A_n \setminus (\cup_{k<n} A_k) = A_n \cap (\cap_{n<k} A_n^c)$, $n > 1$ and $B_1 = A_1$. As $\mathcal{C}$ is closed under complementation and we supposed that $\mathcal{C}$ is closed under finite intersection, it follows that (B_n) is a sequence of elements in $\mathcal{C}$. From $\cup B_n = \cup A_n$ and (B_n) pairwise disjoint, it follows from the λ-system assumption on $\mathcal{C}$ that $\cup A_n = \cup B_n \in \mathcal{C}$.
Now, it clearly holds $\lambda(\mathcal{P}) \subseteq \sigma(\mathcal{P})$. From what we just showed, we just have to check that $\lambda(\mathcal{P})$ is closed under finite intersection, since then $\lambda(\mathcal{P})$ would be a σ-algebra containing $\mathcal{P}$ and so $\sigma(\mathcal{P}) \subseteq \lambda(\mathcal{P})$. For $D \in \lambda(\mathcal{P})$, define $\mathcal{D}_D = \{A \subseteq \Omega \colon A \cap D \in \lambda(\mathcal{P})\}$ which is a λ-system. Indeed $\Omega \in \mathcal{D}_D$. 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, $A \cap D \in \lambda(\mathcal{P})$ and 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 of pairwise disjoints elements in $\mathcal{D}_D$. From the stability of $\lambda(\mathcal{P})$ under countable union of 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 indeed a λ-system. 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 $\mathcal{P} \subseteq \mathcal{D}_A$ for every $A \in \lambda(\mathcal{P})$. Hence $\lambda(\mathcal{P}) \subseteq \mathcal{D}_A$ 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. □

From their definition, σ-algebras as well as π-systems or λ-systems are structures of set, similar to another very important structure of set, namely topologies.

Definition 1.7. A topological space is a tuple $(\Omega, \mathfrak{T})$ where $\mathfrak{T}$ is a collection of subsets of a set Ω such that

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

(ii) $\mathfrak{T}$ is closed under finite intersection, that is, $O_1 \cap O_2 \in \mathfrak{T}$ whenever $O_1, O_2 \in \mathfrak{T}$;

(iii) $\mathfrak{T}$ is closed under arbitrary union, that is, $\cup O_i \in \mathfrak{T}$ for any arbitrary family (O_i) of elements in $\mathfrak{T}$.

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

A topology 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 $\mathfrak{T}(\mathfrak{B})$ generated by a collection $\mathfrak{B}$ 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 1.8. A topological base on a set Ω is a collection $\mathfrak{B}$ of subsets of Ω such that

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

(ii) for every $x \in O_1 \cap O_2$ for $O_1, O_2 \in \mathfrak{B}$, there exists $O_3 \in \mathfrak{B}$ with $x \in O_3$ and such that $O_3 \subseteq O_1 \cap O_2$.

Lemma 1.9. *Let $\mathfrak{B}$ be a topological base, and $\mathfrak{T}(\mathfrak{B})$ be the topology generated by $\mathfrak{B}$. It follows that $\mathfrak{T}(\mathfrak{B})$ is exactly the collection of arbitrary union of elements in $\mathfrak{B}$.*

Proof. Denote by $\mathcal{U}(\mathfrak{B})$ the collection of arbitrary unions of elements in $\mathfrak{B}$. By definition of $\mathfrak{T}(\mathfrak{B})$, it follows that $\mathfrak{B} \subseteq \mathcal{U}(\mathfrak{B}) \subseteq \mathfrak{T}(\mathfrak{B})$. Since $\mathfrak{T}(\mathfrak{B})$ is the smallest topology containing $\mathfrak{B}$, we just have to show that $\mathcal{U}(\mathfrak{B})$ is a topology itself. First $\Omega \in \mathcal{U}(\mathfrak{B})$ due to the first assumption of a topological base. As any union over an empty family is empty, it also follows that $\emptyset \in \mathcal{U}(\mathfrak{B})$. By definition, $\mathcal{U}(\mathfrak{B})$ is stable under arbitrary union. We are left to show that $\mathcal{U}(\mathfrak{B})$ is stable under intersection. Let $\tilde{O}_1 = \cup O_i, \tilde{O}_2 = \cup O_j \in \mathcal{U}(\mathfrak{B})$ for families $(O_i), (O_j)$ of elements in $\mathfrak{B}$. It follows that $\tilde{O}_1 \cap \tilde{O}_2 = \cup_{i,j}(O_i \cap O_j)$. By definition of a topological base, for every i, j and every $x \in O_i \cap O_j$, there exists $O_{i,j}^x \in \mathfrak{B}$ such that $x \in O_{i,j} \subseteq O_i \cap O_j$. Hence, $\cup_{x \in O_i \cap O_j} O_{i,j}^x = O_i \cap O_j$ from which follows that $\tilde{O}_1 \cap \tilde{O}_2 = \cup_{i,j,x \in O_i \cap O_j} O_{i,j}^x \in \mathcal{U}(\mathfrak{B})$ showing that $\mathcal{U}(\mathfrak{B})$ is a topology. ☐

For a set $A \subseteq \Omega$, we define the interior and closure of A as

$$\text{Int}(A) = \cup\{O \colon O \text{ open with } O \subseteq A\}, \qquad \text{Cl}(A) = \cap\{F \colon F \text{ closed and } A \subseteq F\}$$

Clearly, A is open or closed if, and only if, $A = \text{Int}(A)$ or $A = \text{Cl}(A)$, respectively.

Lemma 1.10. *Let $A \subseteq \Omega$ be a subset of Ω which topology is generated by a topological base $\mathfrak{B}$. Then it holds*

$$\begin{aligned} Int(A) &= \cup\{O \colon O \subseteq A, O \in \mathfrak{B}\} = [Cl(A^c)]^c \\ Cl(A) &= \{\omega \colon \omega \in O, O \in \mathfrak{B} \text{ and } O \cap A \neq \emptyset\} \end{aligned}$$

Proof. The first equality for the interior follows from the fact that every open set in Ω is an arbitrary union of elements in $\mathfrak{B}$. The second equality follows from de Morgan's law

$$\begin{aligned} \text{Int}(A) &= [[\cup\{O \colon O \subseteq A, O \text{ open}\}]^c]^c = [\cap\{O^c \colon A^c \subseteq O^c, O \text{ open}\}]^c \\ &= [\cap\{F \colon A^c \subseteq F, F \text{ closed}\}]^c = [\text{Cl}(A^c)]^c. \end{aligned}$$

As for the closure equality, it holds

$$\begin{aligned} \text{Cl}(A) &= [\text{Int}(A^c)]^c = [\cup\{O \colon O \subseteq A^c, O \in \mathfrak{B}\}]^c = [\{\omega \colon \omega \in O, O \subseteq A^c \text{ for some } O \in \mathfrak{B}\}]^c \\ &= [\{\omega \colon \omega \in O, O \cap A = \emptyset \text{ for some } O \in \mathfrak{B}\}]^c = \{\omega \colon \omega \in O, O \in \mathfrak{B} \text{ and } O \cap A \neq \emptyset\}. \end{aligned}$$ ☐

Example 1.11. A metric space (Ω, d) is a set Ω together with a function $d : \Omega \times \Omega \to [0, \infty[$ – called *metric* or *distance* – with the properties

(i) *identity:* $d(\omega, \omega') = 0$ if, and only if, $\omega = \omega'$;

(ii) *symmetry:* $d(\omega, \omega') = d(\omega', \omega)$;

(iii) *triangular inequality:* $d(\omega, \omega'') \leq d(\omega, \omega') + d(\omega', \omega'')$.

The collection $\mathfrak{B}$ of *open balls* $B_n(\omega) := \{\omega' \colon d(\omega, \omega') < 1/n\}$ for $n \in \mathbb{N}$ and $\omega \in \Omega$ is a topological base due to the triangular inequality, and the topology generated by this family is called the metric topology. The nice thing about metric spaces is that we can characterize the topology by converging sequences. Indeed, a set $F \subseteq \Omega$ is closed if, and only if, the limit of each converging sequence[2] of elements in F also belongs to F. This follows from Lemma 1.10, the definitions of the open balls and the triangular inequality.
In $\mathbb{R}^d$, the balls $B_{1/n}(q) = \{x \in \mathbb{R}^2 \colon d(x, q) = \|x - q\| < 1/n\}$ where $n \in \mathbb{N}$ and $q \in \mathbb{Q}^d$ constitute a topological basis.[3] The resulting topology is the usual euclidean topology on $\mathbb{R}^d$. A particularity of this topology is that the topological base is countable.[4] 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[5] (ω_n) of elements in Ω, then the countable collection of open balls $B_{1/n}(\omega_n) = \{\omega \in \Omega \colon d(\omega_n, \omega) < 1/m\}$ for $m, n \in \mathbb{N}$ is a countable topological base of Ω. Such spaces play a central role in probability theory since the Borel σ-algebra, see following definition, coincide with the σ-algebra generated by this countable family of balls. $\diamondsuit$

Definition 1.12. Let $(\Omega, \mathfrak{T})$ be a topological space. The σ-algebra $\mathcal{B}(\mathfrak{T})$ generated by the open sets of Ω is called the *Borel* σ-algebra.

Remark 1.13. In the case of $\mathbb{R}^d$, since it is generated by the countable family $\mathfrak{B}$ of open balls centered around a rational, it follows that the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}^d$ is fully generated by the topological base, that is $\mathcal{B} = \sigma(\mathfrak{B})$. $\blacklozenge$

Exercice 1.14. 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

$$\mathcal{A} = \{F \colon F \text{ closed subset of } \mathbb{R}\}$$

$$\begin{aligned}
&\mathcal{A} = \{]a, b[\colon a \leq b \text{ with } a, b \in \mathbb{R}\} && \mathcal{A} = \{[a, b] \colon a \leq b \text{ with } a, b \in \mathbb{R}\}\\
&\mathcal{A} = \{]a, b] \colon a \leq b \text{ with } a, b \in \mathbb{R}\} && \mathcal{A} = \{[a, b[\colon a \leq b \text{ with } a, b \in \mathbb{R}\}\\
&\mathcal{A} = \{]-\infty, b] \colon b \in \mathbb{R}\} && \mathcal{A} = \{]-\infty, b[\colon b \in \mathbb{R}\}\\
&\mathcal{A} = \{[a, \infty[\colon a \in \mathbb{R}\} && \mathcal{A} = \{]a, \infty[\colon a \in \mathbb{R}\}\\
&\mathcal{A} = \{]a, b[\colon a \leq b \text{ with } a, b \in \mathbb{Q}\} && \mathcal{A} = \{[a, b] \colon a \leq b \text{ with } a, b \in \mathbb{Q}\}\\
&\mathcal{A} = \{]a, b] \colon a \leq b \text{ with } a, b \in \mathbb{Q}\} && \mathcal{A} = \{[a, b[\colon a \leq b \text{ with } a, b \in \mathbb{Q}\}\\
&\mathcal{A} = \{]-\infty, b] \colon b \in \mathbb{Q}\} && \mathcal{A} = \{]-\infty, b[\colon b \in \mathbb{Q}\}\\
&\mathcal{A} = \{[a, \infty[\colon a \in \mathbb{Q}\} && \mathcal{A} = \{]a, \infty[\colon a \in \mathbb{Q}\}
\end{aligned}$$

$\diamondsuit$

[2] Naturally, a sequence (ω_n) converges to ω, and denoted by $\omega_n \to \omega$ if $d(\omega_n, \omega) \to 00$.
[3] Check this using the triangular inequality and the density of $\mathbb{Q}^d$ in $\mathbb{R}^d$.
[4] Such topologies generated by a countable base are called second countable topologies.
[5] A subseteq $A \subseteq \Omega$ is called dense if $\mathrm{Cl}(A) = \Omega$.

Exercice 1.15. Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces and $X : \Omega \to \Omega'$ a function. Show that

- $\{X^{-1}(B) \colon B \in \mathcal{F}'\}$ 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. $\diamondsuit$

Definition 1.16. Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces. A measurable function[6] is a function $X : \Omega \to S$ such that

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

If $\Omega' = \mathbb{R}$ and $\mathcal{F}'$ is the Borel σ-algebra, we call X a *random variable*.

Remark 1.17. 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

$$\{X \in B\} := X^{-1}(B), \qquad \{X = x\} := X^{-1}(\{x\}), \qquad \{X \leq x\} := X^{-1}(]-\infty, x]), \qquad \cdots \quad \blacklozenge$$

Remark 1.18. If Ω is a state space without a predefined σ-algebra, and $X : \Omega \to \Omega'$ is a function with value in a measurable space $(\Omega', \mathcal{F}')$, then

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

is the smallest σ-algebra for which X is a measurable function. In other words, in the framework of the Definition 1.16 of measurable function, it holds $\sigma(X) \subseteq \mathcal{F}$. More generally, let (X_i) be a family of functions $X_i : \Omega \to \Omega_i'$ where $(\Omega_i', \mathcal{F}_i')$ is a family of measurable spaces, then $\sigma(X_i \colon i) = \sigma(\{X_i^{-1}(B) \colon B \in \mathcal{F}_i', i\})$ is the smallest σ-algebra such that each X_i is measurable. $\blacklozenge$

Lemma 1.19. *The composition of measurable functions is measurable.*

Proof. Let $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ and $(\Omega'', \mathcal{F}'')$ be three measurable spaces and $X : \Omega \to \Omega'$, $Y : \Omega' \to \Omega''$ be two measurable functions. Define $Z = Y \circ X : \Omega \to \Omega''$, $\omega \mapsto Z(\omega) = Y(X(\omega))$ the composition of X and Y. For every $A \in \mathcal{F}''$, it holds $Z^{-1}(A) = X^{-1}(Y^{-1}(A)) = X^{-1}(B)$ where $B = Y^{-1}(A)$. Since Y is measurable, it follows that $B = Y^{-1}(A) \in \mathcal{F}'$. Further, the measurability of X implies that $Z^{-1}(A) = X^{-1}(B) \in \mathcal{F}$ showing that Z is measurable. $\square$

Proposition 1.20. *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces. If $\mathcal{C}'$ is a collection of subsets of Ω' such that $\mathcal{F}' = \sigma(\mathcal{C}')$, then for $X : \Omega \to \Omega'$, 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 $\mathcal{D}' := \{A \in \mathcal{F}' \colon X^{-1}(A) \in \mathcal{F}\}$. To show measurability of X, we just have to show that $\mathcal{F}' = \mathcal{D}'$. By assumption, $\mathcal{C}' \subseteq \mathcal{D}'$, therefore, $\mathcal{F}' = \sigma(\mathcal{C}') \subseteq \sigma(\mathcal{D}') \subseteq \mathcal{F}'$. We are left to show that $\sigma(\mathcal{D}') = \mathcal{D}'$. This is however immediate since X^{-1} commutes with complements, arbitrary union and $X^{-1}(\emptyset) = \emptyset$. $\square$

[6] If necessary, we say $\mathcal{F}$-$\mathcal{F}'$-measurable if the context is not clear with respect to which we are measurable.

Combined with Exercise 1.14, 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 1.21. Let $(\Omega, \mathfrak{T}), (\Omega', \mathfrak{T}')$ be two topological spaces. A function $X : \Omega \to \Omega'$ is called *continuous* if $X^{-1}(O')$ is open for every open set $O' \subseteq \Omega'$.[7]
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 $t \in \mathbb{R}$.

Remark 1.22. 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 every $\omega_n \to \omega$, respectively.

As for the first assertion, suppose that X is continuous and pick a converging sequence $\omega_n \to \omega$. By continuity of X, $X^{-1}(]X(\omega) - 1/m, X(\omega) + 1/m[)$ is an open set containing ω for every integer m. Hence, there exists $\delta > 0$ such that $B_\delta(\omega) \subseteq X^{-1}(]X(\omega) - 1/m, X(\omega) + 1/m[)$. Since $\omega_n \to \omega$, there exists n_0 such that $\omega_n \in B_\delta(\omega)$ for every $n \geq n_0$. All together, it implies that for every $m \in \mathbb{N}$, there exists n_0 such that $|X(\omega) - X(\omega_n)| \leq 1/m$ for every $n \geq n_0$. It shows that $X(\omega_n) \to X(\omega)$ for every $\omega_n \to \omega$. Reciprocally, let $F \subseteq \mathbb{R}$ be a closed set and (ω_n) be a sequence in $X^{-1}(F)$ converging to ω. By assumption, it follows that the sequence $(X(\omega_n))$ of elements in F converges to $X(\omega)$. Since F is closed, it follows that $X(\omega) \in F$ and therefore $\omega \in X^{-1}(F)$ showing that $X^{-1}(F)$ is closed. Thus X is continuous.
Let us show the characterization of lower semi-continuity. Suppose that X is lower semi-continuous and let (ω_n) be a sequence in Ω converging to ω. Let $a = \liminf X(\omega_n) = \sup_n \inf_{k \geq n} X(\omega_k)$. It follows that $X^{-1}(] - \infty, a]) \supseteq \cap_n \cup_{k \geq n} \{\omega_k\}$. Since $X^{-1}(] - \infty, a])$ is closed, it follows that $X^{-1}(] - \infty, a]) \supseteq \mathrm{Cl}(\cap_n \cup_{k \geq n} \{\omega_k\})$. Since $\omega_n \to \omega$, it follows that $\omega \in \mathrm{Cl}(\cap_n \cup_{k \geq n} \{\omega_k \colon k \geq n\})$, and therefore $X(\omega) \leq a = \liminf X(\omega_n)$. Reciprocally, let $F = X^{-1}(] - \infty, a])$ and let (ω_n) be a sequence in F converging to ω. It follows that $X(\omega_n) \leq a$ for every n and therefore $\liminf X(\omega_n) \leq a$. Since $X(\omega) \leq \liminf X(\omega_n) \leq a$ it follows that $\omega \in F$ showing that $X^{-1}(] - \infty, a])$ is closed. ♦

Remark 1.23. As for measurable functions, you can define topologies generated by family of functions, analogue to Remark 1.18 as the smallest topology that makes functions continuous. Also, the composition of continuous functions is continuous. ♦

Corollary 1.24. *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;*[8]
- *X is a lower semi-continuous function;*[9]

[7] Or equivalently $X^{-1}(F')$ is closed for every closed set $F' \subseteq \Omega'$.
[8] That is $\{\omega \colon X(\omega) \geq x\}$ is closed for every $x \in \mathbb{R}$.
[9] That is $\{\omega \colon X(\omega) \leq x\}$ is closed for every $x \in \mathbb{R}$.

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 1.20 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 $t \in \mathbb{R}$. $\square$

Definition 1.25. 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

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

Exercice 1.26. Show that in 2 dimensions, it holds $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A \times B \colon A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$. $\Diamond$

Under the notations of Definition 1.25, a *product cylinder set* is a set $A \subseteq \Omega$ of the form – assuming that the index set is directed[10] –

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

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 1.27. 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 $b_k \in \{-1, 1\}$, $k = 1, \ldots, n$. $\Diamond$

We now focus mainly on random variables. The following propositions and theorems use a lot the structure of $\mathbb{R}$, in particular its complete order that generates the topology.[11] 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 1.28. *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;*

[10] Which is always possible from the general theory of boolean algebra, see appendix A.1 where more is said about product σ-algebras.
[11] Think why for each of the following assertions, the structure of $\mathbb{R}$ is so important.

- $\max(X, Y)$ *and* $\min(X, Y)$ *are random variables;*
- $\sup X_n$ *and* $\inf X_n$ *are extended real valued random variables;*[12]
- $\liminf X_n := \inf_n \sup_{k\geq n} X_k$ *and* $\limsup X_n := \inf_n \sup_{k\geq n} X_k$ *are extended real valued random variables;*
- $A := \{\lim X_n$ *exists*$\} := \{\omega \colon \lim X_n(\omega)$ *exists*$\} = \{\liminf X_n = \limsup X_n\}$ *is measurable.*

Proof. First, let $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function. It follows that the function $g(X, Y)$ is measurable for the following reason. First, the mapping $T : \Omega \to \mathbb{R} \times \mathbb{R}$, $\omega \mapsto (X(\omega), Y(\omega))$ is measurable with respect to product Borel σ-algebra on $\mathbb{R} \times \mathbb{R}$. Indeed, for every two Borel sets A, B of the real line, it follows that $T^{-1}(A \times B) = \{X \in A\} \cap \{Y \in B\}$ which is an element in $\mathcal{F}$ by measurability of X and Y. Now, since the product sets $A \times B$ for A, B Borel sets in $\mathbb{R}$ generates the Borel product σ-algebra on $\mathbb{R}^2$, see Exercise 1.26, it follows that T is measurable. By continuity of g together with Corollary 1.24, it follows that g is measurable and therefore $g \circ T$ is a random variable by Lemma 1.19. Taking $g(x, y) = ax + b$, $g(x, y) = xy$, $g(x, y) = \max(x, y)$ and $g(x, y) = \min(x, y)$, the third tree points follows.
Let $a \in \mathbb{R}$, it holds $\{\sup_n X_n \leq a\} = \{X_n \leq a\colon \text{ for every } n\} = \cap_n\{X_n \leq a\}$ which is measurable since $\{X_n \leq a\}$ is measurable. Since $]-\infty, a]$ generates the Borel σ-algebra, it follows that $\sup_n X_n$ is measurable. The same argumentation for $\inf X_n$ follows with $\{\inf X_n \geq a\}$.[13] Let $a \in \mathbb{R}$, it holds

$$\{\liminf X_n \leq a\} = \{\sup_n \inf_{k\geq n} X_n \leq a\} = \{X_k \leq a \colon \text{for some } k \geq n \text{ for all } n\} = \cap_n \cup_{k\geq n} \{X_k \leq a\}$$

and so the measurability of $\liminf X_n$ follows by the same argumentation as above and the stability of the σ-algebra under countable intersection and union. Finally $A = \{\liminf X_n = \limsup X_n\} = \{Z = 0\}$ for the random variable $\liminf X_n - \limsup X_n$ and therefore is measurable. □

1.2. Probability Measures

Definition 1.29. 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[14] 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.

Lemma 1.30. *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* $A \subseteq B$*;*
- $P[A^c] = 1 - P[A]$*;*
- $P[A \cup B] + P[A \cap B] = P[A] + P[B]$*;*

[12] 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 $\mathbb{R}$.
[13] Or the fact that $\inf_n X_n = -\sup X_n$ and $-x$ is a continuous function.
[14] That is $A_n \cap A_m = \emptyset$ for every $m \neq n$.

- σ-subadditivity: $P[\cup A_n] \leq \sum P[A_n]$
- lower semi-continuity: $\lim_n P[\cup_{k\leq n} A_k] = P[\cup A_n]$
- upper semi-continuity: $\lim_n P[\cap_{k\leq n} A_k] = P[\cup A_n]$

Proof. The last three properties follows from Lemma 1.35. 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 $P[\Omega] = 1$. If we drop the assumption that $P[\Omega] = 1$, then P is a measure – traditionally denoted with the Greek letters $\mu, \nu, \ldots$

- 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 $A_1 \subseteq A_2 \subseteq \ldots$ 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[15] $N \subseteq \Omega$ 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 $\mathcal{F}$.[16]
- 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 $P \ll Q$.

In probability theory, we often adopt the following short handwritings

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

Example 1.31 (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 $\sum p(\omega) = 1$.[17] An important example of which is when $\Omega = \{1, \ldots, N\}$ for some integer N and we take $p(n) = 1/N$ for every $n = 1, \ldots, N$. The resulting probability measure is called the *uniform probability distribution* on Ω.

[15] Not necessarily measurable
[16] Be careful that the completed σ-algebra depends on P.
[17] Why?

2) **Dirac measure**. The Dirac measure at $\omega_0 \in \Omega$ is defined as the set value function

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

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

3) **Counting measure**. Define

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

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 $\mathcal{F}$ the Borel σ-algebra of the real line, we define

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

where λ is the Lebesgue measure on $\mathbb{R}$. This is the famous normal distribution, and you certainly already showed in your mathematical life that $P[\mathbb{R}] = 1$ so that P is 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 1.27, 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 then head twice 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 1.32. 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. Inspection shows that the ring generated by a semi-ring $\mathcal{R}$ is exactly the collection of sets $A = \cup_{k \leq n} A_k$ for every finite family $(A_k)_{k \leq n}$ of pairwise disjoint elements in $\mathcal{R}$.

Definition 1.33. Let $\mathcal{R}$ be a semi-ring. A function $P : \mathcal{R} \to [0, \infty]$ is called a *content* if

- $P[\emptyset] = 0$;

- $P[A\cup B]=P[A]+P[B]$ for every $A,B\in\mathcal{R}$ such that $A\cap B=\emptyset$ and $A\cup B\in\mathcal{R}$.

If a content P satisfies

- $P[\cup_n A_n]=\sum P[A_n]$ for every sequence (A_n) of pairwise disjoint elements in $\mathcal{R}$ and such that $\cup_n A_n\in\mathcal{R}$.

then P is called a *premeasure*.

Recall that unlike rings, semi-rings are in general not closed under union. If P is a content on a ring, then it is finitely additive with respect to finite family of disjoints events. Furthermore, if P is a content on a semi-ring $\mathcal{R}$, it can easily be extended to a content on the ring generated by $\mathcal{R}$. Indeed, as mention above, the ring generated by the semi-ring $\mathcal{R}$ is the collection $A=\cup_{k\leq n}A_k$ for finite pairwise disjoints family $(A_n)_{n\leq k}$ of elements in $\mathcal{R}$. So defining $P[A]:=\sum_{k\leq n}P[A_k]$ provides the desired extension.

Remark 1.34. Note that a content on a ring is automatically

- monotone: indeed for $A\subseteq B$ it holds $P[B]=P[B\setminus A\cup A]=P[B\setminus A]+P[A]\geq P[A]$. In particular $P[B\setminus A]=P[B]-P[A]$.

- sub-additivity: that is for $(A_k)_{k\leq n}$ finite family of elements in $\mathcal{R}$ and $A\subseteq\cup_{k\leq n}A_k$ it holds $P[A]\leq\sum_{k\leq n}P[A_k]$. Indeed, define $B_1=A\cap A_1$ and recursively $B_k=A\cap(A_k\setminus(\cup_{l<k}A_l))$ for $k\leq n$. By definition of a ring, it defines a finite disjoint family of elements in $\mathcal{R}$ and it holds $B_k\subseteq A_k$ for every k as well as $A=\cup_{k\leq n}B_k$. Hence, by additivity and monotony from the previous point, it follows that $P[A]=P[\cup_{k\leq n}B_k]=\sum P[B_k]\leq\sum P[A_k]$. ♦

A probability measure on a measurable space in particular a content on a ring. The following central lemma for holds for probability measures, but it also holds for the broader class of finite content on a ring. Since we will need it in the appendix for the proof of Caratheodory theorem, we state it in its generality here.

Lemma 1.35. *Let $\mathcal{R}$ be a ring and $P:\mathcal{R}\to[0,\infty]$ a finite content, that is $P[A]<\infty$ for every $A\in\mathcal{R}$. Then the following are equivalent*

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

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

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

(iv) Continuous at $\emptyset$: $\inf_n P[A_n]=0$ *for every countable family (A_n) of decreasing elements in $\mathcal{R}$ such that $\cap A_n=\emptyset$;*

(v) σ-sub-additivity: $P[A]\leq\sum P[A_n]$ *for every countable family (A_n) of elements in $\mathcal{R}$, $A\in\mathcal{R}$ such that $\cup A_n\in\mathcal{R}$.*

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 $B_n = A_1 \setminus A_n$ defines 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[18]

$$\begin{aligned} P[A_1] - \inf P[A_n] &= \sup(P[A_1] - P[A_n]) = \sup P[A_1 \setminus A_n] \\ &= P[\cup A_1 \setminus A_n] = P[A_1 \setminus A] = P[A_1] - P[A] \end{aligned}$$

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 $B_n = A \setminus A_n$ defines 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.
We show that (i) implies (v). Let (A_n) be a countable family of elements in $\mathcal{R}$ and $A \in \mathcal{R}$ such that $A \subseteq \cup A_n$. Define $B_1 = A \cap A_1$ and $B_n = A \cap (A_n \setminus \cup_{k<n} A_k)$ which by induction and the definition of a σ-ring is countable family of disjoint elements in $\mathcal{R}$ such that $A = \cup B_n \in \mathcal{R}$ and $B_n \subseteq A_n$ for every n. Further, since P is a premeasure it is in particular monotone, see Remark 1.34, hence

$$P[A] = P[\cup B_n] = \sum P[B_n] \leq \sum P[A_n]$$

showing the σ-sub-additivity. Reciprocally, let P be a σ-subadditive content on $\mathcal{R}$. It follows in particular that it is monotone, see Remark 1.34. Let (A_n) be a countable family of pairwise disjoint events in $\mathcal{R}$ such that $A = \cup A_n \in \mathcal{R}$. It follows that

$$\sum P[A_n] = \sup_n \sum_{k\leq n} P[A_k] = \sup_n P[\cup_{k\leq n} A_k] \leq \sup_n P[A] = P[A].$$

The σ-sub-additivity yields the reverse equality, showing σ-additivity. $\square$

Example 1.36. 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! $\lozenge$

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 hard, 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.

Theorem 1.37 (Caratheordory Extension Theorem). *Let Ω be a non empty-set, $\mathcal{R}$ a semi-ring such that $\Omega = \cup A_n$ for some countable family (A_n) of elements in $\mathcal{R}$. Suppose that $P : \mathcal{R} \to [0,\infty]$ is a content such that*

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

(i) $P[A] < \infty$ *for every* A*;*

(ii) P *is* σ***-sub-additive****, that is* $P[\cup A_n] \leq \sum P[A_n]$ *whenever* (A_n) *is a countable family of elements in* $\mathcal{R}$ *such that* $\cup A_n \in \mathcal{R}$*, and* $A \in \mathcal{R}$ *with* $A \subseteq \cup A_n$*.*

Then P *can be extended to a measure* P *on* $\mathcal{F} = \sigma(\mathcal{R})$*.*

The proof of which is done in the appendix as well as the construction of several important measures, among others such as the probability coinciding with the fair coin toss when the experience is conducted infinitely many times. The main question though is if such an extension is unique. This follows however from Dynkin Theorem 1.6.

Proposition 1.38. *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 1.6, it follows that $\mathcal{F} = \sigma(\mathcal{P}) \subseteq \mathcal{C} \subseteq \mathcal{F}$ showing that $P = Q$. □

1.3. Integration

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

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

for a $A_1, \ldots, A_n \in \mathcal{F}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Note that this representation is not unique![19] 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 simple random variable X with respect to P as

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

Exercice 1.39. Show that the definition of the expectation is a well defined operator on $\mathcal{L}^{0,step}$. [20] ♢

Proposition 1.40. *On* $\mathcal{L}^{0,step}$*, the following properties hold*

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

Proof. Let $X = \sum_{k \leq n} \alpha_k 1_{A_k}$ and $Y = \sum_{k \leq m} \beta_k 1_{B_k}$ be two simple random variables. Without loss of generality by taking a finer partition that contains both families, we may assume that $m = n$, $A_k = B_k$ and (A_k) is a pairwise disjoint family. If $X \leq Y$, it follows that $\alpha_k = X(\omega) \leq Y(\omega) = \beta_k$ for every $\omega \in A_k$ and $k = 1, \ldots, n$. Hence, $\hat{E}[X] = \sum_{k \leq n} \alpha_k P[A_k] \leq \sum_{k \leq n} \beta_k P[A_k] = E[Y]$. For $a, b \in \mathbb{R}$, it holds $\hat{E}[aX + bY] = \sum_{k \leq n} (a\alpha_k + b\beta_k) P[A_k] = a \sum_{k \leq n} \alpha_k P[A_k] + b \sum_{k \leq n} \beta_k P[A_k] = a\hat{E}[X] + b\hat{E}[Y]$.

[19] Why?

[20] Why?

Definition 1.41. For $X \in \mathcal{L}^0_+ := \{X \in \mathcal{L}^0 \colon X \geq 0\}$ we define

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

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

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

By definition, E is an extension of $\hat{E}$ to the space $\mathcal{L}^0_+$ since for $Y \in \mathcal{L}^{0,step}_+$ it holds $E[Y] = \hat{E}[Y]$. This is the same on $\mathcal{L}^1$, as $\mathcal{L}^{0,step} \subseteq \mathcal{L}^1$ and it holds $E[Y] = \hat{E}[Y]$ for every $Y \in \mathcal{L}^{0,step}$. We therefore remove the hat on the top of the expectation symbol everywhere. Finally, if X is a positive extended real valued random variable the expectation as given by the definition above is also well defined.

Theorem 1.42. *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 $X = \sup X_n$ is an 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 $E[X] - \varepsilon \leq E[Y]$.[21] For $0 < c < 1$ define the sets $A_n = \{X_n \geq cY\}$. Since X^n is increasing to X, it follows that A_n is an increasing sequence of events. 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

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

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 from the lower semi-continuity of probability measures, see Lemma 1.30, 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

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

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

Proposition 1.43. *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 $M \in \mathbb{R}$.*

[21] Why is it possible?

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

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

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 left as an exercise, ends the proof. □

Proposition 1.44. *The expectation operator is a monotone on $\mathcal{L}_+^0$ and $E[aX + bY] = aE[X] + bE[Y]$ for every $a, b \in \mathbb{R}$ with $a, b \geq 0$. Further, the space $\mathcal{L}^1$ is a linear space and the expectation operator E is a monotone on it.*

Proof. Let $X, Y \in \mathcal{L}_+^0$, real numbers $a, b \geq 0$ and, according to proposition 1.43, denote by X_n, Y_n two increasing sequence of simple random variable such that $X_n \nearrow X$ and $Y^n \nearrow Y$. If $X \leq Y$ it follows from the construction in 1.43 that $X_n \leq Y_n$. From the monotonicity of E on $\mathcal{L}^{0,step}$ and the monotone convergence theorem that $E[X] = \lim E[X_n] \leq \lim E[Y_n] = E[Y]$. The same argumentation using the linearity of E on $\mathcal{L}^{0,step}$, it follows that $E[aX + bY] = \lim E[aX_n + bY_n] = \lim aE[X_n] + bE[Y_n] = a \lim E[X_n] + b \lim E[Y_n] = aE[X] + bE[Y]$. The case of $\mathcal{L}^1$ follows the same lines. □

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

Theorem 1.45. *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 $X = \lim X_n = \liminf X_n = \limsup X_n$, 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. □

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

Proposition 1.46. *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 $E[X] = E[Y]$. Also, if $X \geq 0$ P-almost surely and $E[X] = 0$, then it follows that $X = 0$ P-almost surely.

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 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[22] $X \sim Y$ on $\mathcal{L}^0$ if, and only if, $X = Y$ P-almost surely.[23] 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 on L^0_+. Similarly, we can define L^1 as the set of equivalence classes of integrable random variable that coincide P-almost surely. Also, since the operator $\|\cdot\|_1$ does not take into account objects defined on negligeable sets, it carries over to L^1 is there 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

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

[22] 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 $x \sim z$.

[23] Verify that this is indeed an equivalence relation.

Theorem 1.47 (Jensen's inequality). *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and X be an integrable random variable. 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. $\square$

Exercice 1.48. Using Jensen's inequality, prove that $(\prod a_i)^{1/n} \leq 1/n \sum a_i$ where $a_1, \ldots, a_n > 0$. $\diamond$

Theorem 1.49 (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:*

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

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

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

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

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

where E_Q is the expectation operator under the measure Q with density $dQ := Y^q dP/E[Y^q]$.[24] Defining the convex function $x \mapsto \varphi(x) = x^p$, Jensen's inequality together with the fact that $p(1-q) + q = 0$ and $1 - 1/p = 1/q$ yields

$$\begin{aligned}
E[XY] = E[Y^q]E_Q[XY^{1-q}] &= E[Y^q]\varphi(E_Q[XY^{1-q}])^{1/p} \leq E[Y^q]E_Q\left[\varphi(XY^{1-q})\right]^{1/p} \\
&= E[Y^q]E_Q\left[X^p Y^{p(1-q)}\right]^{1/p} = E[Y^q]E\left[X^p Y^{p(1-q)} Y^q / E[Y^q]\right]^{1/p} \\
&= E[X^p]^{1/p} E[Y^q]^{1-1/p} = E[X^p]^{1/p} E[Y^q]^{1/q}
\end{aligned}$$

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 $|x+y|^p \leq 1/2\, |2x|^p + 1/2\, |2y|^p = 2^{p-1}\left(|x|^p + |y|^p\right)$. 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

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

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

[24] Verify that Q defined as such is indeed a probability measure, that is $Q(A) = \int_A dQ = \int_A Y^q/E[Y^q]dP = E[1_A Y^q/E[Y^q]]$ is a σ additive measure and it holds $E_Q[Z] = E[ZY^q/E[Y^q]]$.

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 1.50. *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 an 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. □

Definition 1.51. 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 1.52. *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 1.53 (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.[25] □

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

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 1.54. *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 1.55 (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*

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

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

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

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

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

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.[26] 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. □

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

$$E[X; A] := E[X 1_A]$$

[26] Check it as an exercise!

Theorem 1.56. *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*

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

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.[27] It follows from Radon-Nikodym Theorem 1.55 that there exists two P-almost surely unique positive random variables $Z^{\pm} \in L^1(\mathcal{G})$ such that

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

Defining $E\left[X|\mathcal{G}\right] = Z^+ - Z^- \in L^1(G)$ 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 1.57. Under the assumptions of the Theorem 1.47, show that for a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, if $\varphi(X)$ is integrable, then it holds

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

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

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

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

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

[27] Verify that these are indeed measures!

Hence

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

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.

1.5. Uniform Integrability

We finish this subsection with some results about uniform integrability. Note that for $X\in L^1$, Lebesgues dominated convergence implies that $E[|X|\,;|X|\geq n]\to 0$. Uniform integrability is a similar requirement but on a whole set of random variables.

Definition 1.58. A set $H\subseteq L^1$ is called uniformly integrable if

$$\sup_{X\in H} E\left[|X|\,;|X|\geq n\right]\to 0$$

Proposition 1.59. *For $H\subseteq L^1$, the following assertions are equivalent*

(i) H is uniformly integrable;

(ii) the following two assertions holds

- *H is bounded in L^1, that is $\sup_{X\in H} E[|X|]<\infty$;*
- *For every $\varepsilon>0$ there exists $\delta>0$ such that*

$$E\left[|X|\,;A\right]\leq\varepsilon$$

for all $X\in H$ and $A\in\mathcal{F}$ such that $P[A]\leq\delta$.

(iii) There exists a Borel measurable function $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ such that $\varphi(x)/x\to\infty$ as $x\to\infty$ for which holds

$$\sup_{X\in H} E\left[\varphi(|X|)\right]<\infty.$$

Proof. Suppose that *(i)* holds. It follows that for n large enough we have $E[|X|\,;|X|\geq n]\leq 1$ for all $X\in H$. Hence $E[|X|]\leq n+1$ for all $X\in H$ showing that H is bounded in L^1. Let further $\varepsilon>0$ and choose n large enough such that $E[|X|\,;|X|\geq n]\leq\varepsilon/2$. Seting $\delta=\varepsilon/(2n)$, for every $A\in\mathcal{F}$ such that $P[A]\leq\delta$, it follows that

$$E\left[|X|\,;A\right]=E\left[|X|\,;A\cap\{|X|\geq n\}\right]+E\left[|X|\,;A\cap\{|X|<n\}\right]\leq nP[A]+\varepsilon/2\leq\varepsilon,$$

showing that *(i)* implies *(ii)*.
Reciprocally, suppose that *(i)* holds. Denote by $M=\sup_X E[|X|]<\infty$, and let $\varepsilon>0$. There exists $\delta>0$ such that $E[X;A]\leq\varepsilon$ for every $A\in\mathcal{F}$ with $P[A]\leq\delta$. Choose then n greater than M/δ. For $X\in H$, Markov inequality yields

$$P\left[|X|\geq n\right]\leq\frac{E[|X|]}{n}\leq\frac{M}{n}\leq\delta$$

Hence

$$\sup_{X\in H} E\left[|X|\,;|X|\geq n\right]\leq\varepsilon$$

showing the uniform integrability of H.

Suppose that *(iii)* holds and denote by $M = \sup_{X\in H} E[\varphi(X)]$. For $\varepsilon > 0$, there exists n_ε such that $\varphi(x) \geq Mx/\varepsilon$ for every $x \geq n_\varepsilon$. Hence

$$M \geq \sup_{X\in H} E\left[\varphi(|X|)\right] \geq \sup_{X\in H} E\left[\varphi(|X|); |X| \geq n_\varepsilon\right] \geq M \sup_{X\in H} E\left[|X|\,; |X| \geq n_\varepsilon\right]/\varepsilon$$

showing that $\sup_n \sup_{X\in H} E\left[|X|\,; |X| \geq n\right] \leq \sup_{X\in H} E\left[|X|\,; |X| \geq n_\varepsilon\right] \leq \varepsilon$ and so the uniform integrability of H.
Reciprocally assume *(i)* and choose a sequence (c_n) which can always be chosen increasing, such that $\sup_{X\in H} E[|X|\,; |X| \geq c_n] \leq 1/n^3$. Define the function $\varphi : \mathbb{R}_+$ as a piecewise linear, equal to 0 on $[0, c_1]$ and the derivative equal to n on $[c_n, c_{n+1}]$ which implies that $\varphi(x)/x \to \infty$ as $x \to \infty$. It follows that

$$E[\varphi(|X|)] = \sum E[\varphi(|X|); c_n \leq |X| \leq c_{n+1}] = \sum n\left(E[|X| \wedge c_{n+1}] - E[|X| \wedge c_n]\right)$$

However,

$$\begin{aligned} &E[|X| \wedge c_{n+1}] - E[|X| \wedge c_n] \\ &\qquad = E[|X|\,; c_n \leq |X| < c_{n+1}] + E[c_{n+1}; |X| \geq c_{n+1}] - E[c_n; |X| \geq c_n] \\ &\qquad\qquad \leq E[|X|\,; |X| \geq c_n] + E[|X|\,; |X| \geq c_{n+1}] \leq 2/n^3 \end{aligned}$$

which shows that $\sup_{X\in H} E[\varphi(|X|)] \leq \sum 2n/n^3 < \infty$. □

Theorem 1.60. *Let $(X_n) \subseteq L^1$ be a sequence of random variables such that X_n converges in probability to a random variable X.*[28] *Then, the following assertions are equivalent*

(i) *the sequence is uniformly integrable;*[29]

(ii) *X_n converges to X in L^1.*

(iii) *$\|X_n\|_1$ converges to $\|X\|_1$*

Proof. We show that (i) implies (ii). By Proposition 1.52, there exists a subsequence (Y_n) of (X_n) that converges P-almost surely to X. In particular, (Y_n) is uniformly integrable. Using Fatou and the L^1 boundedness of the family (X_n), see Proposition 1.59, it follows that $E[|Y|] \leq \liminf E[|Y|_n] \leq \sup_n E[|Y|_n] < \infty$ showing that $X \in L^1$. It follows that the sequence $(X_n - X)$ is uniformly integrable and therefore without loss of generality we can assume that (X_n) is a uniform integrable family converging in probability to 0. For $\varepsilon > 0$ it holds

$$E\left[|X_n|\right] = E\left[|X_n|\,; |X_n| \leq \varepsilon/2\right] + E\left[|X_n|\,; |X_n| > \varepsilon/2\right] \leq \varepsilon/2 + E\left[|X_n|\,; |X_n| > \varepsilon/2\right]$$

By uniform integrability of the family (X_n), making use of Proposition 1.59, let $\delta > 0$ such that $\sup_n E[|X_n|\,; A] \leq \varepsilon/2$ for every $A \in \mathcal{F}$ with $P[A] \leq \delta$. Further, by convergence of (X_n) in probability to 0, there exists n_0 such that $P[|X_n| > \varepsilon/2] \leq \delta$ for every $n \geq n_0$. Thus, for every $n \geq n_0$, it holds $E[|X_n|] \leq \varepsilon/2 + \sup_{k\geq n_0} E[|X_n|\,; |X_n| > \varepsilon/2] \leq \varepsilon$ showing that X_n converges in L^1 to 0.
The fact that (ii) implies (iii) is trivial from $||x| - |y|| \leq |x - y|$, and therefore we finish the proof by showing that (iii) implies (i). For $M > 0$, define φ_M as being the identity on $[0, M-1]$, 0 on $[M, \infty[$ and linearly interpolated on the remaining part of the real line. Let $\varepsilon > 0$ and using the dominated convergence theorem, choose M such that $E[|X|] - E[\varphi_M(|X|)] \leq \varepsilon/2$ since $\varphi_M(|X|)$ converges to and is dominated by $|X| \in L^1$. By continuity of φ_M, it follows that $\varphi_M(|X_n|) \to \varphi_M(|X|)$ also

[28] That is $P[|X_n - X| \geq \varepsilon] \to 0$ for every ε.
[29] That is $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.

in probability. Now, since $\varphi_M(|X_n|) \leq M$ for every n, the dominated convergence theorem in its convergence in probability fashion, see Proposition 1.52 yields $E[\varphi(|X_n|)] \to E[\varphi_M(|X|)]$. Hence, together with $E[|X_n|] \to E[|X|]$, there exists some integer n_0 such that

$$E[|X_n|] - E[|X|] \leq \varepsilon/4 \quad \text{and} \quad E[\varphi_M(|X|)] - E[\varphi(|X_n|)] \leq \varepsilon/4$$

for every $n \geq \varepsilon$. Henceforth

$$E[|X_n|\,; |X_n| \geq M] \leq E[|X_n|] - E[\varphi_M(|X_n|)] \leq \varepsilon/2 + E[|X|] - E[\varphi_M(|X|)] \leq \varepsilon$$

for every $n \geq n_0$. Increases the value of M so that this inequality remains true for the remaining $n \geq n_0$, to conclude the uniform integrability of (X_n). $\square$