MA400. September Introductory Course
(Financial Mathematics, Risk & Stochastics)
P1. Probability Spaces

Preliminary considerations

1. A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ that can be described informally as follows:

   - Ω is the sample space. We can think of Ω as the set of all possible outcomes in "nature" or in a "random experiment" that we want to model. In this context, "nature" chooses exactly one point $\omega \in \Omega$, but we do not know which one, otherwise, we would have no uncertainty and we would know exactly what is going to happen.
   - $\mathcal{F}$ is a collection of event of interests. An event is a subset of Ω, so $\mathcal{F}$ is a set of subsets of Ω. We can think of $\mathcal{F}$ as all the *information* that "nature" has or all the *information* that can be associated with the modelling of a "random experiment".
   - $\mathbb{P}$ is a function that assigns a probability $\mathbb{P}(A)$ to each event $A \in \mathcal{F}$. In particular, given an event $A \in \mathcal{F}$, $\mathbb{P}(A)$ is a number in the interval $[0, 1]$ that represents our belief on how likely the event A is to occur.

   Mathematically, a probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ such that

   - Ω is a set,
   - $\mathcal{F}$ is a σ-algebra on Ω (see Definition 10 below), and
   - $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ (see Definition 29 below).

2. *Example.* Consider tossing a coin that lands heads with probability $p \in (0,1)$ twice. In this context, we can choose the sample space, which is the set of all possible outcomes, to be the set $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, where, e.g.,

$$\begin{aligned}
&\omega_1 \text{ identifies with observing heads first and then tails,}\\
&\omega_2 \text{ identifies with observing heads first and then heads,}\\
&\omega_3 \text{ identifies with observing tails first and then tails, and}\\
&\omega_4 \text{ identifies with observing tails first and then heads.}
\end{aligned}$$

The family of all events of interest that can arise in this random experiment is the set

$$\begin{aligned}
\mathcal{F} = \big\{&\Omega, \emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\},\\
&\{\omega_2, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}\big\}.
\end{aligned}$$

In fact, the elements of this set $\mathcal{F}$ have a simple description in everyday language. For instance, $\{\omega_1, \omega_2\}$ is the event that we observe heads in the first toss, $\{\omega_1\}$ is the event that the coin lands heads first and then tails, $\{\omega_1, \omega_3, \omega_4\}$ is the event that either we observe tails in the second toss or we observe tails first and then heads, Ω is the event that we observe something, and $\emptyset$ is the event that we observe nothing.

Based on everyday intuition, we can assign a probability $\mathbb{P}(A)$ to each event $A \in \mathcal{F}$ in a consistent way, so that, e.g., $\mathbb{P}\big(\{\omega_3, \omega_4\}\big) = 1 - p$, while $\mathbb{P}\big(\{\omega_1\}\big) = p(1-p)$.

3. Plainly, the probabilistic aspects of tossing a coin twice can be described with more simplicity than in 1 above. In this example, the family of all events of interest $\mathcal{F}$ is the set of all subsets of Ω, so, its introduction may seem unnecessary. Indeed, just the sample space Ω and the probability measure $\mathbb{P}$ assigning the appropriate value in $[0,1]$ to each subset of Ω provide a sufficient probabilistic model.

   However, this is not the case when one addresses the probabilistic modelling of more complex random situations arising in several areas, including finance. To appreciate this claim, we consider the following example.

4. *Example.* Consider drawing a number from the interval $(0,1)$ in a completely random way. In this case, we can identify the sample space Ω with $(0,1)$, and every subset A of Ω is an event: A identifies with the event that the number that we draw happens to be in the set A. Given any $a, b \in (0,1)$ such that $a < b$, intuition suggests that the probability of the event (a,b) is $b-a$, because the number that we draw is equally likely to be anywhere in $(0,1)$. In light of this simple observation, any event (i.e., subset of Ω) should have probability equal to its "length".

   The question that thus arises is: can we assign a length to every subset of $(0,1)$? The answer is no: in Example 7 below, we construct a subset of $(0,1)$ to which it is not possible to assign a length. As a result, we cannot assign a probability to every subset of $\Omega \equiv (0,1)$. To develop a meaningful theory, we therefore need to restrict our attention to those subsets of Ω that do have a well-defined length.

This example illustrates why we need to consider families $\mathcal{F}$ of events of "interest" (in the context of the example, such families should include only events that do have a length). Is this a serious restriction? Not really: it turns out that we can always choose an appropriate collection $\mathcal{F}$ of events of "interest" that contains every event of practical interest.

5. It is worth keeping in mind that there are many different probability spaces: we have seen two different ones in Examples 2 and 4 above. The probabilistic modelling of different random situations may require probability spaces of different structures. Indeed, the probability space that we considered in Example 2 is not sufficiently rich to model the random drawing of a number from the interval $(0, 1)$. On the other hand, a probability space that is appropriate for the example discussed in Example 4 could be viewed as too complex to use for the modelling of tossing a coin twice. In this context, the abstract nature of the probability theory we consider is one of its most appealing features.

   We should also note that we can have many different probability measures $\mathbb{P}$ defined on the same pair $(\Omega, \mathcal{F})$. For example, each value in $(0, 1)$ of the parameter p appearing in Example 2 is associated with a different probability measure. In fact, different consistent belief systems concerning the same random situation correspond to different probability measures. For instance, two financial agents might consistently assign different probabilities to the same market events (e.g., the two agents might be a bull and a bear).

6. It is extremely important to note that the probability theory we consider is concerned with *countably* many operations. "Finite" is too restrictive relative to real life applications. Indeed, every property that holds in the limit involves infinite considerations and / or quantities. On the other hand, we should always keep in mind that the theory we consider in this course is incompatible with uncountably many operations.

A subset of $(0,1)$ to which we can assign no length

7. *Example*. Suppose that we can assign a length to every subset of the real line, and denote by $L(A)$ the length of the set $A \subseteq \mathbb{R}$, so that, e.g.,

$$L\big((a,b)\big) = b - a, \quad L\big(\{a\}\big) = 0 \quad \text{and} \quad L\big((-\infty, a)\big) = L\big((a, \infty)\big) = \infty \tag{1}$$

for all real numbers $a < b$. Intuition suggests that the length function L should be positive, i.e., $L(A) \geq 0$ for all $A \subseteq \mathbb{R}$, increasing in the sense that, given any sets $A, B \subseteq \mathbb{R}$,

$$A \subseteq B \quad \Rightarrow \quad L(A) \leq L(B), \tag{2}$$

and countably additive, so that, if (A_n) is a sequence of pairwise disjoint subsets of $\mathbb{R}$, i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$L\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} L(A_n). \tag{3}$$

Also, the length of a set should be translation invariant, so that

$$L(A^a) = L(A) \quad \text{for all } A \subseteq \mathbb{R} \text{ and } a \in \mathbb{R}, \tag{4}$$

where A^a is the translation of A by a, which is defined by $A^a = \{a + x \mid \ x \in A\}$.

Now, we consider the equivalence relation $\sim$ on the real line defined by

$$x \sim y \quad \text{if} \quad x - y \in \mathbb{Q},$$

and split the interval $(0,1)$ in equivalence classes. In this context, the numbers $x, y \in (0,1)$ belong to the same equivalence class if and only if $x \sim y$, i.e., if and only if $x - y \in \mathbb{Q}$, while, if the numbers $x, y \in (0,1)$ belong to different equivalence classes, then $x \not\sim y$, i.e., $x - y \notin \mathbb{Q}$. Also, the equivalence classes are pairwise disjoint, so each number in $(0,1)$ belongs to exactly one equivalence class.

By appealing to the axiom of choice, we next consider a set C that contains exactly one representative from each equivalence class. Since C contains only one point from each equivalence class, any distinct points $x, y \in C$ belong to different equivalence classes, so $x \not\sim y$. Furthermore, given any point $z \in (0,1)$, if x is the representative in C of the equivalence class in which z belongs, then $z \sim x$, so there exists $q \in \mathbb{Q}$ such that $z = q + x$. In view of these observations, it follows that, if we define

$$C^q = \{q + x \mid \; x \in C\}, \quad \text{for } q \in (-1,1) \cap \mathbb{Q},$$

then

$$C^{q_1} \cap C^{q_2} = \emptyset \quad \text{for all } q_1 \neq q_2, \tag{5}$$

and

$$(0,1) \subseteq \bigcup_{q \in (-1,1) \cap \mathbb{Q}} C^q \subseteq (-1,2). \tag{6}$$

Now, we argue by contradiction to conclude that the set C has no length. If $L(C) = 0$, then

$$1 \overset{(1)}{=} L\big((0,1)\big) \overset{(2),(6)}{\leq} L\left(\bigcup_{q \in (-1,1) \cap \mathbb{Q}} C^q\right) \overset{(3),(5)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C^q) \overset{(4)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C) = 0,$$

which is not possible. So, if $L(C)$ exists, we must have $L(C) > 0$ because L is a positive function. In this case, we can see that

$$3 \overset{(1)}{=} L\big((-1,2)\big) \overset{(2),(6)}{\geq} L\left(\bigcup_{q \in (-1,1) \cap \mathbb{Q}} C^q\right) \overset{(3),(5)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C^q) \overset{(4)}{=} \sum_{q \in (-1,1) \cap \mathbb{Q}} L(C) = \infty,$$

which cannot be true. We conclude that $L(C)$ does not exist.

π-systems, algebras and σ-algebras

8. *Definition.* A collection $\mathcal{I}$ of subsets of Ω is a *π-system* if it is stable under finite intersections, i.e.,

$$A_1, A_2 \in \mathcal{I} \quad \Rightarrow \quad A_1 \cap A_2 \in \mathcal{I}.$$

9. *Definition.* An *algebra* on Ω is a collection $\mathcal{A}$ of subsets of Ω such that

$$\begin{aligned} &\text{(i) } \Omega \in \mathcal{A}, \\ &\text{(ii) } A \in \mathcal{A} \quad \Rightarrow \quad A^c \equiv \Omega \setminus A \in \mathcal{A}, \\ &\text{(iii) } A_1, A_2 \in \mathcal{A} \quad \Rightarrow \quad A_1 \cup A_2 \in \mathcal{A}. \end{aligned}$$

10. *Definition.* A *σ-algebra* on Ω is a collection $\mathcal{F}$ of subsets of Ω such that

$$\begin{aligned} &\text{(i) } \Omega \in \mathcal{F}, \\ &\text{(ii) } A \in \mathcal{F} \quad \Rightarrow \quad A^c \equiv \Omega \setminus A \in \mathcal{F}, \\ &\text{(iii) } A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F} \quad \Rightarrow \quad \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}. \end{aligned}$$

11. A simple induction argument reveals that:

   - If $\mathcal{I}$ is a π-system, then

$$A_1, A_2, \ldots, A_n \in \mathcal{I} \quad \Rightarrow \quad \bigcap_{j=1}^{n} A_j \in \mathcal{I}.$$

   - If $\mathcal{A}$ is an algebra, then

$$A_1, A_2, \ldots, A_n \in \mathcal{A} \quad \Rightarrow \quad \bigcup_{j=1}^{n} A_j \in \mathcal{A}.$$

12. *Lemma.* Given a set Ω and a σ-algebra $\mathcal{F}$ on Ω,

$$\emptyset \in \mathcal{F}, \tag{7}$$

$$A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F} \quad \Rightarrow \quad \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}. \tag{8}$$

In particular, *a σ-algebra is stable under countable set operations.*

Similar properties are true for an algebra $\mathcal{A}$ on Ω. In particular, *an algebra is stable under finite set operations.*

Proof. Fix any set Ω and any σ-algebra $\mathcal{F}$ on Ω. Since $\emptyset = \Omega^c$, (7) follows immediately by properties (i) and (ii) of Definition 10. To prove (8), we consider any sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$, and we observe that

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c .$$

The event appearing on the right hand side of this expression belongs to $\mathcal{F}$ because

$$\begin{aligned} A_n \in \mathcal{F} \text{ for all } n \geq 1 \quad &\Rightarrow \quad A_n^c \in \mathcal{F} \text{ for all } n \geq 1 && \text{(by property 10.(ii))} \\ &\Rightarrow \quad \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F} && \text{(by property 10.(iii))} \\ &\Rightarrow \quad \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F} && \text{(by property 10.(ii))} \end{aligned}$$

and (8) follows.

13. *Example*. The *power set* $\mathcal{P}(\Omega)$ of any set Ω, namely, the collection of all subsets of Ω, is a π-system, as well as an algebra on Ω, as well as a σ-algebra on Ω.

    This should be obvious!

14. *Example.* Suppose that $\Omega = \mathbb{R}$. We can easily verify that the family of sets

$$\mathcal{I} = \big\{ (-\infty, a] \mid \ a \in \mathbb{R} \big\}$$

is a π-system. However, $\mathcal{I}$ is neither an algebra nor a σ-algebra because, e.g.,

$$(-\infty, a]^c \equiv \mathbb{R} \setminus (-\infty, a] = (a, \infty) \notin \mathcal{I}.$$

15. *Example.* Suppose that $\Omega = \mathbb{R}$. We can easily verify that the family of sets

$$\mathcal{I} = \big\{ (a, b) \mid \ a, b \in \mathbb{R}, \ a \leq b \big\},$$

where we adopt the convention that $(a, a) = \emptyset$, is a π-system.

16. Every σ-algebra is an algebra. However, an algebra is *not necessarily* a σ-algebra.

    Example. Suppose that $\Omega = [0, \infty)$, and let

$$\mathcal{A} = \big\{ A \subseteq \Omega \mid \ \text{either } A \text{ or } A^c \text{ is a finite set} \big\}.$$

Here, we adopt the convention that the empty set is “finite”.

We can verify that $\mathcal{A}$ is an algebra by checking the properties (i), (ii) and (iii) of Definition 9 as follows:

(i) $\Omega \in \mathcal{A}$ because $\Omega = \emptyset^c$ and we have assumed that the empty set $\emptyset$ is finite.

(ii) By the definition of $\mathcal{A}$, given any event $A \in \mathcal{A}$, we also have $A^c \in \mathcal{A}$.

(iii) Given events $A_1, A_2 \in \mathcal{A}$, we can see that $A_1 \cup A_2 \in \mathcal{A}$ by considering the following two possibilities:

- Each of the sets A_1 and A_2 is finite. In this case, the set $A_1 \cup A_2$ is also finite, and so, $A_1 \cup A_2 \in \mathcal{A}$.
- At least one of the sets A_1, A_2, say A_1, is infinite. In this case, we observe that

$$(A_1 \cup A_2)^c = A_1^c \cap A_2^c \subseteq A_1^c.$$

  Since A_1^c is finite (because $A_1 \in \mathcal{A}$ and A_1 is infinite), it follows that $(A_1 \cup A_2)^c$ is finite, and so, $A_1 \cup A_2 \in \mathcal{A}$.

On the other hand, $\mathcal{A}$ is *not* a σ-algebra. To see this, consider, e.g., the sequence of events defined by

$$A_n = \{n\}, \quad n = 0, 1, 2, \ldots.$$

Clearly, $A_n \in \mathcal{A}$ for all $n \geq 0$. However,

$$\bigcup_{n=1}^{\infty} A_n = \mathbb{N} \notin \mathcal{A},$$

because neither $\mathbb{N}$ nor $\mathbb{N}^c = [0, \infty) \setminus \mathbb{N}$ is finite. This shows that property (iii) of Definition 10 is not necessarily satisfied.

Generating $\boldsymbol{\sigma}$-algebras

17. *Lemma.* Let $\{\mathcal{F}_i,\ i \in I\}$ be a family of σ-algebras on a set Ω indexed by a set $I \neq \emptyset$. The collection $\bigcap_{i\in I} \mathcal{F}_i$ is a σ-algebra on Ω.

    Proof. We have to check the defining properties of a σ-algebra. To this end, we note that the family of events $\bigcap_{i\in I} \mathcal{F}_i$ satisfies property (iii) of Definition 10 because

    $$\begin{aligned} A_1, A_2, \ldots, A_n, \ldots \in \bigcap_{i\in I} \mathcal{F}_i & \\ \Rightarrow \quad & A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}_i \text{ for all } i \in I \\ \Rightarrow \quad & \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_i \text{ for all } i \in I \quad \text{(because each } \mathcal{F}_i \text{ is a } \sigma\text{-algebra)} \\ \Rightarrow \quad & \bigcup_{n=1}^{\infty} A_n \in \bigcap_{i\in I} \mathcal{F}_i. \end{aligned}$$

    Similarly, we can verify properties (i) and (ii) of Definition 10 [*Exercise*].

18. Given two σ-algebras $\mathcal{F}$ and $\mathcal{G}$, the collection of events $\mathcal{F} \cup \mathcal{G}$ is *not necessarily* a σ-algebra. To see this, it suffices to consider an example such as the following.

    Example. Suppose that $\Omega = \{1, 2, 3, 4\}$, and let

    $$\mathcal{F} = \big\{\Omega, \emptyset, \{1, 2\}, \{3, 4\}\big\},$$
    $$\mathcal{G} = \big\{\Omega, \emptyset, \{1\}, \{2, 3, 4\}\big\}.$$

    Then

    $$\mathcal{F} \cup \mathcal{G} = \big\{\Omega, \emptyset, \{1, 2\}, \{3, 4\}, \{1\}, \{2, 3, 4\}\big\}$$

    is *not* a σ-algebra. To see this, consider the events $\{3, 4\}$ and $\{1\}$, which both belong to $\mathcal{F} \cup \mathcal{G}$, and observe that

    $$\{3, 4\} \cup \{1\} = \{1, 3, 4\} \notin \mathcal{F} \cup \mathcal{G}.$$

19. *Definition.* Given a collection $\mathcal{C}$ of subsets of Ω, the σ-algebra $\sigma(\mathcal{C})$ on Ω *generated by* $\mathcal{C}$ is the *smallest* σ-algebra on Ω containing $\mathcal{C}$: it is the intersection of all σ-algebras on Ω which have $\mathcal{C}$ as a subclass.

20. Observe that, if $\mathcal{C}$ is a family of sets and $\mathcal{H}$ is a σ-algebra, then

    $$\mathcal{C} \subseteq \mathcal{H} \quad \Rightarrow \quad \sigma\big(\mathcal{C}\big) \subseteq \sigma\big(\mathcal{H}\big) = \mathcal{H},$$

    because, by definition, $\sigma(\mathcal{C})$ is the intersection of all σ-algebras containing $\mathcal{C}$. In other words, $\sigma(\mathcal{C})$ is a subset of *any* σ-algebra containing $\mathcal{C}$.

21. In Definition 19, the σ-algebra $\sigma(\mathcal{C})$ generated by a given family of events $\mathcal{C}$ is "constructed" from "above" or from "outside". It should be noted that attempting to construct $\sigma(\mathcal{C})$ from "inside" by repeating seqentialy countable operations may fail! Indeed, constructions from "inside" are in general possible only if the family $\mathcal{C}$ is "small", e.g., if it has countable elements (see the following two examples).

22. *Example*. Given a subset A of Ω, the smallest σ-algebra containing A is $\{\Omega, \emptyset, A, A^c\}$.

23. *Example*. Suppose that $\Omega = \{1, 2, 3, 4\}$, and let

$$\mathcal{C} = \big\{\{1\}, \{1, 3, 4\}\big\}.$$

Then

$$\sigma(\mathcal{C}) = \big\{\Omega, \emptyset, \{1\}, \{1, 3, 4\}, \{2, 3, 4\}, \{2\}, \{1, 2\}, \{3, 4\}\big\}.$$

24. *Definition*. The *Borel* σ-algebra $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ is the σ-algebra on $\mathbb{R}$ generated by the family of all open intervals (a, b), i.e.,

$$\mathcal{B}(\mathbb{R}) = \sigma\big(\{(a, b) \mid \ a, b \in \mathbb{R},\ a < b\}\big).$$

More generally, consider any topological space S. The *Borel* σ-algebra $\mathcal{B}(S)$ on S is the σ-algebra on S generated by the family of all open sets, i.e.,

$$\mathcal{B}(S) = \sigma\big(\{A \subset S \mid \ A \text{ is open}\}\big).$$

Borel σ-algebras are very important: they contain every event of practical interest!

25. *Example*. $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I})$, where $\mathcal{I}$ is the π-system of Example 14.

Proof. In view of the Definition 24 of the Borel σ-algebra on $\mathbb{R}$ and the observation in Paragraph 20 above, we can prove this claim as follows.

(i) $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{I})$ will follow if we show that $(a, b) \in \sigma(\mathcal{I})$ for all real numbers $a < b$. This is true because

$$\begin{aligned}(a, b) &= (a, \infty) \cap (-\infty, b) \\ &= (-\infty, a]^c \cap \bigcup_{n=1}^{\infty} (-\infty, b - \tfrac{1}{n}]\,.\end{aligned}$$

(ii) $\mathcal{B}(\mathbb{R}) \supseteq \sigma(\mathcal{I})$ will follow if we show that $(-\infty, a] \in \mathcal{B}(\mathbb{R})$ for every real number a. This follows from the observation that

$$\begin{aligned}(-\infty, a] &= \bigcup_{m=1}^{\infty} (a - m, a] \\ &= \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} (a - m, a + \tfrac{1}{n})\,.\end{aligned}$$

Measures

26. *Definition.* A pair $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F}$ is a σ-algebra on Ω, is called *measurable space*.

27. *Definition.* Let $(S, \mathcal{S})$ be a measurable space, so that $\mathcal{S}$ is a σ-algebra on the set S. A *measure* defined on $(S, \mathcal{S})$ is a function $\mu : \mathcal{S} \to [0, \infty]$ that is *countably additive*, i.e., it is such that

    (i) $\mu(\emptyset) = 0$, and
    (ii) if $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{S}$ is any sequence of pairwise disjoint sets (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then
    $$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

    The triplet $(S, \mathcal{S}, \mu)$ is then called a *measure space*.

28. *Definition.* Given a measure space $(S, \mathcal{S}, \mu)$, we say that

    μ is a *probability* measure if $\mu(S) = 1$,
    μ is a *finite* measure if $\mu(S) < \infty$, and
    μ is a *σ-finite* measure if there is a sequence $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{S}$ such that
    $$\mu(A_n) < \infty \text{ for all } n \geq 1 \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = S.$$

    In this course, we will consider *only* σ-finite measures.

29. Due to its particular interest, we repeat the definition of a probability measure:

    Definition. A *probability measure* defined on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

    (i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, and
    (ii) if $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$ is any sequence of pairwise disjoint events (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then
    $$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

30. *Lemma.* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then,

$$\text{if } A \subseteq B, \text{ then } \mathbb{P}(A) \leq \mathbb{P}(B), \tag{9}$$

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \tag{10}$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \tag{11}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{12}$$

[*Exercise*: How do these properties change if we consider a more general measure space $(S, \mathcal{S}, \mu)$?]

Proof. If $A \subseteq B$, then

$$\mathbb{P}(B) = \mathbb{P}\big(A \cup (B \setminus A)\big) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

Also, (10) follows immediately from the calculations

$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(S) = 1.$$

Given any events A and B, if we define

$$K = A \cap B^c, \qquad L = A \cap B, \qquad M = A^c \cap B,$$

then K, L, M are pairwise disjoint,

$$A = K \cup L \quad \text{and} \quad B = L \cup M.$$

Therefore,

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(K \cup L \cup M) \\
&= \mathbb{P}(K) + \mathbb{P}(L) + \mathbb{P}(M) \\
&= \mathbb{P}(K) + \mathbb{P}(L) + \mathbb{P}(M) + \mathbb{P}(L) - \mathbb{P}(L) \\
&= \mathbb{P}(K \cup L) + \mathbb{P}(M \cup L) - \mathbb{P}(L) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),
\end{aligned}$$

which proves (11).

In view of (11) and the positivity of probabilities,

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

Using this inequality and a straightforward induction argument, we can then obtain inequality (12) [*Exercise*].

31. *Lemma* (*"Continuity" of a measure*). Let $(S, \mathcal{S}, \mu)$ be a measure space. Given an *increasing sequence* $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ of events in $\mathcal{S}$, we can define the *limit* of the sequence by

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

In this context,

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{13}$$

Similarly, if $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ is a *decreasing* sequence of events in $\mathcal{S}$, the *limit* of the sequence is defined by

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

In this case, $\underline{\text{if } \mu(A_1) < \infty}$, then

$$\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n). \tag{14}$$

Proof. Given an increasing sequence $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ of events in $\mathcal{S}$, let $B_1 = A_1$, and define recursively $B_n = A_n \setminus A_{n-1}$, for $n \geq 2$. By construction, the events $B_1, B_2, \ldots, B_n, \ldots$ are pairwise disjoint,

$$A_n = \bigcup_{k=1}^{n} B_k \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = \bigcup_{k=1}^{\infty} B_k.$$

As a consequence,

$$\begin{aligned}
\mu\left(\lim_{n\to\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \\
&= \mu\left(\bigcup_{k=1}^{\infty} B_k\right) \\
&= \sum_{k=1}^{\infty} \mu(B_k) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_k\right) \\
&= \lim_{n\to\infty} \mu(A_n).
\end{aligned}$$

Consider any decreasing sequence $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ of events in $\mathcal{S}$ such that $\mu(A_1) < \infty$. Since $\emptyset \subseteq A_1 \setminus A_2 \subseteq \cdots \subseteq A_1 \setminus A_n \subseteq \cdots$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_1 \setminus A_n\right) = \lim_{n\to\infty} \mu(A_1 \setminus A_n).$$

Noting that

$$\bigcup_{n=1}^{\infty} A_1 \setminus A_n = A_1 \setminus \bigcap_{n=1}^{\infty} A_n,$$

we can see that this implies that

$$\mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \bigl[\mu(A_1) - \mu(A_n)\bigr],$$

which establishes (14).

[*Exercise*: Where is the assumption $\mu(A_1) < \infty$ needed in the proof of (14) above?]

32. In the previous result, the validity of (14) relies heavily on the assumption $\mu(A_1) < \infty$ (in fact, on the assumption that $\mu(A_k) < \infty$, for some $k \geq 1$). To appreciate this claim, we consider the following example.

    Example. Suppose that $S = \mathbb{R}$, $\mathcal{S} = \mathcal{B}(\mathbb{R})$ and $\mu = L$, where L is the Lebesgue measure that maps each set $C \in \mathcal{B}(\mathbb{R})$ to its length $L(C)$ (see also Paragraph 36 below). If we define $A_n = [n, \infty)$, for $n \geq 1$, then we can see that

$$\mu\left(\lim_{n\to\infty} A_n\right) = L\left(\bigcap_{n=1}^{\infty} [n,\infty)\right) = L(\emptyset) = 0 < \infty = \lim_{n\to\infty} L\bigl([n,\infty)\bigr) = \lim_{n\to\infty} \mu(A_n).$$

Construction and uniqueness of measures

33. Although σ-algebras often do not include every possible subset of the underlying sample spaces, they can be sufficiently "rich" to accommodate all events of practical interest. As a consequence, σ-algebras can be awkward to handle when proving mathematical results because they might be "too big". It is for this reason that we consider "coarser" classes of events such as π-systems and algebras.

34. *Theorem* (*Uniqueness of extension*). Suppose that $\mathcal{I}$ is a π-system on a set Ω and let $\mathcal{F} = \sigma(\mathcal{I})$. Also, let μ_1 and μ_2 be two measures on $(\Omega, \mathcal{F})$ such that

$$\mu_1(A) = \mu_2(A) \quad \text{for all } A \in \mathcal{I}$$

and suppose that there exists a sequence $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{I}$ such that

$$\mu_1(A_n) = \mu_2(A_n) < \infty \text{ for all } n \geq 1 \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = \Omega.$$

Then

$$\mu_1(A) = \mu_2(A) \quad \text{for all } A \in \mathcal{F}.$$

Informally, *if two measures agree and are σ-finite on a π-system, then they agree on the σ-algebra generated by that π-system.*

35. *Theorem* (*Carathéodory's extension theorem*). Suppose that $\mathcal{A}$ is an algebra on a set Ω and let $\mathcal{F} = \sigma(\mathcal{A})$. Also, suppose that a function $\mu_0 : \mathcal{A} \to [0, \infty]$ is countably additive in the sense that

(i) $\mu_0(\emptyset) = 0$, and

(ii) if $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A}$ is any sequence of pairwise disjoint sets such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, then

$$\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n).$$

Then there exists a measure μ on $(\Omega, \mathcal{F})$ such that

$$\mu(A) = \mu_0(A) \quad \text{for all } A \in \mathcal{A}.$$

36. *The Lebesgue measure.* The Lebesgue measure maps sets $C \subseteq \mathbb{R}$ to which we can assign a length to their length $L(C)$ (recall Examples 4 and 7).

Let $\mathcal{A}$ be the collection of sets $C \subseteq \mathbb{R}$ that admit the representation

$$C = (a_1, b_1] \cup \cdots \cup (a_k, b_k] \tag{15}$$

or

$$C = (-\infty, b_0] \cup (a_1, b_1] \cup \cdots \cup (a_k, b_k] \tag{16}$$

or

$$C = (a_1, b_1] \cup \cdots \cup (a_k, b_k] \cup (a_{k+1}, \infty) \tag{17}$$

or

$$C = (-\infty, b_0] \cup (a_1, b_1] \cup \cdots \cup (a_k, b_k] \cup (a_{k+1}, \infty), \tag{18}$$

for some $k \geq 1$ and $-\infty < b_0 < a_1 < b_1 < \cdots < a_k < b_k < a_{k+1} < \infty$, together with $\mathbb{R}$ and $\emptyset$. We can check that this collection is an algebra on $\mathbb{R}$ and that $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ [*Exercise*].

We define the function $L_0 : \mathcal{A} \to [0, \infty]$ by associating each set $C \in \mathcal{A}$ with its length, so that

$$L_0(C) = \infty \quad \text{if } C \text{ is as in (16), (17) or (16)}$$

and

$$L_0(C) = \sum_{j=1}^{k} b_j - a_j \quad \text{if } C \text{ is as in (15).}$$

It can be shown that L_0 is countably additive. By appealing to the the Carathéodory's extension Theorem 35 and the uniqueness Theorem 34 (with the π-system of Example 15), we can then see that L_0 has a unique extension to a measure L on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which is called *Lebesgue measure*.

Given an interval (α, β), where $-\infty \leq \alpha < \beta \leq \infty$, we can follow the same steps to construct the Lebesgue measure on $\big((\alpha, \beta), \mathcal{B}\big((\alpha, \beta)\big)\big)$.