

Measure theory

Kristian Wichmann

August 27, 2016

1 Prerequisites

1.1 Infimum and supremum

For a subset of the reals $A \subseteq \mathbb{R}$, and *upper bound* is a number U that is larger than or equal to all elements of A :

$$U \text{ is an upper bound on } A \Leftrightarrow \forall a \in A : U \geq a \quad (1.1)$$

Similarly, L is a *lower bound* on A if it is smaller than or equal to all elements of A :

$$L \text{ is a lower bound on } A \Leftrightarrow \forall a \in A : L \leq a \quad (1.2)$$

Definition 1.1. *The infimum of $A \subseteq \mathbb{R}$ (if it exists) is the lowest upper bound on A . Similarly, the supremum (if it exists) is the highest lower bound of A .*

It is a fundamental property of the real numbers that a lower (upper) bound subset $A \subseteq \mathbb{R}$ has an infimum (supremum). There's no "holes" in the real numbers.

1.2 The extended reals \mathbb{R}^*

By the extended reals, we mean the usual reals \mathbb{R} as well as $\pm\infty$:

$$\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\} \quad (1.3)$$

Any non-empty subset of \mathbb{R}^* has an infimum and a supremum. This follows from the usual properties of bounded sets of reals.

Definition 1.2. For a sequence $a_1, a_2, \dots \in \mathbb{R}^*$ we define limes inferior and limes superior by:

$$\liminf_{n \rightarrow \infty} a_n = \inf_{p \geq 1} \sup\{a_n | n \geq p\} \quad (1.4)$$

$$\limsup_{n \rightarrow \infty} a_n = \sup_{p \geq 1} \inf\{a_n | n \geq p\} \quad (1.5)$$

$$(1.6)$$

In other words, limes inferior is the infimum of the non-increasing sequence $\lambda_p = \sup\{a_n | n \geq p\}$. Similarly, limes superior is the supremum of the non-decreasing sequence $\delta_p = \inf\{a_n | n \geq p\}$.

Theorem 1.1. A sequence $a_1, a_2, \dots \in \mathbb{R}^*$ is convergent if and only if:

$$\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n = a \quad (1.7)$$

The limit is then equal to a .

Proof. Let's split into the two cases of the limit being finite and non-finite.

- Finite limit, \Rightarrow : The series is convergent to $a \in \mathbb{R}$, which means that:

$$\forall \varepsilon > 0 \exists p \forall n \geq p : |a_n - a| < \varepsilon \quad (1.8)$$

So $\{a_n | n \geq p\}$ takes values in the interval $(a - \varepsilon, a + \varepsilon)$. This means that $a + \varepsilon$ is an upper bound and that no upper bound can be below $a - \varepsilon$. In other words $a + \varepsilon \geq \lambda_p \geq a - \varepsilon$. Since this is true for all $\varepsilon > 0$ this means that limes inferior must be equal to a . A similar argument shows that limes superior must also be a .

- Finite limit, \Leftarrow : Here:

$$\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n = a \in \mathbb{R} \quad (1.9)$$

This means that $\forall p : \lambda_p \geq a$ and $\forall p : \delta_p \leq a$. Now, let an $\varepsilon > 0$ be given. There must now be a p' where $\lambda_{p'} - a < \varepsilon$ and another p'' where $a - \delta_{p''} < \varepsilon$. Set $p = \max\{p', p''\}$. Then by monotonicity of the λ 's and δ 's we have:

$$\lambda_p - a < \varepsilon, \quad a - \delta_p < \varepsilon \quad (1.10)$$

Now, since $\lambda_p = \sup\{a_n | n \geq p\}$ we have: $\forall n \geq p : a_n \leq \lambda_p$. This is equivalent to $a_n - a \leq \lambda_p - a < \varepsilon$. Similarly, since $\delta_p = \inf\{a_n | n \geq p\}$ we have: $\forall n \geq p : a_n \geq \delta_p$. Which is equivalent to $a_n - a \geq \delta_p - a \Leftrightarrow a - a_n \leq a - \delta_p$. So $a - a_n \leq \varepsilon$. All in all we have $|a_n - a| < \varepsilon$, which shows that $\lim_{n \rightarrow \infty} a_n = a$.

- Infinite limit, \Rightarrow : Assume the series converges to ∞ . This means that:

$$\forall r \in \mathbb{R} \exists p \forall n \geq p : a_n > r \quad (1.11)$$

Since $\{a_n | n \geq p\}$ is not upper bound, the only upper bound is ∞ . So $\lambda_n = \infty$, which means limes inferior must be ∞ . A lower bound for $\{a_n | n \geq p\}$ is r , so $\delta_p \geq r$. Since this is true for all r , limes superior must be ∞ . Similarly for convergence to $-\infty$.

- Infinite limit, \Leftarrow : Assume that:

$$\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n = \infty \quad (1.12)$$

Since limes inferior is infinite, this means that the δ 's get arbitrarily large. In other words:

$$\forall r \in \mathbb{R} \exists p \forall n \geq p : \delta_n > r \quad (1.13)$$

But since δ_n is an infimum we must have $a_r \geq \delta_r$, so $a_n > r$, which shows the desired convergence. For convergence to $-\infty$ a similar argument can be made using limes superior.

□

Summation in \mathbb{R}^* is extended through:

$$a + \infty = \infty + a = \infty, \quad a \neq -\infty \quad (1.14)$$

$$a + (-\infty) = -\infty + a = -\infty, \quad a \neq \infty \quad (1.15)$$

We leave $\infty + (-\infty)$ and $-\infty + \infty$ undefined. Furthermore, we extend multiplication in \mathbb{R}^* as:

$$0 \cdot (\pm\infty) = \pm\infty \cdot 0 = 0 \quad (1.16)$$

$$c \cdot (\pm\infty) = (\pm\infty) \cdot c = \pm\infty, \quad 0 < c \leq \infty \quad (1.17)$$

$$c \cdot (\pm\infty) = (\pm\infty) \cdot c = \mp\infty, \quad -\infty < c \leq 0 \quad (1.18)$$

Often we will work in $\mathbb{R}_+^* = [0, \infty]$.

1.3 Infinite sums

We now consider an infinite sum of elements of \mathbb{R}_+^* :

$$\sum_{j \in J} a_j, \quad a_j \in \mathbb{R}_+^* \quad (1.19)$$

The index set J can be of arbitrary cardinality. We now set:

Definition 1.3.

$$\sum_{j \in J} a_j = \sup \left\{ \sum_{i \in I} a_i \mid I \subseteq J, I \text{ finite} \right\} \quad (1.20)$$

If $J = \emptyset$ we set the sum equal to 0.

The following theorem for suprema in \mathbb{R}_+^* will come in handy:

Theorem 1.2. *Let $a \in \mathbb{R}_+^*$ and $A, B \subseteq \mathbb{R}_+^*$. Then:*

$$(a) \sup(a + B) = a + \sup B \quad (1.21)$$

$$(b) \sup(a \cdot B) = a \cdot \sup B \quad (1.22)$$

$$(c) \sup(A + B) = \sup A + \sup B \quad (1.23)$$

$$(d) \sup(A \cdot B) = \sup A \cdot \sup B \quad (1.24)$$

Here $a + B = \{a + b \mid b \in B\}$, $a \cdot B = \{a \cdot b \mid b \in B\}$, $A + B = \{a + b \mid a \in A, b \in B\}$ and $A \cdot B = \{a \cdot b \mid a \in A, b \in B\}$.

Proof. There's four parts to be shown:

- (a): If U is an upper bound on B , then $U \geq \sup B$. Now $U' = a + U$ is an upper bound on $a + B$. So for all upper bounds on $a + B$ we have $U' \geq a + \sup B$, and so is the minimum upper bound, which is the supremum is $a + \sup B$.
- (b): Similar to (a), but here the bounds are multiplied by a instead of added.
- (c): Let U_A and U_B be upper bounds on A and B respectively. This means that $U_A \geq \sup A$ and $U_B \geq \sup B$. So any upper bound of $(A + B)$ can be written $U' = U_A + U_B \geq \sup A + \sup B$. This shows that $\sup(A + B) = \sup A + \sup B$.
- (d): Similar to (c) but with multiplication instead of addition.

□

We can now prove the following:

Theorem 1.3. *Let $(a_j)_{j \in J}$ be elements of \mathbb{R}_+^* . And let $J_1, J_2 \subseteq J$ so that $J_1 \cup J_2 = J$ and $J_1 \cap J_2 = \emptyset$. Then:*

$$\sum_{j \in J} a_j = \sum_{j_1 \in J_1} a_{j_1} + \sum_{j_2 \in J_2} a_{j_2} \quad (1.25)$$

Proof. By definition, the sums are equal to suprema over finite subsets of the respective families:

$$\sum_{j \in J} a_j = \sup \left\{ \sum_{i \in I} a_i \mid I \subseteq J, I \text{ finite} \right\} \quad (1.26)$$

$$\sum_{j_1 \in J_1} a_j = \sup \left\{ \sum_{i_1 \in I_1} a_{i_1} \mid I_1 \subseteq J_1, I_1 \text{ finite} \right\} \quad (1.27)$$

$$\sum_{j_2 \in J_2} a_j = \sup \left\{ \sum_{i_2 \in I_2} a_{i_2} \mid I_2 \subseteq J_2, I_2 \text{ finite} \right\} \quad (1.28)$$

Since the right hand side of equation (1.25) is now a sum of suprema over two sets, we can use theorem (1.2), (c) to write it as:

$$\sup \left\{ \sum_{i_1 \in I_1} a_{i_1} + \sum_{i_2 \in I_2} a_{i_2} \mid I_1 \subseteq J_1, I_2 \subseteq J_2; I_1, I_2 \text{ finite} \right\} \quad (1.29)$$

But this set is just the same as finite sums from J , so the two sides of equation (1.25) are equal. \square

By induction this generalizes to:

Theorem 1.4. *Let $(a_j)_{j \in J}$ be elements of \mathbb{R}_+^* . And let $J_1, J_2, \dots, J_k \subseteq J$ so that $\bigcup_{j=1}^k J_j = J$ and $J_i \cap J_j = \emptyset$ when $i \neq j$. Then:*

$$\sum_{j \in J} a_j = \sum_{i=1}^k \sum_{j_i \in J_i} a_{j_i} \quad (1.30)$$

In fact, it turns out that we may generalize to an arbitrary family of subsets:

Theorem 1.5. *Let $(a_j)_{j \in J}$ be elements of \mathbb{R}_+^* . And let $J_i \subseteq J, i \in K$ so that $\bigcup_{k \in K} J_k = J$ and $J_i \cap J_j = \emptyset$ when $i \neq j$. Then:*

$$\sum_{j \in J} a_j = \sum_{k \in K} \sum_{j_k \in J_k} a_{j_k} \quad (1.31)$$

Proof. Now, the sum over K is again defined as a supremum. For brevity, set $S_k = \sum_{j_k \in J_k} a_{j_k}$. The right hand side is now:

$$\sum_{k \in K} S_k = \sup \left\{ \sum_{k' \in K'} S_{k'} \mid K' \subseteq K, K' \text{ finite} \right\} \quad (1.32)$$

Since each S_k is itself a supremum over a finite subset, the set on the right hand side is in fact just the set of finite sums from J itself, and therefore equation (1.31) must be true. \square

One special case of theorem 1.5 will be particularly useful:

Theorem 1.6. *Let $a_{ij} \in \mathbb{R}_+^*$ for $i, j \in \mathbb{N}$. Then:*

$$\sum_{ij} a_{ij} = \sum_i \sum_j a_{ij} = \sum_j \sum_i a_{ij} \quad (1.33)$$

Finally, consider the following:

Theorem 1.7. *Let $(a_j)_{j \in J}$ be elements of \mathbb{R}_+^* , so that:*

$$\sum_{j \in J} a_j = s < \infty \quad (1.34)$$

Then $a_j \neq 0$ for at most countably many j .

Proof. For $n \in \mathbb{N}$, consider the number of $a_j \geq \frac{1}{n}$. Since the sum is s , there can be at most sn (rounded down) of such elements, i.e. a finite amount. The amount of non-zero elements is clearly bounded by counting these for all n . But this is at most countably many, since \mathbb{N} is countable. \square

2 Sigma algebras

Definition 2.1. *An algebra \mathbb{E} over a set X is a non-empty collection of subsets of X such that:*

$$(i) A \in \mathbb{E} \Rightarrow A^c \in \mathbb{E} \quad (2.1)$$

$$(ii) A, B \in \mathbb{E} \Rightarrow A \cup B \in \mathbb{E} \quad (2.2)$$

In other words, \mathbb{E} is closed under complements and finite unions.

For any X we may always form the *trivial algebra*, which is simply $\mathbb{E} = \mathcal{P}(X)$, the set of all subsets of X .

Theorem 2.1. *Given an algebra \mathbb{E} over X , $X \in \mathbb{E}$ and $\emptyset \in \mathbb{E}$.*

Proof. Since an algebra is non-empty, there exists an $A \in \mathbb{E}$. By (i) $A^c \in \mathbb{E}$. By (ii) $A \cup A^c = X \in \mathbb{E}$. By (i) $X^c = \emptyset \in \mathbb{E}$. \square

In fact, for any X , the set $\{\emptyset, X\}$ is an algebra. If $A \subseteq X$, so is the set $\{\emptyset, A, A^c, X\}$.

Theorem 2.2. *An algebra \mathbb{E} over X is closed under finite intersections:*

$$A, B \in \mathbb{E} \Rightarrow A \cap B \in \mathbb{E} \quad (2.3)$$

Proof. Let $A, B \in \mathbb{E}$. By (i) $A^c, B^c \in \mathbb{E}$. By (ii) $A^c \cup B^c \in \mathbb{E}$. By (i) $(A^c \cup B^c)^c = A \cap B \in \mathbb{E}$. Here one of de Morgan's laws have been used. \square

Definition 2.2. *A σ -algebra is an algebra that is closed under countably infinite unions instead of merely finite ones:*

$$A_1, A_2, \dots \in \mathbb{E} \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathbb{E} \quad (2.4)$$

The tuple (X, \mathbb{E}) is collectively known as a measurable space. The elements of \mathbb{E} are called \mathbb{E} -measurable or simply measurable, when there's no chance of confusion.

In general, sigma refers to countable infinity.

Theorem 2.3. *A σ -algebra \mathbb{E} over X is closed under countably infinite intersections:*

$$A_1, A_2, \dots \in \mathbb{E} \Rightarrow \bigcap_{i \in \mathbb{N}} A_i \in \mathbb{E} \quad (2.5)$$

Proof. The proof is completely analogous to the one for theorem 2.2, since de Morgan's laws hold for any collection of sets. \square

Theorem 2.4. *Given a family of σ -algebras $(\mathbb{E}_i)_{i \in I}$ over X , the intersection $\bigcap_{i \in I} \mathbb{E}_i$ is also a σ -algebra.*

Proof. Since all σ -algebras contain \emptyset and X , the intersection is non-empty. Consider $A \in \bigcap_{i \in I} \mathbb{E}_i$. Then A must be an element in all \mathbb{E}_i . Therefore A^c is also an element in all \mathbb{E}_i , and thus an element of $\bigcap_{i \in I} \mathbb{E}_i$. Similarly, if $A_n \in \bigcap_{i \in I} \mathbb{E}_i$ for $n \in \mathbb{N}$, then $A_n \in \mathbb{E}_i$, and therefore $\bigcup_{n \in \mathbb{N}} A_n \in \mathbb{E}_i$ for all $i \in I$. So $\bigcup_{n \in \mathbb{N}} A_n \in \bigcap_{i \in I} \mathbb{E}_i$. \square

We can use theorem (2.4) to make the following useful definition:

Definition 2.3. *Let \mathbb{D} be a subset of a σ -algebra \mathbb{E} . We then define the σ -algebra generated by \mathbb{D} as:*

$$\sigma(\mathbb{D}) = \bigcap \{ \mathbb{G} | \mathbb{D} \subseteq \mathbb{G} \subseteq \mathbb{E}, \mathbb{G} \text{ is a } \sigma\text{-algebra} \} \quad (2.6)$$

Note: The set to be intersected is non-empty, as $\mathcal{P}(\Omega)$ always is a σ -algebra over Ω . $\sigma(\mathbb{D})$ can be thought of as the smallest σ -algebra containing \mathbb{D} .

Definition 2.4. *If there is a finite \mathbb{D} such that $\sigma(\mathbb{D}) = \mathbb{E}$, \mathbb{E} is said to be finitely generated.*

2.1 The Borel algebra

For any topological space, the open sets induce a σ -algebra known as the *Borel algebra*. The sets in this algebra are known as *Borel sets*. Our concern will mainly be the cases where the space is simply \mathbb{R} or \mathbb{R}^n .

Definition 2.5. *The σ -algebra generated by the open subsets of \mathbb{R} is called the Borel algebra \mathbb{B} . Similarly, the σ -algebra generated by the open subsets of \mathbb{R}^n is denoted \mathbb{B}_n .*

From this definition immediately follows:

Theorem 2.5. *In \mathbb{R}^n , all closed, finite and countable sets are Borel sets.*

Proof. Since all open sets are Borel sets, so are their complements, which are exactly the closed sets. All finite sets are closed and so are Borel sets. Specifically all singletons are Borel sets. Any countable set can be written as a countable union of singletons, and so are also a Borel set. \square

The open intervals $(a, b) \subseteq \mathbb{R}$ are Borel sets. It turns out that closed and half-closed intervals are also Borel sets:

Theorem 2.6. *For $a, b \in \mathbb{R}$, the intervals $(a, b], [a, b), [a, b] \in \mathbb{B}$.*

Proof. For the first interval, consider the following countable family of open intervals:

$$I_n = \left(a, b + \frac{1}{n} \right) \quad (2.7)$$

The intersection $\bigcap_{n \in \mathbb{N}} I_n = (a, b]$ and so is a Borel set. For the other intervals, consider the families $I_n = (a - \frac{1}{n}, b)$ and $I_n = (a - \frac{1}{n}, b + \frac{1}{n})$ respectively. \square

The Borel algebra(s) has a lot of different generator systems. Systems of intervals being some of the most important.

Definition 2.6. *An box I in \mathbb{R}^n is a cartesian product of ordinary intervals I_1, I_2, \dots, I_n :*

$$I = I_1 \times I_2 \times \dots \times I_n \quad (2.8)$$

The intervals I_1, I_2, \dots, I_n may be bounded, unbounded, open, closed or half-open individually. However, when all the intervals are half-open of the form $(a_i, b_i]$ - which we will refer to as a standard interval - we will refer to the box as a standard box. If $b_i - a_i$ is also the same for all I_i , we will refer to the resulting standard box I as a standard cube.

It turns out that the collection of standard cubes generates the Borel algebra(s):

Theorem 2.7. *Let $G \subseteq \mathbb{R}^n$ be an open set. Then G can be written as a union of countably many standard cubes.*

Proof. Start by considering what we will call the 0th order mesh:

$$M_0 = \{I_1 \times I_2 \times \cdots \times I_n | I_i = (z_i, z_i + 1], z_i \in \mathbb{Z}\} \quad (2.9)$$

All the countably many elements of M_0 are clearly disjoint standard cubes of side length 1, and $\bigcup M_0 = \mathbb{R}^n$. Now, pick out the elements of M_0 that are fully contained within G . Next, consider the 1st order mesh:

$$M_1 = \left\{ I_1 \times I_2 \times \cdots \times I_n | I_i = \left(\frac{1}{2}z_i, \frac{1}{2}(z_i + 1) \right], z_i \in \mathbb{Z} \right\} \quad (2.10)$$

Again, the countably many elements of M_1 are disjoint standard cubes with side length $\frac{1}{2}$, and $\bigcup M_1 = \mathbb{R}^n$. Now pick out the elements of M_1 that are fully contained in G minus the elements we picked from M_0 . In general, we can define the p 'th mesh as:

$$M_p = \{I_1 \times I_2 \times \cdots \times I_n | I_i = (2^{-p}z_i, 2^{-p}(z_i + 1)], z_i \in \mathbb{Z}\} \quad (2.11)$$

M_p consists of countably many disjoint standard cubes of length 2^{-p} and $\bigcup M_p = \mathbb{R}^n$. For each new mesh, we pick out the cubes that are fully contained within what "remains" of G . We can continue this process to the limit $p \rightarrow \infty$. Because G is open, each point in G is an inner point, and we will in fact get all of G this way. \square

We can now state a number of sets that generate \mathbb{B} :

Theorem 2.8. *The following sets generate the Borel algebra \mathbb{B} :*

$$(i) \{(a, b] | a, b \in \mathbb{R}, a < b\} \quad (2.12)$$

$$(ii) \{(a, b) | a, b \in \mathbb{R}, a < b\} \quad (2.13)$$

$$(iii) \{[a, b] | a, b \in \mathbb{R}, a < b\} \quad (2.14)$$

$$(iv) \{[a, b) | a, b \in \mathbb{R}, a < b\} \quad (2.15)$$

$$(v) \{(a, \infty) | a \in \mathbb{R}\} \quad (2.16)$$

$$(vi) \{[a, \infty) | a \in \mathbb{R}\} \quad (2.17)$$

$$(vii) \{(-\infty, a) | a \in \mathbb{R}\} \quad (2.18)$$

$$(viii) \{(-\infty, a] | a \in \mathbb{R}\} \quad (2.19)$$

Proof. There's eight sets to prove for. The proofs will be somewhat similar:

- (i): This follows directly from theorem 2.7.

- (ii): Consider the family of sets $I_n = (a, b + \frac{1}{n})$. Now $\bigcap_{n \in \mathbb{N}} I_n = (a, b]$ (if $b - a < 1$, start from the smallest n where $\frac{1}{n} < b - a$ - use similar strategies below). The result now follows from theorem 2.7.
- (iii): Consider the family of sets $I_n = [a + \frac{1}{n}, b]$. Now $\bigcup_{n \in \mathbb{N}} I_n = (a, b]$. The result now follows from theorem 2.7.
- (iv): Consider the family of sets $I_n = [a, b + \frac{1}{n})$. Now $\bigcap_{n \in \mathbb{N}} I_n = [a, b]$. The result now follows from (iii).
- (v): Let $a, b \in \mathbb{R}, a < b$. By complement $(-\infty, b]$ is generated. So is the intersection $(a, \infty) \cap (-\infty, b] = (a, b]$. The result now follows from (i).
- (vi): Let $a, b \in \mathbb{R}, a < b$. By complement $(-\infty, b)$ is generated. So is the intersection $[a, \infty) \cap (-\infty, b) = [a, b)$. The result now follows from (iv).
- (vii): Let $a, b \in \mathbb{R}, a < b$. By complement $[a, \infty)$ is generated. So is the intersection $(-\infty, b) \cap [a, \infty) = [a, b)$. The result now follows from (iv).
- (viii): Let $a, b \in \mathbb{R}, a < b$. By complement $(a, \infty]$ is generated. So is the intersection $(-\infty, b] \cap (a, \infty] = (a, b]$. The result now follows from (i).

□

2.2 The Borel algebra on \mathbb{R}^*

We will also construct a similar σ -algebra on \mathbb{R}^* . To do so, we need some way to specify open sets. Here, we will use a metric.

Definition 2.7. \mathbb{B}^* , the Borel algebra on \mathbb{R}^* is defined as induced by the following metric:

$$d(x, y) = |\text{Arctan}(x) - \text{Arctan}(y)| \quad (2.20)$$

Here we set $\text{Arctan}(\pm\infty) = \pm\frac{\pi}{2}$.

Theorem 2.9. Let \mathbb{B}^* as defined above. Then:

$$\mathbb{B}^* = \sigma(\{(a, \infty] \mid a \in \mathbb{R}\}) \quad (2.21)$$

Furthermore, for $A \subseteq \mathbb{R}^*$:

$$A \in \mathbb{B}^* \Leftrightarrow A \cap \mathbb{R} \in \mathbb{B} \quad (2.22)$$

Proof. The open neighbourhoods around $\pm\infty$ in this metric are of the form $(a, \infty]$, $a \neq \infty$ and $[-\infty, a)$, $a \neq -\infty$, respectively. If none of these points are in a set, it is open exactly when it's open in the usual metric. These sets are generated by the open intervals (a, b) , $a, b \in \mathbb{R}$. So, if these three types of intervals are generated we get all of \mathbb{B}^* .

- $(a, \infty]$, $a \neq \infty$: All of these are in the generating set pr. definition, except for $a = -\infty$. However, this is just all of \mathbb{R}^* which is always generated.
- $[-\infty, a)$, $a \neq -\infty$: Since $(a, \infty]$ is in the set of generators (except when $a = \infty$, but this gives all of \mathbb{R}^* like above), so is $(a, \infty]^c = [-\infty, a]$. But then, so is $[-\infty, a - \frac{1}{n}]$, $n \in \mathbb{N}$. The countable union over $n \in \mathbb{N}$ is $[-\infty, a)$.
- (a, b) , $a, b \in \mathbb{R}$: According to the items above, $(a, \infty]$ and $[-\infty, b)$ are both generated. But then, so is $(a, \infty] \cap [-\infty, b) = (a, b)$.

Now for the second statement: According to theorem (2.8) (v) \mathbb{B} is generated by $\{(a, \infty) | a \in \mathbb{R}\}$. This shows that the two σ -algebras \mathbb{B} and \mathbb{B}^* contain the same sets when disregarding the two points $\pm\infty$. This shows the " \Rightarrow " part. As for " \Leftarrow ": By taking the countable intersections $[-\infty, -n)$ and $(n, \infty]$ we see that that singletons $\{\pm\infty\}$ are both Borel sets for \mathbb{R}^* . If a set is a Borel set for \mathbb{R} , possibly adding a singleton or two will still yield another Borel set for \mathbb{R}^* . \square

3 Measurable functions

Definition 3.1. Given two measurable spaces (X, \mathbb{E}) and (Y, \mathbb{F}) , a function $\varphi : X \rightarrow Y$ is called $\mathbb{E} - \mathbb{F}$ -measurable (or simply measurable) if:

$$F \in \mathbb{F} \Rightarrow \varphi^{-1}(F) \in \mathbb{E} \quad (3.1)$$

Here $\varphi^{-1}(F)$ is the *pre-image* of the set F . Note the parallel to the definition of continuity of maps between topological space. In particular, this immediately implies that a continuous function is $\mathbb{B}_X - \mathbb{B}_Y$ -measurable.

The following properties of pre-images will come in handy in a moment:

Theorem 3.1. Let $\varphi : A \rightarrow B$ and $\psi : B \rightarrow C$. Then the following is true of the pre-images φ^{-1} and ψ^{-1} :

$$(i) \ F \subseteq B \Rightarrow \varphi^{-1}(F^c) = (\varphi^{-1}(F))^c \quad (3.2)$$

$$(ii) \ F_i \subseteq B, i \in \mathbb{N} \Rightarrow \varphi^{-1} \left(\bigcup_{i \in \mathbb{N}} F_i \right) = \bigcup_{i \in \mathbb{N}} \varphi^{-1}(F_i) \quad (3.3)$$

$$(iii) \ G \subseteq C \Rightarrow \varphi^{-1}(\psi^{-1}(G)) = (\varphi \circ \psi)^{-1}(G) \quad (3.4)$$

Proof. There's three parts:

- (i): Assume $F \subseteq Y$. Then $\varphi^{-1}(F) = \{a \in A | \varphi(a) \in F\}$. Start from the right:

$$(\varphi^{-1}(F))^c = A \setminus \varphi^{-1}(F) = A \setminus \{a \in A | \varphi(a) \in F\} = \quad (3.5)$$

$$\{a \in A | \varphi(a) \notin F\} = \varphi^{-1}(F^c) \quad (3.6)$$

- (ii): Assume $F_i \subseteq Y$. Then $\varphi^{-1}(F_i) = \{a \in A | \varphi(a) \in F_i\}$. Start from the right:

$$\bigcup_{i \in \mathbb{N}} \varphi^{-1}(F_i) = \bigcup_{i \in \mathbb{N}} \{a \in A | \varphi(a) \in F_i\} = \quad (3.7)$$

$$\left\{ a \in A | \varphi(a) \in \bigcup_{i \in \mathbb{N}} F_i \right\} = \varphi^{-1} \left(\bigcup_{i \in \mathbb{N}} F_i \right) \quad (3.8)$$

- (iii): We have $\psi^{-1}(G) = \{b \in B | \psi(b) \in G\}$. Now consider:

$$\varphi^{-1}(\psi^{-1}(G)) = \varphi^{-1}(\{b \in B | \psi(b) \in G\}) = \quad (3.9)$$

$$\{a \in A | \varphi(a) \in \{b \in B | \psi(b) \in G\}\} = \{a \in A | \varphi(a) \in \varphi(\psi^{-1}(G))\} \quad (3.10)$$

This is exactly the pre-image $(\varphi \circ \psi)^{-1}(G)$.

□

Luckily, it turns out, that to show measurability of a function, it is enough to show equation (3.1) for a set that generates \mathbb{F} :

Theorem 3.2. Let (X, \mathbb{E}) and (Y, \mathbb{F}) be measurable spaces. Let $\mathbb{D} \subseteq \mathbb{F}$ such that $\sigma(\mathbb{D}) = \mathbb{F}$. Now consider a function $\varphi : X \rightarrow Y$. φ is then $\mathbb{E} - \mathbb{F}$ -measurable if and only if:

$$D \in \mathbb{D} \Rightarrow \varphi^{-1}(D) \in \mathbb{E} \quad (3.11)$$

Proof. 'Only if' follows directly from the definition of measurability. For 'if' define the following subset of \mathbb{F} :

$$\tilde{\mathbb{F}} = \{F \in \mathbb{F} \mid \varphi^{-1}(F) \in \mathbb{E}\} \quad (3.12)$$

If we can show $\tilde{\mathbb{F}} = \mathbb{F}$, we're done. It turns out that $\tilde{\mathbb{F}}$ is a σ -algebra:

- Since $\varphi^{-1}(Y) = X \in \mathbb{E}$, $\tilde{\mathbb{F}}$ is non-empty, as $Y \in \tilde{\mathbb{F}}$.
- Let $F \in \tilde{\mathbb{F}}$, i.e. $\varphi^{-1}(F) \in \mathbb{E}$ and therefore $(\varphi^{-1}(F))^c \in \mathbb{E}$. By theorem 3.1 (i): $\varphi^{-1}(F^c) = (\varphi^{-1}(F))^c$, so $F^c \in \tilde{\mathbb{F}}$.
- Let $F_i \in \tilde{\mathbb{F}}, i \in \mathbb{N}$, i.e. $\varphi^{-1}(F_i) \in \mathbb{E}$ and therefore $\bigcup_{i \in \mathbb{N}} \varphi^{-1}(F_i) \in \mathbb{E}$. By theorem 3.1 (ii): $\varphi^{-1}(\bigcup_{i \in \mathbb{N}} F_i) = \bigcup_{i \in \mathbb{N}} \varphi^{-1}(F_i)$, so $\bigcup_{i \in \mathbb{N}} F_i \in \tilde{\mathbb{F}}$.

$\mathbb{D} \subseteq \tilde{\mathbb{F}}$ by assumption, but since $\tilde{\mathbb{F}}$ is a σ -algebra and \mathbb{D} generates \mathbb{F} , we must have $\tilde{\mathbb{F}} \subseteq \mathbb{F}$. So we must have $\tilde{\mathbb{F}} = \mathbb{F}$. \square

This means that we immediately get a sufficient condition for measurability of a function $f : X \rightarrow \mathbb{R}$ for each set of generators of \mathbb{B} we have found. Similarly for functions into \mathbb{R}^* and \mathbb{B}^* . For instance, theorem 2.8 gives us eight such sufficient conditions:

Theorem 3.3. *Let (X, \mathbb{E}) be a measurable space, and associate the Borel algebra with the real numbers to form the measurable space (\mathbb{R}, \mathbb{B}) . Then the following condition are all equivalent to the function $f : X \rightarrow \mathbb{R}$ being $\mathbb{E} - \mathbb{B}$ -measurable:*

- (i) $\forall a, b \in \mathbb{R}, a < b : \{x \in X \mid a < f(x) \leq b\} \in \mathbb{E}$
- (ii) $\forall a, b \in \mathbb{R}, a < b : \{x \in X \mid a < f(x) < b\} \in \mathbb{E}$
- (iii) $\forall a, b \in \mathbb{R}, a < b : \{x \in X \mid a \leq f(x) \leq b\} \in \mathbb{E}$
- (iv) $\forall a, b \in \mathbb{R}, a < b : \{x \in X \mid a \leq f(x) < b\} \in \mathbb{E}$
- (v) $\forall a \in \mathbb{R} : \{x \in X \mid a < f(x)\} \in \mathbb{E}$
- (vi) $\forall a \in \mathbb{R} : \{x \in X \mid a \leq f(x)\} \in \mathbb{E}$
- (vii) $\forall a \in \mathbb{R} : \{x \in X \mid f(x) < a\} \in \mathbb{E}$
- (viii) $\forall a \in \mathbb{R} : \{x \in X \mid f(x) \leq a\} \in \mathbb{E}$

Proof. All of these follow from theorem 2.8 and theorem 3.2. We'll show (i) explicitly - the rest follow completely analogously.

- (i): Here $\mathbb{D} = \{(a, b] | a, b \in \mathbb{R}, a < b\}$. By equation (3.11) the corresponding criterion is:

$$D \in \{(a, b] | a, b \in \mathbb{R}, a < b\} \Rightarrow f^{-1}(D) \in \mathbb{E} \quad (3.13)$$

But $f^{-1}(D) = \{x \in X | a < f(x) \leq b\}$ and we're done.

□

Similarly for \mathbb{B}^* :

Theorem 3.4. *Let (X, \mathbb{E}) be a measurable space. A function $f : (X, \mathbb{E}) \rightarrow (\mathbb{R}^*, \mathbb{B}^*)$ is measurable if and only if:*

$$\forall a \in \mathbb{R} : \{x \in X | a < f(x)\} \in \mathbb{E} \quad (3.14)$$

Proof. This follows from theorems 2.9 and 3.2. □

Theorem 3.5. *Let (X, \mathbb{E}) , (Y, \mathbb{F}) and (Z, \mathbb{G}) be measurable spaces. Let $\varphi : X \rightarrow Y$ be $\mathbb{E} - \mathbb{F}$ -measurable and $\psi : Y \rightarrow Z$ be $\mathbb{F} - \mathbb{G}$ -measurable. Then $\varphi \circ \psi : X \rightarrow Z$ is $\mathbb{E} - \mathbb{G}$ -measurable.*

Proof. Let $G \in \mathbb{G}$. Since ψ is $\mathbb{F} - \mathbb{G}$ -measurable this means $F = \psi^{-1}(G) \in \mathbb{F}$. Since φ is $\mathbb{E} - \mathbb{F}$ -measurable $E = \varphi^{-1}(F) \in \mathbb{E}$. But $E = \varphi^{-1}(\psi^{-1}(G)) = (\varphi \circ \psi)^{-1}(G)$ by theorem 3.1 (iii), so $\varphi \circ \psi$ is $\mathbb{E} - \mathbb{G}$ -measurable. □

If the target space is \mathbb{R}^n instead of \mathbb{R} it turns out that the question of measurability reduces to measurability of the coordinate functions:

Theorem 3.6. *Consider the function $f : (X, \mathbb{E}) \rightarrow (\mathbb{R}^n, \mathbb{B}_n)$. In vector form: $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$. Then f is $\mathbb{E} - \mathbb{B}_n$ -measurable if and only if all f_i are $\mathbb{E} - \mathbb{B}$ -measurable.*

Proof. Let \mathbb{D} be the system of open half-spaces, i.e:

$$\mathbb{D} = \{(x_1, x_2, \dots, x_n) | x_i > a, i \in \{1, 2, \dots, n\}, a \in \mathbb{R}\} \quad (3.15)$$

Then $\sigma(\mathbb{D})$ must contain sets like:

$$S_i(a', a'') = \{(x_1, x_2, \dots, x_n) | x_i > a'\} \cap \{(x_1, x_2, \dots, x_n) | x_i > a''\}^c = \quad (3.16)$$

$$\{(x_1, x_2, \dots, x_n) | a' < x_i \leq a''\} \quad (3.17)$$

Here $a' < a''$. $S_i(a', a'')$ can be thought of as a slice of the space normal to the i 'th axis. If we intersect one such slice for each axis we get a standard interval - from definition 2.6. So all standard intervals are in $\sigma(\mathbb{D})$ and thus $\sigma(\mathbb{D}) = \mathbb{B}_n$. Now we can use theorem 3.2 to get an equivalent condition to f being measurable:

$$\forall i \in \{1, 2, \dots, n\} \forall a \in \mathbb{R} : \{x \in \mathbb{R} | f_i(x) > a\} \in \mathbb{E} \quad (3.18)$$

But according to theorem 3.3 (v) this simply means that the coordinate function f_1, f_2, \dots, f_n are measurable. □

3.1 Limits of measurable functions

Theorem 3.7. *Let (X, \mathbb{E}) be a non-empty measurable space. Let f_1, f_2, \dots be a series of $\mathbb{E} - \mathbb{B}^*$ -measurable functions, $f_i : (X, \mathbb{E}) \rightarrow (\mathbb{R}^*, \mathbb{B}^*)$. Then the functions (defined in a pointwise sense) $\sup_n f_n, \inf_n f_n, \limsup_n f_n$ and $\liminf_n f_n$ are all $\mathbb{E} - \mathbb{B}^*$ -measurable.*

Proof. Let $f = \sup_n f_n$. That is $\forall x \in X : f(x) = \sup\{f_i(x) | i \in \mathbb{N}\}$. Now consider the inequality:

$$f(x) > a \Leftrightarrow \exists n \in \mathbb{N} : f_n(x) > a \quad (3.19)$$

Here $a \in \mathbb{R}$. Now construct the set of all $x \in X$ which satisfy the inequality:

$$\{x \in X | f(x) > a\} = \{x \in X | \exists n \in \mathbb{N} : f_n(x) > a\} \quad (3.20)$$

But this can also be written as:

$$\bigcup_{n \in \mathbb{N}} \{x \in X | f_n(x) > a\} \quad (3.21)$$

But since f_n is measurable, according to theorem 3.4 this set is an element of \mathbb{E} . By the same theorem, f is also measurable. A similar argument can be made for $f = \inf_n f_n$. Limes superior and limes inferior are both successive applications of infimum and supremum, and so are both measurable as well. \square

It now immediately follows that the pointwise limit of a series of measurable functions is itself measurable:

Theorem 3.8. *Let (X, \mathbb{E}) be a non-empty measurable space. Let f_1, f_2, \dots be a series of $\mathbb{E} - \mathbb{B}^*$ -measurable functions, $f_i : (X, \mathbb{E}) \rightarrow (\mathbb{R}^*, \mathbb{B}^*)$. Then if the function $f : (X, \mathbb{E}) \rightarrow (\mathbb{R}^*, \mathbb{B}^*)$, where $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ exists, then it is $\mathbb{E} - \mathbb{B}^*$ -measurable.*

Proof. If f exists, then by theorem 1.1: $\limsup_n f_n = \liminf_n f_n = f$, which is then measurable by theorem 3.7. \square

3.2 Combinations of measurable functions

Theorem 3.9. *Let (X, \mathbb{E}) be a non-empty measurable space. Let $f, g : (X, \mathbb{E}) \rightarrow (\mathbb{R}^*, \mathbb{B})$ be $\mathbb{E} - \mathbb{B}$ -measurable functions. Then the following functions are also $\mathbb{E} - \mathbb{B}$ -measurable:*

$$|f|, cf, f + g, f - g, f \vee g, f \wedge g, fg \quad (3.22)$$

Here $c \in \mathbb{R}$, $f \vee g(x) = \max(f(x), g(x))$ and $f \wedge g(x) = \min(f(x), g(x))$ respectively.

Proof. All of these function can be made by composites of f or $\varphi : X \rightarrow \mathbb{R}^2, \varphi(x) = (f(x), g(x))$ and continuous functions. By theorem 3.5 these are all measurable. \square

3.2.1 Generating measurable sets using \mathbb{R}

One use of measurable functions and their combinations is to create sets that are elements of the σ -algebra \mathbb{E} . If for instance $f, g : (X, \mathbb{E}) \rightarrow \mathbb{R}$ are measurable, then the following sets are elements of \mathbb{E} :

$$\{x \in X | f(x) < g(x)\}, \{x \in X | f(x) \leq g(x)\}, \{x \in X | f(x) = g(x)\} \quad (3.23)$$

This can be seen by considering the function $g - f$. By theorem 3.9 the function is measurable. But the sets above can be expressed as pre-images of $(g - f) : (g - f)^{-1}((0, \infty)), (g - f)^{-1}([0, \infty))$, and $(g - f)^{-1}(\{0\})$ respectively. Since the sets $(0, \infty), [0, \infty), \{0\} \in \mathbb{B}$, these are all \mathbb{E} -measurable.

3.3 Subspaces

Theorem 3.10. *Let (X, \mathbb{E}) be a measurable space and let $X' \subseteq X$. Then the following set is a σ -algebra over X' :*

$$\mathbb{E}' = \{X' \cap E | E \in \mathbb{E}\} \quad (3.24)$$

Proof. There's three things to show:

- \mathbb{E}' is non-empty since $X' \cap X = X' \in \mathbb{E}'$.
- If $E \in \mathbb{E}$, then $E^c \in \mathbb{E}$. And so $X' \cap E, X' \cap E^c \in \mathbb{E}'$. Remember that complement in \mathbb{E}' is different from in \mathbb{E} . In X' , the complement of $X' \cap E$ is:

$$X' \setminus (X \cap E) = X' \cap (X \cap E)^c = X' \cap (X^c \cup E^c) = \quad (3.25)$$

$$X' \cap (\emptyset \cup E^c) = X' \cap E^c \in \mathbb{E}' \quad (3.26)$$

- If $E_1, E_2, \dots \in \mathbb{E}$, then $\bigcup_{i \in \mathbb{N}} E_i \in \mathbb{E}$. And so $X' \cap E_1, X' \cap E_2, \dots \in \mathbb{E}'$ and $X' \cap \bigcup_{i \in \mathbb{N}} E_i \in \mathbb{E}'$. Consider the union:

$$\bigcup_{i \in \mathbb{N}} (X' \cap E_i) = X' \cap \bigcup_{i \in \mathbb{N}} E_i \quad (3.27)$$

\square

Definition 3.2. The σ -algebra \mathbb{E}' from theorem 3.10 is referred to as the σ -algebra induced by or inherited from X' . The measurable space (X', \mathbb{E}') is called the induced or inherited subspace.

Theorem 3.11. Let (X, \mathbb{E}) be a measurable space, $X' \subseteq X$ and (X', \mathbb{E}') the inherited subspace. Then:

$$X' \in \mathbb{E} \Leftrightarrow \mathbb{E}' \subseteq \mathbb{E} \quad (3.28)$$

Proof. " \Rightarrow " follows from σ -algebras being closed under intersection. " \Leftarrow " is true because $X' \in \mathbb{E}'$. \square

Theorem 3.12. Let (X, \mathbb{E}) and (Y, \mathbb{F}) be measurable spaces, $X' \subseteq X$ and (X', \mathbb{E}') the inherited subspace from X . Let further $\varphi : X \rightarrow Y$ be measurable. Then the following functions are measurable:

- The inclusion map $i : X' \rightarrow X, i(x) = x$
- The restriction $\varphi|_{X'} : X' \rightarrow Y, (\varphi|_{X'})(x) = \varphi(x)$

Proof. Let $E \in \mathbb{E}$, then $i^{-1}(E) = X' \cup E \in \mathbb{E}'$. So i is measurable. $\varphi|_{X'} = \varphi \circ i$, so by theorem 3.5, $\varphi|_{X'}$ is also measurable. \square

Theorem 3.13. Let (X, \mathbb{E}) and (Y, \mathbb{F}) be measurable spaces. Let furthermore $A_1, \dots, A_n \in \mathbb{E}$ be pairwise disjoint sets so that $\bigcup_{i=1}^n A_i = X$. Let further $\varphi_i : A_i \rightarrow Y$, and define:

$$\phi : X \rightarrow Y, \phi(x) = \begin{cases} \varphi_1(x) & \text{for } x \in A_1 \\ \varphi_2(x) & \text{for } x \in A_2 \\ \vdots \\ \varphi_n(x) & \text{for } x \in A_n \end{cases} \quad (3.29)$$

If all the φ_i 's are $\mathbb{E}_{A_i} - \mathbb{F}$ -measurable, then ϕ is measurable.

Proof. Let $F \in \mathbb{F}$. Consider the pre-image:

$$\varphi^{-1}(F) = \bigcup_{i=1}^n (\varphi^{-1}(F) \cap A_i) = \bigcup_{i=1}^n \varphi_i^{-1}(F) \quad (3.30)$$

But since φ_i is measurable, $\varphi_i^{-1}(F) \in \mathbb{E}$ and so is the finite union. \square

3.3.1 Generating measurable sets using \mathbb{R}_+^*

Above we used measurable function to generate measurable subsets using \mathbb{R} . We can do something similar for \mathbb{R}_+^* . Let $f, g : (X, \mathbb{E}) \rightarrow (\mathbb{R}_+^*, \mathbb{B}_+^*)$ be measurable. Here, \mathbb{B}_+^* is the σ -algebra inherited from \mathbb{B}^* under restriction to \mathbb{R}_+^* . Then the following sets are elements of \mathbb{E} :

$$\{x \in X | f(x) < g(x)\}, \{x \in X | f(x) \leq f(x)\}, \{x \in X | f(x) = g(x)\} \quad (3.31)$$

We can't use the same strategy as we did above. The trouble is, that $g - f$ is not well-defined when the functions can take on values both $-\infty$ and ∞ . Instead, define the four following set:

$$F = \{x \in X | f(x) < \infty\} = f^{-1}((0, \infty)) \quad (3.32)$$

$$F_\infty = \{x \in X | f(x) = \infty\} = f^{-1}(\{\infty\}) \quad (3.33)$$

$$G = \{x \in X | g(x) < \infty\} = g^{-1}((0, \infty)) \quad (3.34)$$

$$G_\infty = \{x \in X | g(x) = \infty\} = g^{-1}(\{\infty\}) \quad (3.35)$$

These are all \mathbb{E} -measurable, since $(0, \infty)$ and $\{\infty\}$ are measurable. Now, consider the restriction of \mathbb{E} to the set $F \cap G$. On this set, $g - f$ is well-defined since we don't run into any infinities. So:

$$(g - f)^{-1}((0, \infty)) = \{x \in F \cap G | f(x) < g(x)\} \in \mathbb{E}_{F \cap G} \subseteq \mathbb{E} \quad (3.36)$$

Now, we only need the $x \in A$ where $f(x) \neq \infty$ and $g(x) = \infty$. This is done by forming a union with $F \cap G_\infty$. The result is still an element of \mathbb{E} . The rest of the sets can be constructed in similar ways.

4 Measures

A *measure* is a function that assigns length, area, volume or other weight to the elements of a σ -algebra.

Definition 4.1. For a σ -algebra \mathbb{E} over X , a *measure* is a function $\mu : \mathbb{E} \rightarrow \mathbb{R}_+^*$ which satisfies:

$$(i) \mu(\emptyset) = 0 \quad (4.1)$$

$$(ii) \mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i), \quad A_i \text{'s pairwise disjoint} \quad (4.2)$$

The tuple (X, \mathbb{E}, μ) is collectively known as a *measure space*. Sometimes referred to as (X, μ) as a shorthand. Further, we define:

- $\mu(A)$ is called the measure of $A \in \mathbb{E}$.
- If $\mu(X) < \infty$, the measure μ is called finite.
- If $\mu(X) = 1$, the measure μ is sometimes referred to as a probability measure or distribution. In this case, it is customary to use the designations (Ω, \mathcal{F}, P) instead of (X, \mathbb{E}, μ) .

Theorem 4.1. Let (X, \mathbb{E}, μ) be a measure space. Let any set named E or E_i below be an element of (X, \mathbb{E}, μ) . Then:

- (i): $\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$ when the A_i 's are pairwise disjoint.
- (ii): $E_1 \subseteq E_2 \Rightarrow \mu(E_1) \leq \mu(E_2)$.
- (iii): $E_1 \subseteq E_2, \mu(E_1) < \infty: \mu(E_2 \setminus E_1) = \mu(E_2) - \mu(E_1)$.
- (iv): $\mu(\bigcup_{i \in \mathbb{N}} E_i) \leq \sum_{i \in \mathbb{N}} \mu(E_i)$
- (v): $\mu(\bigcup_{i=1}^n E_i) \leq \sum_{i=1}^n \mu(E_i)$
- (vi): When $E_1 \subseteq E_2 \subseteq E_3 \subseteq \dots: \mu(E_i) \nearrow \mu(\bigcup_{i \in \mathbb{N}} E_i)$
- (vii): When $E_1 \supseteq E_2 \supseteq E_3 \supseteq \dots, \mu(E_1) < \infty: \mu(E_i) \searrow \mu(\bigcap_{i \in \mathbb{N}} E_i)$

Proof. There's seven statements:

- (i): This is simply the finite version of definition 4.1 (ii). Formally it can be shown by setting $E_{n+1} = E_{n+2} = \dots = \emptyset$.
- (ii): Since $E_1, E_2 \in \mathbb{E}$, $E_2 \setminus E_1 \in \mathbb{E}$. Since $E_1 \subseteq E_2$, we may write $E_2 = E_1 \cup (E_2 \setminus E_1)$. Now by (ii) we have $\mu(E_2) = \mu(E_1) + \mu(E_2 \setminus E_1)$. Since $\mu(E_2 \setminus E_1) \geq 0$ we must have $\mu(E_1) \leq \mu(E_2)$.
- (iii): Same situation as (ii): $\mu(E_2) = \mu(E_1) + \mu(E_2 \setminus E_1)$. But now that we know $\mu(E_1)$ is finite, we may subtract it on both sides to yield the desired result.
- (iv): Define a series of sets as follows:

$$F_1 = E_1, F_2 = E_2 \setminus E_1, F_3 = E_3 \setminus (E_1 \cup E_2), \dots, F_n = E_n \setminus \left(\bigcup_{i=1}^{n-1} E_i \right) \quad (4.3)$$

These F_i are pairwise disjoint and $\bigcup_{i \in \mathbb{N}} F_i = \bigcup_{i \in \mathbb{N}} E_i$. For all $n \in \mathbb{N}$ we have $F_n \subseteq E_n$. So using definition 4.1 (ii) and (ii) above we get:

$$\mu\left(\bigcup_{i \in \mathbb{N}} E_i\right) = \mu\left(\bigcup_{i \in \mathbb{N}} F_i\right) = \sum_{i \in \mathbb{N}} \mu(F_i) \leq \sum_{i \in \mathbb{N}} \mu(E_i) \quad (4.4)$$

- (v): This is just the finite version of (iv).
- (vi): Because of (ii) $\mu(E_1) \leq \mu(E_2) \leq \mu(E_3) \leq \dots$. Now define a sequence of sets:

$$F_1 = E_1, F_2 = E_2 \setminus E_1, F_3 = E_3 \setminus E_2, \dots, F_n = E_n \setminus E_{n-1}, \dots \quad (4.5)$$

These are pairwise disjoint, and $E_n = \bigcup_{i=1}^n F_i$. This also means that $\bigcup_{i \in \mathbb{N}} F_i = \bigcup_{i \in \mathbb{N}} E_i$. So using (i) above and 4.1 (ii) we get:

$$\mu(E_n) = \mu\left(\bigcup_{i=1}^n F_i\right) = \sum_{i=1}^n \mu(F_i) \nearrow \sum_{i \in \mathbb{N}} \mu(F_i) = \quad (4.6)$$

$$\mu\left(\bigcup_{i \in \mathbb{N}} F_i\right) = \mu\left(\bigcup_{i \in \mathbb{N}} E_i\right) \quad (4.7)$$

- (vii): Define a sequence of sets by $F_n = E_1 \setminus E_n$. This sequence has the properties from (vi), so:

$$\mu(F_i) \nearrow \mu\left(\bigcup_{i \in \mathbb{N}} F_i\right) \quad (4.8)$$

But $\mu(F_i) = \mu(E_1) - \mu(E_n)$ by (iii) and $\bigcup_{i \in \mathbb{N}} F_i = E_1 \setminus \bigcap_{i \in \mathbb{N}} E_i$. So again by (iii) $\mu\left(\bigcap_{i \in \mathbb{N}} F_i\right) = \mu(E_1) - \mu\left(\bigcap_{i \in \mathbb{N}} E_i\right)$. This means that:

$$\mu(E_1) - \mu(E_n) \nearrow \mu(E_1) - \mu\left(\bigcap_{i \in \mathbb{N}} E_i\right) \Leftrightarrow \mu(E_n) \searrow \mu\left(\bigcap_{i \in \mathbb{N}} E_i\right) \quad (4.9)$$

□

4.1 Examples

4.1.1 The Lebesgue measure

This will be the most important example: The *Lebesgue measure* on m_n on $(\mathbb{R}^n, \mathbb{B}_n)$ is the (unique) measure that assigns the usual length/area/volume to intervals. So for instance $m_1([a, b]) = b - a$. We'll devote an entire section to this subject below.

4.1.2 The counting measure

For any set X , the *counting measure* μ is defined on the σ -algebra $\mathcal{P}(X)$. $\mu(E)$ is equal to the number of elements in E for finite E , and set to ∞ otherwise.

We want to show that μ is a measure:

- (i): $\mu(\emptyset) = 0$ since the empty set contains zero elements.
- (ii): A union of disjoint sets contains as many elements as the individual sets combined.

4.1.3 Restriction of a measure

If (X, \mathbb{E}, μ) is a measure space, and $A \in \mathbb{E}$. We might consider restricting the measure to the inherited σ -algebra $\mathbb{E}_A = \{A \cap F | F \in \mathbb{E}\}$, setting $\mu_A(F) = \mu(A \cap F)$. The measure space (A, \mathbb{E}_A, μ_A) is denoted the *restriction* of μ to A .

We want to show that μ_A is a measure. Both are shown using the corresponding property of μ :

- (i): $\mu_A(\emptyset) = \mu(A \cap \emptyset) = \mu(\emptyset) = 0$.
- (ii):

$$\begin{aligned} \mu_A \left(\bigcup_{i \in \mathbb{N}} E_i \right) &= \mu \left(A \cap \bigcup_{i \in \mathbb{N}} E_i \right) = \mu \left(\bigcup_{i \in \mathbb{N}} (A \cap E_i) \right) = \\ &= \sum_{i \in \mathbb{N}} \mu(A \cap E_i) = \sum_{i \in \mathbb{N}} \mu_A(E_i) \end{aligned}$$

4.1.4 Linear combinations of measures

Let $(\mu_j)_{j \in J}$ be a family of measure on the measurable space (X, \mathbb{E}) , and $(a_j)_{j \in J}$ an associated family of numbers, $a_j \in \mathbb{R}_+^*$. We may then define a new measure by:

$$\mu(F) = \sum_{j \in J} a_j \mu_j(F) \tag{4.10}$$

We want to show that μ is a measure:

- (i): $\mu(\emptyset) = \sum_{j \in J} a_j \mu_j(\emptyset) = \sum_{j \in J} 0 = 0$.

- (ii): Here, we'll need theorem 1.6 (for the penultimate step):

$$\begin{aligned}\mu\left(\bigcup_{i \in \mathbb{N}} E_i\right) &= \sum_{j \in J} a_j \mu_j\left(\bigcup_{i \in \mathbb{N}} E_i\right) = \sum_{j \in J} \left(a_j \sum_{i \in \mathbb{N}} \mu_j(E_i)\right) = \\ &= \sum_{i \in \mathbb{N}} \sum_{j \in J} a_j \mu_j(E_i) = \sum_{i \in \mathbb{N}} \mu(E_i)\end{aligned}$$

4.1.5 The Dirac measure

For a set X and an $a \in X$, the *Dirac measure* ε_a is a measure on the measurable space $(X, \mathcal{P}(X))$ defined by:

$$\varepsilon_a(F) = \begin{cases} 1 & \text{when } a \in F \\ 0 & \text{otherwise} \end{cases} \quad (4.11)$$

We note that this is also a probability distribution, since $\mu(X) = 1$.

We want to show that ε_a is a measure:

- (i): $\varepsilon_a(\emptyset) = 0$, since $a \notin \emptyset$.
- (ii): $\varepsilon_a\left(\bigcup_{i \in \mathbb{N}} E_i\right)$. This is 1 if $\exists i : a \in E_i$ and 0 otherwise. Since the E_i 's are disjoint a can be in at most one E_i . So this is equal to $\sum_{i \in \mathbb{N}} \mu(E_i)$.

4.2 The term "almost everywhere"

Definition 4.2. Let (X, \mathbb{E}, μ) be a measure space. If for an $F \in \mathbb{E}$ we have $\mu(A) = 0$, then A and any subset of A , is called a *null set* with respect to μ .

Definition 4.3. Let (X, \mathbb{E}, μ) be a measure space. Let $p(x)$ be an open statement about a point $x \in X$. We use the phrase p " μ -almost everywhere" (sometimes shortened to "*a.o.*" or "*p.p.*" for *preque partout*) to mean that:

$$\{x \in X \mid \neg p(\omega)\} \text{ is a null set with respect to } \mu \quad (4.12)$$

For instance, a real function $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be zero m_1 -almost everywhere if there is a m_1 -null set $N \subseteq \mathbb{R}$ so that $f(x) = 0$ for all $x \in \mathbb{R} \setminus N$.

4.2.1 Functions equal almost everywhere

Definition 4.4. Let (X, \mathbb{E}, μ) be a measure space, and let f, g be functions from X to some set Y . We now define the following equivalence relation:

$$f \sim g \Leftrightarrow f(x) = g(x) \text{ } \mu\text{-almost everywhere} \quad (4.13)$$

The functions of a particular equivalence class are in a sense "essentially the same" with respect to the measure μ .

5 Integration

Given a measure space (X, \mathbb{E}, μ) , we wish to be to associate a given function $f : X \rightarrow \mathbb{R}$ with an *integral*: $\int f(x) d\mu$. When μ is the Lebesgue measure, this should be equal to the usual Riemann integral for well-behaved functions. The core of the idea is, that for $A \in \mathbb{E}$, the integral of the indicator function should be:

$$\int 1_A d\mu = \mu(A) \quad (5.1)$$

The Lebesgue integral essentially follows by demanding linearity and appropriate continuity conditions of the integral.

5.1 Simple functions

Definition 5.1. *Let (X, \mathbb{E}) be a measurable space. A function $s : X \rightarrow \mathbb{R}$ is called simple if the image consist only of a finite set of different values.*

An example of simple functions are indicator functions, since they only take on the values 0 or 1. If a simple function s takes on the n values $a_1, a_2, \dots, a_n \in \mathbb{R}$, X can be partitioned into the n sets $A_1 = s^{-1}(\{a_1\})$, $A_2 = s^{-1}(\{a_2\})$, \dots , $A_n = s^{-1}(\{a_n\})$. Now, s can be expressed in terms of indicator functions:

$$s = \sum_{i=1}^n a_i \cdot 1_{A_i} \quad (5.2)$$

Now assume μ is a measure on (X, \mathbb{E}) . If s is measurable, the sets A_i are also measurable. In light of equation 5.1, it seems natural to require linearity from the integral and set:

$$\int s d\mu = \sum_{i=1}^n a_i \int 1_{A_i} d\mu = \sum_{i=1}^n a_i \mu(A_i) \quad (5.3)$$

However, we might run into some problems if both $-\infty$ and ∞ are terms in the sum. But if s is non-negative, things are fine. This hints that it's easier if we work only with non-negative, simple, measurable functions:

Definition 5.2. *Let (X, \mathbb{E}) be a measurable space. Then the collection of all simple, measurable functions from X to $\mathbb{R}_+ = [0, \infty)$ is denoted $\mathcal{S}^+(X, \mathbb{E})$. If μ is a measure on (X, \mathbb{E}) , we define the integral on $\mathcal{S}^+(X, \mathbb{E})$ through equation (5.3).*

Let's explore some properties of simple functions and their integrals.

Theorem 5.1. *Let (X, \mathbb{E}, μ) be a measure space, and let $s \in \mathcal{S}^+$. Assume s can be written in the form:*

$$s = \sum_{j=1}^m b_j \cdot 1_{B_j} \quad (5.4)$$

Here $b_j \in \mathbb{R}_+$, $B_j \in \mathbb{E}$, $\bigcup_{j=1}^m B_j = X$ and the B_j 's are pairwise disjoint. Note that the b_j 's need not be distinct. Then:

$$\sum_{j=1}^m b_j \mu(B_j) = \int s \, d\mu \quad (5.5)$$

Proof. Like above we can write:

$$s = \sum_{i=1}^n a_i \cdot 1_{A_i} \quad (5.6)$$

Here again $A_1 = s^{-1}(\{a_1\})$, $A_2 = s^{-1}(\{a_2\})$, \dots , $A_n = s^{-1}(\{a_n\})$. Comparing with equation (5.4) it's clear that $b_j \in \{a_1, \dots, a_n\}$ and that $A_i = \bigcup_{k=1}^{p_i} B_{q_{ik}}$. Now, calculate:

$$\sum_{j=1}^m b_j \mu(B_j) = \sum_{i=1}^n \left(a_i \sum_{k=1}^{p_i} \mu(B_{q_{ik}}) \right) = \sum_{i=1}^n a_i \mu(A_i) = \int s \, d\mu \quad (5.7)$$

In the penultimate step, we have used theorem 4.1 (i). \square

This ensures that if we can write a simple function as a sum of indicator functions of a partition of X , we can evaluate the integral in the usual way, without worrying that the coefficients are distinct or that the partition sets A_i happen to be of the form $s^{-1}(\{a_i\})$. This will be useful in the proof of the following:

Theorem 5.2. *Let (X, \mathbb{E}, μ) be a measure space and let $s, t \in \mathcal{S}^+(X, \mathbb{E})$ and $c \in \mathbb{R}_+ = [0, \infty)$. Then $c \cdot s$ and $s + t$ are also simple, measurable functions and:*

$$\int (c \cdot s) d\mu = c \int s \, d\mu, \quad \int (s + t) d\mu = \int s \, d\mu + \int t \, d\mu \quad (5.8)$$

Proof. s is simple and has m distinct values a_1, a_2, \dots, a_m . Similar for t with n distinct values b_1, b_2, \dots, b_n . Set $A_i = s^{-1}(\{a_i\})$, $B_j = t^{-1}(\{b_j\})$. Then $c \cdot s$ still has m distinct values (unless $c = 0$ where there's only one distinct value: 0) and is therefore simple. $s + t$ has at most nm distinct values $a_i + b_j$

and is also simple. Measurability follows from theorem 3.9. The first integral is:

$$\int c \cdot s \, d\mu = \int c \sum_{i=1}^m a_i 1_{A_i} d\mu = \int \sum_{i=1}^m (c \cdot a_i) 1_{A_i} d\mu = \quad (5.9)$$

$$\sum_{i=1}^m (c \cdot a_i) \mu(A_i) = c \sum_{i=1}^m a_i \mu(A_i) = c \int s \, d\mu \quad (5.10)$$

For the sum things are less clear. But we can use this trick:

$$X = X \cap X = \left(\bigcup_{i=1}^n A_i \right) \cap \left(\bigcup_{j=1}^m B_j \right) = \bigcup_{i=j=1}^{i=m, j=n} (A_i \cap B_j) \quad (5.11)$$

The sets $A_i \cap B_j$ are disjoint, and the value of $s + t$ is constant and equal to $a_i + b_j$ on each set. Therefore:

$$s + t = \sum_{i=j=1}^{i=m, j=n} (a_i + b_j) 1_{A_i \cap B_j} \quad (5.12)$$

But we can also use the intersections to express the integrals of s and t :

$$\int s \, d\mu + \int t \, d\mu = \int \left(\sum_{i=j=1}^{i=m, j=n} a_i 1_{A_i \cap B_j} \right) + \int \left(\sum_{i=j=1}^{i=m, j=n} b_j 1_{A_i \cap B_j} \right) = \quad (5.13)$$

$$\sum_{i=j=1}^{i=m, j=n} a_i \mu(A_i \cap B_j) + \sum_{i=j=1}^{i=m, j=n} b_j \mu(A_i \cap B_j) = \quad (5.14)$$

$$\sum_{i=j=1}^{i=m, j=n} (a_i + b_j) \mu(A_i \cap B_j) = \int (s + t) d\mu \quad (5.15)$$

In the last step we have used theorem 5.1. □

Theorem 5.3. *Let (X, \mathbb{E}, μ) be a measure space and let $s, t \in \mathcal{S}^+(X, \mathbb{E})$ so that $s \leq t$. Then:*

$$\int s \, d\mu \leq \int t \, d\mu \quad (5.16)$$

Proof. Write $t = s + (t - s)$ and use theorem 5.2:

$$\int t \, d\mu = \int (s + (t - s)) \, d\mu = \int s \, d\mu + \underbrace{\int (t - s) \, d\mu}_{\geq 0} \quad (5.17)$$

□

5.2 The Lebesgue integral and the monotone convergence theorem

We now wish to extend the integral to a class of non-simple, measurable functions:

Definition 5.3. Let (X, \mathbb{E}) be a measurable space. Then the collection of all measurable functions from X to $\mathbb{R}_+^* = [0, \infty]$ is denoted $\mathcal{M}^+(X, \mathbb{E})$.

Note the extension of the co-domain to \mathbb{R}_+^* . These functions are promising in this regard, as they can be expressed as a limit of simple functions:

Theorem 5.4. Let (X, \mathbb{E}) be a measurable space. If $f \in \mathcal{M}^+(X, \mathbb{E})$, there is a non-decreasing sequence of functions $s_1 \leq s_2 \leq \dots \in \mathcal{S}^+(X, \mathbb{E})$, such that it converges pointwise to f

Proof. One such sequence is given by setting:

$$s_n(x) = \begin{cases} 0 & \text{when } 0 \leq f(x) < \frac{1}{2^n} \\ \frac{1}{2^n} & \text{when } \frac{1}{2^n} \leq f(x) < \frac{2}{2^n} \\ \frac{2}{2^n} & \text{when } \frac{2}{2^n} \leq f(x) < \frac{3}{2^n} \\ \vdots & \\ \frac{k}{2^n} & \text{when } \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \\ \vdots & \\ \frac{n2^n-1}{2^n} & \text{when } \frac{n2^n-1}{2^n} \leq f(x) < n \\ n & \text{when } n \leq f(x) \end{cases} \quad (5.18)$$

In words: For the n 'th function $[0, n)$ is chopped into $n2^n$ slices of length $\frac{1}{2^n}$ and the $f(x)$ -values are then rounded down to the nearest such value. Larger $f(x)$ -values are assigned to n . This sequence satisfies all the conditions above. \square

With this in mind, it seems reasonable to define the integral for a non-simple function as follows:

Definition 5.4. (Lebesgue integral) Let (X, \mathbb{E}, μ) be a measure space. For $f \in \mathcal{M}^+(X, \mathbb{E})$ we set:

$$\int f \, d\mu = \sup \left\{ \int s' \, d\mu \mid s' \leq f, s' \in \mathcal{S}^+(X, \mathbb{E}) \right\} \quad (5.19)$$

Theorem 5.5. Let (X, \mathbb{E}, μ) be a measure space and let $s \in \mathcal{S}^+(X, \mathbb{E})$. Setting $f = s$ in equation (5.19) yields the same integral value as the one from equation (5.3).

Proof. By setting $s' = s$ in equation (5.19), we see that the "non-simple integral" \geq "simple integral". On the other hand, if $s' \leq s$ we know from theorem 5.3 that $\int s' d\mu \leq \int s d\mu$ ("simple" integrals). So we must also have "non-simple integral" \leq "simple integral". Hence the two must be equal. \square

This means that we can use the integral symbol without ambiguity in the following. We can now extend theorem 5.3:

Theorem 5.6. *Let (X, \mathbb{E}, μ) be a measure space. For $f, g \in \mathcal{M}^+(X, \mathbb{E})$, $f \leq g$ we have:*

$$\int f d\mu \leq \int g d\mu \quad (5.20)$$

Proof. Let $s \in \mathcal{S}^+(X, \mathbb{E})$. If $s < f$, then $s < g$. So by equation (5.19) we must have $\int f d\mu \leq \int g d\mu$. \square

That definition 5.4 is reasonable is made clear by the following:

Theorem 5.7. *Let (X, \mathbb{E}, μ) be a measure space. Assume a function $I_\mu : \mathcal{M}^+(X, \mathbb{E}) \rightarrow \mathbb{R}^*$ satisfies the following conditions:*

- (i) For $E \in \mathbb{E} : I_\mu(1_E) = \mu(E)$.
- (ii) For $f, g \in \mathcal{M}^+(X, \mathbb{E}) : I_\mu(f + g) = I_\mu(f) + I_\mu(g)$
- (iii) For a sequence in $\mathcal{M}^+(X, \mathbb{E})$ where $f_n \nearrow f : \lim_{n \rightarrow \infty} I_\mu(f_n) = I_\mu(f)$

Then $I_\mu(f) = \int f d\mu$ as defined through equation (5.19).

The proof is rather complex, so we will break it down into a series of steps. First, let's analyze the situation.

Analysis: One might wonder why the homogeneity condition is absent from the list. It turns out it is not needed. Set $f = g$ and (ii) becomes $I_\mu(2f) = 2I_\mu(f)$. By repeated application $I_\mu(nf) = nI_\mu(f)$ for all $n \in \mathbb{R}$. For a rational number $r = \frac{p}{q} \in \mathbb{Q}$:

$$pI_\mu(f) = I_\mu(pf) = I_\mu(q(rf)) = qI_\mu(rf) \Leftrightarrow I_\mu(rf) = rI_\mu(f) \quad (5.21)$$

This can be done for any linear map. But (iii) allows us to generalize to all positive real numbers in this case. For $c \in \mathbb{R}_+$ there exist an increasing Cauchy sequence of rationals $(q_i)_{i \in \mathbb{N}}$ such that $q_i \nearrow c$. Now apply (iii) to the sequence $q_i f \nearrow cf$ to get:

$$\lim_{n \rightarrow \infty} I_\mu(q_i f) = \lim_{n \rightarrow \infty} (q_i I_\mu(f)) = cI_\mu(f) = I_\mu(cf) \quad (5.22)$$

(A similar argument can be made for $c = \infty$ using the sequence $(n)_{n \in \mathbb{N}}$.) So homogeneity follows from (ii) and (iii), but for clarity, we will add it as a separate condition:

$$(iv) \text{ For } f \in \mathcal{M}^+(X, \mathbb{E}), c \in \mathbb{R}_+^* : I_\mu(cf) = cI_\mu(f) \quad (5.23)$$

Now, above we have seen that (i), (ii) and (iv) leads to equation (5.3). If we further require (iii) to be true. From theorem 5.4 we know that such a sequence of simple limits exists, and since the limit is equal to supremum for a non-decreasing sequence, equation (5.19) follows.

So, our analysis shows, that if I_μ exists, it is unique and must be of the form given by equation (5.19). The big question is now whether I_μ actually satisfies the general conditions of theorem 5.7?

Proof of theorem 5.7: (i) is clearly true. Below we will prove (iii), so for now we will assume it to be true. Then (iii) \Rightarrow (ii): Let $f, g \in \mathcal{M}^+(X, \mathbb{E})$. Then pr. theorem 5.4 there exists sequences of simple, measurable functions $(s_i)_{i \in \mathbb{N}}$ and $(t_i)_{i \in \mathbb{N}}$ such that $s_i \nearrow f, t_i \nearrow g$ and therefore $s_i + t_i \nearrow f + g$. So according to (iii) and theorem 5.2:

$$\int (f + g) d\mu = \lim_{i \rightarrow \infty} \int (s_i + t_i) d\mu = \lim_{n \rightarrow \infty} \left(\int s_i d\mu + \int t_i d\mu \right) = \quad (5.24)$$

$$\lim_{n \rightarrow \infty} \int s_i d\mu + \lim_{n \rightarrow \infty} \int t_i d\mu = \int f d\mu + \int g d\mu \quad (5.25)$$

So finally, we need to prove (iii). This is a famous result, so we restate it in its own theorem:

Theorem 5.8. (*Lebesgue's monotone convergence theorem*) Let (X, \mathbb{E}, μ) be a measure space and $f_1, f_2, \dots \in \mathcal{M}^+(X, \mathbb{E})$ so that $f_1 \leq f_2 \leq \dots$. Then:

$$\int \left(\lim_{n \rightarrow \infty} f_i \right) d\mu = \lim_{n \rightarrow \infty} \int f_i d\mu \quad (5.26)$$

Proof. Since the sequence is non-decreasing, we have:

$$f = \lim_{n \rightarrow \infty} f_i = \lim_{n \rightarrow \infty} \sup f_i \quad (5.27)$$

Start by noting the following: According to theorem 3.7, $f \in \mathcal{M}^+(X, \mathbb{E})$. According to theorem 5.6, $\int f_1 d\mu \leq \int f_2 d\mu \leq \dots$. So both sides in the equation are well-defined. Since $f \geq f_i$ for all $i \in \mathbb{N}$ we have - again by theorem 5.6:

$$\int f d\mu \geq \int f_i d\mu \geq \lim_{n \rightarrow \infty} \int f_i d\mu \quad (5.28)$$

If we can show the opposite inequality, we're done. This amounts to showing, that for all $s \in \mathcal{S}^+(X, \mathbb{E})$, $s \leq f$:

$$\int s \, d\mu \leq \lim_{n \rightarrow \infty} \lim \int f_i \, d\mu \quad (5.29)$$

We now use an odd trick: If we could show the following for all $\alpha \in (0, 1)$, this would also show equation (5.29):

$$\alpha \int s \, d\mu = \int (\alpha \cdot s) d\mu \leq \lim_{n \rightarrow \infty} \int f_i \, d\mu \quad (5.30)$$

The reason for the need for α will become clear in a minute. Now, consider the $x \in X$ for which $f(x) = 0$. In this case $s(x) = f_1(x) = \dots = 0$. If not, we have $s(x) \leq f(x)$. If $s(x) = f(x)$, this means $f_n(x) \nearrow s$. In this case, there might not be an n where $s(x) \leq f_n(x)$ (this can only happen if $s(x) = f_i(x)$ from a certain point). So we would like a sharp inequality. This is where α comes in, because then $\alpha \cdot s(x) < f(x)$. Now, when this is true, there must be some n where $\alpha \cdot s(x) \leq f_n(x)$. Now we construct the following sequence of sets:

$$E_n = \{x \in X \mid \alpha \cdot s(x) \leq f_n(x)\} \quad (5.31)$$

$E_n \in \mathbb{E}$ according to (3.31), and $E_1 \subseteq E_2 \subseteq \dots$. And because of the use of α , $\bigcup_{i \in \mathbb{N}} E_i = X$. Now define the following map

$$\nu : \mathbb{E} \rightarrow \mathbb{R}_+^*, \mu(E) = \alpha \int (s \cdot 1_E) d\mu \quad (5.32)$$

We wish to show that ν is a measure. To see this, write s as $s = \sum_{i=1}^n a_i \cdot 1_{A_i}$. Now:

$$\nu(E) = \alpha \int \left(\sum_{i=1}^n a_i \cdot 1_{A_i \cap E} \right) d\mu = \alpha \sum_{i=1}^n a_i \mu(A_i \cap E) = \alpha \int s \, d\mu(A_i \cap E) \quad (5.33)$$

Here $\mu(A_i \cap E)$ is the restriction measure of μ to $(A_i \cap E)$. As shown in section 4.1.4, a linear combination of measures, ν is itself a measure. Now we can use theorem 4.1 (vi) to get:

$$\nu(E_n) = \alpha \int (s \cdot 1_{E_n}) d\mu \nearrow \nu(X) = \alpha \int (s \cdot 1_X) d\mu = \alpha \int s \, d\mu \quad (5.34)$$

By construction $\alpha \cdot s \cdot 1_{E_n} \leq f_n$, so $\alpha \int s \cdot 1_{E_n} d\mu \leq \int f_n \, d\mu$, by theorem 5.6. So restating equation (5.34) from the right:

$$\alpha \int s \, d\mu = \lim_{n \rightarrow \infty} \alpha \int (s \cdot 1_{E_n}) d\mu \leq \lim_{n \rightarrow \infty} \int f_i \, d\mu \quad (5.35)$$

□

From equation (5.23) and theorem 5.6 - both "spoils" from proving the monotone convergence theorem - the following result follows rather easily:

Theorem 5.9. *Let (X, \mathbb{E}, μ) be a measure space and $f \in \mathcal{M}^+(X, \mathbb{E})$. Then:*

$$(i) \int f \, d\mu = 0 \Leftrightarrow f(x) = 0 \text{ } \mu\text{-almost everywhere} \quad (5.36)$$

$$(ii) \int f \, d\mu < \infty \Rightarrow f(x) < \infty \text{ } \mu\text{-almost everywhere} \quad (5.37)$$

Proof. There's two parts:

- Let $A = \{x \in X | f(x) > 0\}$. Then we may write $\infty \cdot f = \infty \cdot 1_A$. Now use equation (5.23):

$$\infty \int f \, d\mu = \int (\infty \cdot f) d\mu = \int (\infty \cdot 1_A) d\mu = \infty \int 1_A \, d\mu = \infty \cdot \mu(A) \quad (5.38)$$

But this is true if and only if $\mu(A) = 0$ (remember, that $0 \cdot \infty = 0$).

- Similarly, let $B = \{x \in X | f(x) < \infty\}$. Then $\infty \cdot 1_B \leq f$. Now use theorem 5.6:

$$\int (\infty \cdot 1_B) d\mu = \infty \cdot \int 1_B \, d\mu = \infty \cdot \mu(B) \leq \int f \, d\mu \quad (5.39)$$

If the integral is finite, the only possibility is $\mu(B) = 0$.

□

The monotone convergence theorem itself can be used to extend theorem 5.7 (ii) to countably many terms:

Theorem 5.10. *Let (X, \mathbb{E}, μ) be a measure space with a sequence $f_1, f_2, \dots \in \mathcal{M}^+(X, \mathbb{E})$. Then:*

$$\int \left(\sum_{i=1}^{\infty} f_i \right) d\mu = \sum_{i=1}^{\infty} \int f_i \, d\mu \quad (5.40)$$

Proof. From $\sum_{i=1}^n f_i \nearrow \sum_{i=1}^{\infty} f_i$ the monotone convergence theorem implies:

$$\int \left(\sum_{i=1}^{\infty} f_i \right) d\mu = \int \left(\lim_{n \rightarrow \infty} \sum_{i=1}^n f_i \right) d\mu = \lim_{n \rightarrow \infty} \int \left(\sum_{i=1}^n f_i \right) d\mu = \quad (5.41)$$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \int f_i \, d\mu = \sum_{i=1}^{\infty} \int f_i \, d\mu \quad (5.42)$$

□

Finally, let's use the monotone convergence theorem to prove yet another famous, useful result:

Theorem 5.11. (*Fatou's lemma*) Let (X, \mathbb{E}, μ) be a measure space and (f_n) a sequence in $\mathcal{M}^+(X, \mathbb{E})$. Then:

$$\int (\liminf f_n) d\mu = \liminf \int f_n d\mu \quad (5.43)$$

Proof. Let $g_p = \inf_{n \geq p} f_n$. So $g_p \leq f_n$ for $p \leq n$. By theorem 5.6:

$$\int g_p d\mu = \int f_n d\mu, \quad p \leq n \quad (5.44)$$

So:

$$\int g_p d\mu \leq \inf_{p \geq n} \int f_n d\mu \leq \liminf \int f_n d\mu \quad (5.45)$$

Since g_p is non-decreasing, and converges to $\liminf f_n$, we can use the monotone convergence theorem to yield the desired result. \square

5.3 Integrals of real functions

We now turn our attention to integrals of functions that may take on positive as well as negative values. Here, having both positive and negative infinities in play will prove problematic, so we will consider only values in \mathbb{R} itself:

Definition 5.5. Let (X, \mathbb{E}) be a measurable space. Then the collection of measurable functions from X to \mathbb{R} is denoted $\mathcal{M}(X, \mathbb{E})$. For $f \in \mathcal{M}(X, \mathbb{E})$ we define (in the pointwise sense):

$$f^+ = f \vee 0, \quad f^- = -(f \wedge 0) \quad (5.46)$$

These are known as the positive and negative parts of f respectively.

It is immediately clear, that $f = f^+ - f^-$, $|f| = f^+ + f^-$, and that $f^+, f^- \in \mathcal{M}^+(X, \mathbb{E})$. We can now define the integral on $\mathcal{M}(X, \mathbb{E})$:

Definition 5.6. Let (X, \mathbb{E}, μ) be a non-empty measure space. We say that an $f \in \mathcal{M}(X, \mathbb{E})$ is μ -integrable if both $\int f^+ d\mu$ and $\int f^- d\mu$ are finite. In that case, we set:

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \quad (5.47)$$

The collection of all μ -integrable functions is denoted $\mathcal{L}(X, \mathbb{E}, \mu)$.

Note that when $f \in \mathcal{M}^+(X, \mathbb{E}) \cap \mathcal{M}(X, \mathbb{E})$, $f^+ = f$ and $f^- = 0$, so the new definition is consistent with equation (5.19). However, since it may be confusing if an integral means the usual \mathcal{M}^+ -definition, and when it means the one defined through equation (5.47), for now we will denote the latter with a \pm sign as the "lower limit" to make the distinction crystal clear. For instance, in this notation equation (5.47) would read:

$$\int_{\pm} f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu \quad (5.48)$$

Also note that for $A \subseteq X$, 1_A is integrable if and only if $A \in \mathbb{E}$ and $\mu(A) < \infty$. In the following X is always assumed to be a measure space (X, \mathbb{E}, μ) , where $X \neq \emptyset$.

Theorem 5.12. *A measurable function $f : X \rightarrow \mathbb{R}$ is integrable if and only if $|f|$ is integrable. In this case:*

$$\left| \int_{\pm} f \, d\mu \right| \leq \int_{\pm} |f| \, d\mu \quad (5.49)$$

Proof. First we show the equivalence:

- \Rightarrow : If f is integrable, then $\int f^+ \, d\mu, \int f^- \, d\mu < \infty$. Since $|f| = f^+ + f^-$, we have $|f|^+ = f^+ + f^-$, and $|f|^- = 0$. Therefore $\int |f|^+ \, d\mu = \int f^+ \, d\mu + \int f^- \, d\mu < \infty$, $\int |f|^- \, d\mu = 0 < \infty$. So $|f|$ is integrable.
- \Leftarrow : If $|f|$ is integrable, then $\int |f|^+ \, d\mu = \int f^+ \, d\mu + \int f^- \, d\mu < \infty$. This is only possible if both terms are finite, and so f is integrable.

The inequality is shown by using the relation $|a - b| \leq a + b$ for real numbers:

$$\left| \int_{\pm} f \, d\mu \right| = \left| \int f^+ \, d\mu - \int f^- \, d\mu \right| \leq \int f^+ \, d\mu + \int f^- \, d\mu = \int_{\pm} |f| \, d\mu \quad (5.50)$$

□

This immediately gives us the following criterion for integrability:

Theorem 5.13. *If $f \in \mathcal{M}(X, \mathbb{E})$, $g \in \mathcal{M}^+(X, \mathbb{E})$, so that $\int g \, d\mu < \infty$. If $|f| \leq g$, then f is integrable.*

Proof. Since $\int |f| \, d\mu \leq \int g \, d\mu < \infty$, $\int |f| \, d\mu$ must be finite, and therefore $|f|$ is integrable. From theorem 5.12 it follows that f is integrable too. □

The integral further has these nice properties:

Theorem 5.14. *If $f, g \in \mathcal{L}(X, \mathbb{E}, \mu)$ and $a \in \mathbb{R}$, then $(a \cdot f), (f + g) \in \mathcal{L}(X, \mathbb{E}, \mu)$ and:*

$$\int_{\pm} (a \cdot f) d\mu = a \int_{\pm} f d\mu, \quad \int_{\pm} (f + g) d\mu = \int_{\pm} f d\mu + \int_{\pm} g d\mu \quad (5.51)$$

Proof. This is shown for each of the new functions:

- $(a \cdot f)$: Split into three cases according to c :
 - $c = 0$: In this case, the result is trivial.
 - $c > 0$: Here $(a \cdot f)^+ = a \cdot f^+$ and $(a \cdot f)^- = a \cdot f^-$. Their integrals:

$$\int (a \cdot f)^+ d\mu = a \int f^+ d\mu, \quad \int (a \cdot f)^- d\mu = a \int f^- d\mu \quad (5.52)$$

Since f is integrable, both are finite, and hence $(a \cdot f) \in \mathcal{L}(X, \mathbb{E}, \mu)$ with integral:

$$\int_{\pm} (a \cdot f) d\mu = a \int f^+ d\mu - a \int f^- d\mu = a \int_{\pm} f d\mu \quad (5.53)$$

- $c < 0$: Now $(a \cdot f)^+ = -a \cdot f^-$ and $(a \cdot f)^- = -a \cdot f^+$. Once again, the integrals are finite:

$$\int (a \cdot f)^+ d\mu = -a \int f^- d\mu, \quad \int (a \cdot f)^- d\mu = -a \int f^+ d\mu \quad (5.54)$$

So $(a \cdot f) \in \mathcal{L}(X, \mathbb{E}, \mu)$ with integral:

$$\int_{\pm} (a \cdot f) d\mu = -a \int f^- d\mu + a \int f^+ d\mu = a \int_{\pm} f d\mu \quad (5.55)$$

- $(f + g)$: Since f and g are integrable, so is $|f|$ and $|g|$. $|f + g| \leq |f| + |g|$. So

$$\int |f + g| d\mu \leq \int (|f| + |g|) d\mu = \int |f| d\mu + \int |g| d\mu < \infty \quad (5.56)$$

This shows that $|f + g|$ and therefore $f + g$ are integrable. For calculating the integral it is not so easy to get a simple expression for $(f + g)^+$ and $(f + g)^-$. But we can write:

$$(f + g)^+ - (f + g)^- = f + g = f^+ - f^- + g^+ - g^- \Leftrightarrow \quad (5.57)$$

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+ \quad (5.58)$$

All these terms are members of $\mathcal{M}^+(X, \mathbb{E})$, so:

$$\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f+g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu \quad (5.59)$$

All of these are finite, since f, g and $f + g$ are integrable. So:

$$\int (f+g)^+ d\mu - \int (f+g)^- d\mu = \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu \quad (5.60)$$

In other words $\int_{\pm}(f+g)d\mu = \int_{\pm} f d\mu + \int_{\pm} g d\mu$.

□

Theorem 5.15. *Let $f, g \in \mathcal{L}(X, \mathbb{E}, \mu)$ so $f \leq g$. Then:*

$$\int_{\pm} f d\mu \leq \int_{\pm} g d\mu \quad (5.61)$$

The equality sign is true if and only if $g = f$ μ -almost everywhere.

Proof. Since $f \leq g$, $g - f \in \mathcal{M}^+(X, \mathbb{E})$. So $\int (g - f) d\mu \geq 0$. Since g and f are both integrable, we may use the "new integral" here instead, and use theorem 5.14:

$$\int_{\pm} (g - f) d\mu \geq 0 = \int_{\pm} g d\mu - \int_{\pm} f d\mu \geq 0 \Leftrightarrow \int_{\pm} g d\mu \geq \int_{\pm} f d\mu \quad (5.62)$$

Furthermore, theorem 5.9 (i) gives us:

$$\int (g - f) d\mu = 0 \Leftrightarrow g - f = 0 \text{ } \mu\text{-almost everywhere} \quad (5.63)$$

This is equivalent to $g = f$ μ -almost everywhere. □

As we have now generalized most basic results to the "new integral", we will drop the \int_{\pm} -notation and simply use the integral sign.

5.4 Lebesgue's dominated convergence theorem

Unfortunately, the limit for a pointwise convergent sequence of integrable functions turn out to not necessarily be integrable. Requiring the associated sequence of integrals to be convergent does not remedy this. However, we can assure such convergence by using the following famous theorem - a cornerstone of Lebesgue integration theory.

Theorem 5.16. (*Lebesgue's dominated convergence theorem*¹)

Let $f_1, f_2, \dots \in \mathcal{M}(X, \mathbb{E})$ be a sequence of functions that converge pointwise to the function f . If there exists a function $g \in \mathcal{M}^+(X, \mathbb{E})$ with $\int g \, d\mu < \infty$ such that $|f_i| \leq g$ for all $i \in \mathbb{N}$. Then:

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \int f \, d\mu \quad (5.64)$$

The function g is sometimes known as a majorant for the sequence.

Proof. By theorem 5.13 we have that all the f_i must be integrable. So must the limit f . At first, we show the theorem for the case where $g(x) < \infty$ for all $x \in X$. This means that $g \in \mathcal{L}(X, \mathbb{E}, \mu)$. Since $|f_n| \leq g$ we have $g + f_n \geq 0$ and $g - f_n \geq 0$. These two sequences are pointwise convergent, since $(f_n)_{i \in \mathbb{N}}$ is. For a convergent sequence, limes superior and limes inferior are both equal to the limit, so:

$$\liminf_{n \rightarrow \infty} g + f_n = \lim_{n \rightarrow \infty} g + f_n = g + f \quad (5.65)$$

$$\liminf_{n \rightarrow \infty} g - f_n = \lim_{n \rightarrow \infty} g - f_n = g - f \quad (5.66)$$

We can now use Fatou's lemma on equations (5.65) and (5.66). First:

$$\int (g + f) d\mu \leq \liminf \int (g + f_n) d\mu \quad (5.67)$$

Using the linearity of the integral:

$$\int g \, d\mu + \int f \, d\mu \leq \liminf \left(\int g \, d\mu + \int f_n \, d\mu \right) \quad (5.68)$$

But since g is constant, the right hand side is $\int g \, d\mu + \liminf \int f_n \, d\mu$. Now subtract $\int g$ from both sides to get:

$$\int f \, d\mu \leq \liminf \int f_n \, d\mu \quad (5.69)$$

Second:

$$\int (g - f) d\mu \leq \liminf \int (g - f_n) d\mu \quad (5.70)$$

By linearity

$$\int g \, d\mu - \int f \, d\mu \leq \liminf \left(\int g \, d\mu - \int f_n \, d\mu \right) \quad (5.71)$$

¹Sometimes simply shortened to DCT.

Once again, the integral of g separates and can be subtracted out:

$$-\int f \, d\mu \leq -\liminf \int f_n \, d\mu \quad (5.72)$$

So here we get:

$$\int f \, d\mu \geq \liminf \int f_n \, d\mu \quad (5.73)$$

So we must have

$$\int f \, d\mu = \liminf \int f_n \, d\mu = \lim \int f_n \, d\mu \quad (5.74)$$

But what if there's $x \in X$ for which $g(x) = \infty$? Let $N = \{x \in X | g(x) = \infty\}$. We can now use $1_{X \setminus N} \cdot g$ as a majorant for the sequence $(1_{X \setminus N} \cdot f_n)_{n \in \mathbb{N}}$ using the argument above, so that:

$$\int (1_{X \setminus N} \cdot f) d\mu = \lim \int (1_{X \setminus N} \cdot f_n) d\mu \quad (5.75)$$

Since $\int g \, d\mu < \infty$, from theorem 5.9 (ii) N is a null set. So the integrals have the same values as the ones without indicator functions, and we refind the desired result. \square

5.4.1 Addendum to the proof of DCT

In the very last step in the proof we used the following:

Theorem 5.17. *If $f \in \mathcal{M}(X, \mathbb{E})$, $g \in \mathcal{L}(X, \mathbb{E}, \mu)$ and $f = g$ μ -almost everywhere, then $f \in \mathcal{L}(X, \mathbb{E}, \mu)$ and:*

$$\int f \, d\mu = \int g \, d\mu \quad (5.76)$$

Proof. Write $f = g + (f - g)$. We know that $f - g$ is zero μ -almost everywhere, but we can't be sure it's an element of $\mathcal{M}^+(X, \mathbb{E})$. However, $|f - g|$ must be, and it is also zero μ -almost everywhere. By theorem 5.9 (i) we have $\int |f - g| d\mu = 0$. By theorem 5.12 $f - g$ is integrable. Finally by theorem 5.14 f as a sum of integrable functions is itself integrable. \square

This has the consequence, that for some theorems, we might weaken some conditions slightly. For instance, for the dominated convergence theorem, we may release the requirement of convergence of f_i for all $x \in X$ to μ -almost everywhere. Related, it would also enough to require that the majorant $|f_i| \leq g$ μ -almost everywhere.

We can also use the theorem to integrate functions that are only defined μ -almost everywhere.

5.5 Integrals over subsets

Let (X, \mathbb{E}, μ) be a measure space, and let $V \in \mathbb{E}$, V non-empty. Consider the restricted measure space (X, \mathbb{E}_V, μ_V) as described in section 4.1.3. To recapitulate:

$$\mathbb{E}_V = \{E \cap V \mid E \in \mathbb{E}\}, \quad \mu_V(E) = \mu|_{\mathbb{E}_V} \quad (5.77)$$

A function $g : V \rightarrow \mathbb{R}$ can be extended so the domain becomes all of X by setting:

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \in V \\ 0 & \text{otherwise} \end{cases} \quad (5.78)$$

By theorem 3.13, \tilde{g} is measurable if and only if g is.

Theorem 5.18. For $g \in \mathcal{M}^+(V, \mathbb{E}_V)$:

$$\int g \, d\mu_V = \int \tilde{g} \, d\mu \quad (5.79)$$

For any \mathbb{E} -measurable g :

$$g \in \mathcal{L}(V, \mathbb{E}_V, \mu_V) \Leftrightarrow \tilde{g} \in \mathcal{L}(X, \mathbb{E}, \mu) \quad (5.80)$$

Furthermore equation (5.79) remains true if this is the case.

Proof. According to theorem 5.7 (i) applied to the two measures, for $E \in \mathbb{E}_V$, $\int \widetilde{1_E} \, d\mu = \mu(E)$, but also $\int 1_E \, d\mu_V = \mu_V(E)$. But since $\widetilde{1_E} = 1_E$ and $\mu(E) = \mu_V(E)$, it follows that for $s \in \mathcal{S}^+(V, \mathbb{E}_V)$:

$$\int \tilde{s} \, d\mu = \int s \, d\mu_V \quad (5.81)$$

Therefore we must have $\int \tilde{g} \, d\mu = \int g \, d\mu_V$. For the second part, assume $g \in \mathcal{L}(V, \mathbb{E}_V, \mu_V)$. According to theorem 5.12 this is equivalent to $\int |g| \, d\mu_V < \infty$. Now "put the tilde" on $|g|$:

$$\widetilde{|g|}(x) = \begin{cases} |g(x)| & \text{if } x \in V \\ 0 & \text{otherwise} \end{cases} = |\tilde{g}(x)| \quad (5.82)$$

So $\widetilde{|g|} = |\tilde{g}|$. Therefore:

$$\int |g| \, d\mu_V = \int |\tilde{g}| \, d\mu < \infty \quad (5.83)$$

But again, according to theorem 5.12 this is equivalent to $\tilde{g} \in \mathcal{L}(X, \mathbb{E}, \mu)$. Now, the integral of g is:

$$\int g \, d\mu_V = \int g^+ \, d\mu_V - \int g^- \, d\mu_V \quad (5.84)$$

Since $\widetilde{g^+}$ must be equal to $(\tilde{g})^+$ and similarly $\widetilde{g^-} = (\tilde{g})^-$, this is equal to:

$$\int (\tilde{g})^+ \, d\mu - \int (\tilde{g})^- \, d\mu = \int \tilde{g} \, d\mu \quad (5.85)$$

□

If $g \in \mathcal{L}(V, \mathbb{E}_V, \mu_V)$, we say that g is μ -integrable over V and use the following notation:

$$\int_V g \, d\mu = \int g \, d\mu_V = \int \tilde{g} \, d\mu \quad (5.86)$$

The notation is used even if g is defined on all of X . Then we set:

$$\int_V g \, d\mu = \int_V g|_V \, d\mu_V \quad (5.87)$$

Since $\widetilde{g|_V} = 1_V \cdot g$ this means:

$$\int_V g \, d\mu = \int 1_V \cdot g \, d\mu \quad (5.88)$$

5.5.1 Integration over intervals

When $(X, \mathbb{E}) = (\mathbb{R}, \mathbb{B})$ and $V = I$ is an interval (a, b) , $(a, b]$, $[a, b)$ or $[a, b]$, we use the well-known notation:

$$\int_I g \, d\mu = \int_a^b g \, d\mu \quad (5.89)$$

(Since single points are \mathbb{B} -null sets, the open-ness of the interval does not matter with regards to the integral). If $\mu = m$ is the Lebesgue measure, we use the (even more well-known) notation:

$$\int_I g \, dm = \int_a^b g(x) \, dx \quad (5.90)$$

5.6 Density with respect to a measure

From a given measure we can easily generate more by combining it with a density function. It is often a practical way to make a measure by specifying a density with respect to the Lebesgue or counting measure.

Theorem 5.19. *Let (X, \mathbb{E}, μ) be a measure space and let $f \in \mathcal{M}^+(X, \mathbb{E})$. Then the following map $\mathbb{E} \rightarrow \mathbb{R}_+^*$ is a measure:*

$$E \mapsto \int_E f \, d\mu \quad (5.91)$$

This measure is denoted $f \cdot \mu$ or simply $f\mu$. f is said to be the measure's density with respect to μ .

Proof. We need to show the two usual conditions:

- (i): $(f \cdot \mu)(\emptyset) = \int_{\emptyset} f \, d\mu = \int 1_{\emptyset} \cdot f \, d\mu = \int 0 \, d\mu = 0$.
- (ii):

$$(f \cdot \mu) \left(\bigcup_{i \in \mathbb{N}} E_i \right) = \int_{\bigcup_{i \in \mathbb{N}} E_i} f \, d\mu = \int 1_{\bigcup_{i \in \mathbb{N}} E_i} \cdot f \, d\mu \quad (5.92)$$

Since the E_i 's are disjoint, we have $1_{\bigcup_{i \in \mathbb{N}} E_i} = \sum_{i \in \mathbb{N}} 1_{E_i}$, so the integral becomes:

$$\int \left(\sum_{i \in \mathbb{N}} 1_{E_i} \right) f \, d\mu = \int \sum_{i \in \mathbb{N}} (1_{E_i} \cdot f) \, d\mu \quad (5.93)$$

Now use theorem 5.10 to get:

$$\sum_{i \in \mathbb{N}} \int (1_{E_i} \cdot f) \, d\mu = \sum_{i \in \mathbb{N}} \int_{E_i} f \, d\mu = \sum_{i \in \mathbb{N}} (f \cdot \mu)(E_i) \quad (5.94)$$

□

Theorem 5.20. *For $f, \varphi \in \mathcal{M}^+(X, \mathbb{E})$ we have:*

$$\int \varphi \, d(f \cdot \mu) = \int (\varphi \cdot f) \, d\mu \quad (5.95)$$

If $f(x) < \infty$ for all $x \in X$, we further have for a $\varphi \in \mathcal{M}(X, \mathbb{E})$:

$$\varphi \in \mathcal{L}(X, \mathbb{E}, f \cdot \mu) \Leftrightarrow \varphi \cdot f \in \mathcal{L}(X, \mathbb{E}, \mu) \quad (5.96)$$

In this case, the formula above remains true.

Proof. Since $(f \cdot \mu)$ is a measure, it is characterized by the three properties of theorem 5.7. Now consider the function

$$I_f : \mathcal{M}^+(X, \mathbb{E}) \rightarrow \mathbb{R}_+^*, \quad I_f(\varphi) = \int (\varphi \cdot f) d\mu \quad (5.97)$$

It turns out that this also satisfies the three properties (through the same properties for μ):

- (i): $I_f(1_E) = \int (1_E \cdot f) d\mu = \int_E f d\mu = (f \cdot \mu)(E)$
- (ii):

$$\begin{aligned} I_f(g_1 + g_2) &= \int ((g_1 + g_2) \cdot f) d\mu = \int (g_1 \cdot f + g_2 \cdot f) d\mu = \\ &= \int (g_1 \cdot f) d\mu + \int (g_2 \cdot f) d\mu = I_f(g_1) + I_f(g_2) \end{aligned}$$

- (iii): Let $g_n \nearrow g$, so $g_n \cdot f \nearrow g \cdot f$:

$$I_f(g_n) = \int (g_n \cdot f) d\mu \nearrow \int (g \cdot f) d\mu = I_f(g) \quad (5.98)$$

So by theorem 5.7, $I_f(\varphi) = \int (\varphi \cdot f) d\mu = \int \varphi d(f \cdot \mu)$.

For the second part, let $\varphi \in \mathcal{M}(X, \mathbb{E})$ be given. The according to theorem 5.12, φ being $(f \cdot \mu)$ -integrable is equivalent to:

$$\int |\varphi| d(f \cdot \mu) = \int (|\varphi| \cdot f) d\mu < \infty \quad (5.99)$$

If f is finite-valued, then $|\varphi|f = |\varphi \cdot f|$, which means that $\varphi \cdot f \in \mathcal{L}(X, \mathbb{E}, \mu)$. For the integrals, start by noting, that since f is non-negative $(\varphi \cdot f)^+ = \varphi^+ \cdot f$ and $(\varphi \cdot f)^- = \varphi^- \cdot f$. So:

$$\int \varphi d(f \cdot \mu) = \int \varphi^+ d(f \cdot \mu) - \int \varphi^- d(f \cdot \mu) = \quad (5.100)$$

$$\int (f \cdot \varphi^+) d\mu - \int (f \cdot \varphi^-) d\mu = \quad (5.101)$$

$$\int (f \cdot \varphi)^+ d\mu - \int (f \cdot \varphi)^- d\mu = \int (\varphi \cdot f) d\mu \quad (5.102)$$

□

5.7 Image measures

In this section, (X, \mathbb{E}, μ) and (Y, \mathbb{F}, ν) are assumed to be measure spaces.

Theorem 5.21. *Let $\varphi : X \rightarrow Y$ be measurable. Then the following map $\mathbb{F} \rightarrow \mathbb{R}_+^*$ is a measure:*

$$F \mapsto \mu(\varphi^{-1}(F)) \quad (5.103)$$

This is called the image measure of μ under φ and is denoted $\varphi(\mu)$.

Proof. We have two properties to show:

- (i): $\varphi(\mu)(\emptyset) = \mu(\varphi^{-1}(\emptyset)) = \mu(\emptyset) = 0$
- (ii):

$$\begin{aligned} \varphi(\mu) \left(\bigcup_{i \in \mathbb{N}} E_i \right) &= \mu \left(\varphi^{-1} \left(\bigcup_{i \in \mathbb{N}} E_i \right) \right) = \mu \left(\bigcup_{i \in \mathbb{N}} \varphi^{-1}(E_i) \right) = \\ &= \sum_{i \in \mathbb{N}} \mu(\varphi^{-1}(E_i)) = \sum_{i \in \mathbb{N}} \varphi(\mu)(E_i) \end{aligned}$$

□

Theorem 5.22. *Let $g \in \mathcal{M}^+(Y, \mathbb{F})$. Then:*

$$\int g \, d\mu(\varphi) = \int g \circ \varphi \, d\mu \quad (5.104)$$

For $g \in \mathcal{M}(Y, \mathbb{F})$:

$$g \in \mathcal{L}(Y, \mathbb{F}, \mu(\varphi)) \Leftrightarrow g \circ \varphi \in \mathcal{L}(X, \mathbb{E}, \mu) \quad (5.105)$$

In this case, the formula above remains valid.

Proof. Since $\mu(\varphi)$ is a measure, the related integral satisfies the three conditions of theorem 5.7. Now, define a map:

$$I_\varphi : \mathcal{M}^+ \rightarrow \mathbb{R}_+^*, \quad I_\varphi(g) = \int g \circ \varphi \, d\mu \quad (5.106)$$

It turns out, that this map satisfies the same three conditions:

- (i): $I_\varphi(1_F) = \int 1_F \circ \varphi \, d\mu = \int 1_{\varphi^{-1}(F)} \, d\mu = \mu(\varphi^{-1}(F))$

- (ii):

$$\begin{aligned} I_\varphi(g_1 + g_2) &= \int (g_1 + g_2) \circ \varphi \, d\mu = \int (g_1 \circ \varphi) + (g_2 \circ \varphi) d\mu = \\ &= \int g_1 \circ \varphi \, d\mu + \int g_2 \circ \varphi \, d\mu = I_\varphi(g_1) + I_\varphi(g_2) \end{aligned}$$

- (iii): Let $g_n \nearrow g$, so $g_n \circ \varphi \nearrow g \circ \varphi$.

$$I_\varphi(g_n) = \int g_n \circ \varphi \, d\mu \nearrow \int g \circ \varphi \, d\mu = I_\varphi(g) \quad (5.107)$$

Now, by theorem 5.7, the two must be the same:

$$\int g \, d\mu(\varphi) = \int g \circ \varphi \, d\mu \quad (5.108)$$

$g \in \mathcal{L}(Y, \mathbb{F}, \mu(\varphi))$ if and only if $\int |g| \, d\mu(\varphi) = \int |g| \circ \varphi \, d\mu < \infty$. But:

$$(|g| \circ \varphi)(x) = |g(\varphi(x))| = |g \circ \varphi|(x) \quad (5.109)$$

So $\int |g \circ \varphi| d\mu < \infty$, which is true if and only if $g \circ \varphi \in \mathcal{L}(X, \mathbb{F}, \mu)$.

For the integrals, note that $(g \circ \varphi)^+ = g^+ \circ \varphi$ and $(g \circ \varphi)^- = g^- \circ \varphi$. So:

$$\begin{aligned} \int g \, d\mu(\varphi) &= \int g^+ \, d\mu(\varphi) - \int g^- \, d\mu(\varphi) = \int g^+ \circ \varphi \, d\mu - \int g^- \circ \varphi \, d\mu = \\ &= \int (g \circ \varphi)^+ \, d\mu(\varphi) - \int (g \circ \varphi)^- \, d\mu(\varphi) = \int g \circ \varphi \, d\mu \end{aligned}$$

□

5.8 Sums as integrals

Using the counting measure, we can express sums as integrals, thereby benefiting from the theoretical framework developed so far. In this subsection μ is always the counting measure.

Theorem 5.23. *For any function $f : J \rightarrow \mathbb{R}_+^*$, we have:*

$$\int_J f \, d\mu = \sum_{j \in J} f(j) \quad (5.110)$$

Proof. For any $j \in J$, we have:

$$\int_{\{j\}} f \, d\mu = \int 1_{\{j\}} \cdot f \, d\mu = f(j) \cdot \mu(\{j\}) = f(j) \quad (5.111)$$

We can extend this to finite and countable J by countable union:

$$\int_J f \, d\mu = \sum_{j \in J} f(j), \quad |J| \leq \aleph_0 \quad (5.112)$$

If J is uncountable, there's two options according to theorem 1.7: Either the sum is infinite, or there's at most a countable amount of finite $J(j)$. We've effectively dealt with the latter case above. In the former case, since the integral is a supremum over a set containing an infinity, it must itself be infinite as well. \square

When we consider functions $f : J \rightarrow \mathbb{R}_{*+}$ instead, according to theorem 5.12 a sufficient and necessary condition for integrability is $\int |J| \, d\mu < \infty$, which leads to the following definition:

Definition 5.7. We denote the collection of integrable functions $f : J \rightarrow \mathbb{R}_{*+}$ as $\ell(J)$ rather than $\mathcal{L}(J, \mathcal{P}(J), \mu)$. For a $f \in \ell(J)$ we define:

$$\sum_{j \in J} f(j) = \int_J f \, d\mu \quad (5.113)$$

When $J = \mathbb{N}$, it is tempting to think of the elements of $\ell(\mathbb{N})$ as the usual infinite sums:

$$\sum_{i=1}^{\infty} a_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n a_i \quad (5.114)$$

There is however a subtle, but important difference: As written above, we sum in the order 1, 2, 3, *ldots* etc. We have made no such assumptions in the integral formulation. However, we see from the integrability criterium, that $\ell(\mathbb{N})$ only contains the absolutely convergent sums. Remember, that for absolutely convergent series (and only those), the order of summation is interchangeable.

5.9 Integration with real parameter

In this section (X, \mathbb{E}, μ) denotes a non-empty measure space, $I \subseteq \mathbb{R}$ a (possibly unbound) interval, and $f : X \times I \rightarrow \mathbb{R}$ a function.

Definition 5.8. For $x \in X$, the function $I \rightarrow \mathbb{R}$ defined by:

$$t \mapsto f(x, t) \quad (5.115)$$

is denoted f_x or $f(x, \cdot)$. Similarly for $t \in I$ the function $X \rightarrow \mathbb{R}$ defined by:

$$x \mapsto f(x, t) \quad (5.116)$$

is denoted f^t or $f(\cdot, t)$. These are also known as intersection functions.

If we assume $f^t \in \mathcal{L}(X, \mathbb{E}, \mu)$ for all $t \in I$ we may define a function F :

$$F : I \rightarrow \mathbb{R}, \quad F(t) = \int_X f^t d\mu = \int_X f(x, t) d\mu(x) \quad (5.117)$$

Here, the notation $\mu(x)$ is used to avoid confusion with measures on I .

Theorem 5.24. Further assume, that for a $t_0 \in I$, all f_x are continuous at t_0 . If $g \in \mathcal{M}^+(X, \mathbb{E})$, such that $\int g d\mu < \infty$ and:

$$\forall t \in I \quad \forall x \in X : |f(x, t)| \leq g(x), \quad (5.118)$$

then F is continuous at t_0 .

Proof. If $(t_n) \rightarrow t_0$ is a convergent sequence in I , then by continuity of f_x at t_0 , we have $(f_x(t_n)) \rightarrow f_x(t_0)$. This is the same as $f(x, t_n) \rightarrow f(x, t_0)$ or $f^{t_n}(x) \rightarrow f^{t_0}(x)$. For all $t \in I$ we have $|f^t| \leq g$. By the theorem of dominated convergence:

$$\lim_{n \rightarrow \infty} \int f^{t_n}(x) d\mu(x) = \int f^{t_0}(x) d\mu(x) \Rightarrow \lim_{n \rightarrow \infty} F(t_n) = F(t_0) \quad (5.119)$$

In other words, F is continuous at t_0 . □

5.9.1 Differentiation under the integral

We may use the theory in this subsection to show that when differentiating an integral expressing, it is sometimes possible to move the differential operator into the integral. The exact conditions will be restated in the following theorem:

Theorem 5.25. Let (X, \mathbb{E}, μ) be a non-empty measure space, $I \subseteq \mathbb{R}$ an interval, and $f : X \times I \rightarrow \mathbb{R}$ which satisfies:

$$\forall t \in I : f^t \in \mathcal{L}(X, \mathbb{E}, \mu), \forall x \in X : \frac{\partial}{\partial t} f(x, t) \text{ exists} \quad (5.120)$$

If there exists a $g \in \mathcal{M}^+(X, \mathbb{E})$ with $\int g \, d\mu < \infty$ that satisfies:

$$\forall t \in I \, \forall x \in X : \left| \frac{\partial}{\partial t} f(x, t) \right| \leq g(x), \quad (5.121)$$

then $x \mapsto \frac{\partial}{\partial t} f(x, t)$ is an element of $\mathcal{L}(X, \mathbb{E}, \mu)$ for all $t \in I$, and F is differentiable in I with:

$$\frac{d}{dt} F(t) = \frac{d}{dt} \int_X f(x, t) \, d\mu = \int_X \frac{\partial}{\partial t} f(x, t) \, d\mu \quad (5.122)$$

Proof. Let $(t_n) \rightarrow t$ be a sequence in I , so that $t_i \neq t$. From this construct the sequence:

$$\frac{F(t_n) - F(t)}{t_n - t} = \frac{1}{t_n - t} \left(\int_X f(t_n, x) \, d\mu - \int_X f(t, x) \, d\mu \right) = \quad (5.123)$$

$$\int_X \frac{f^{t_n}(x) - f^t(x)}{t_n - t} d\mu \quad (5.124)$$

As $n \rightarrow \infty$, the sequence converges to $\frac{d}{dt} F(t)$. The integrand converges pointwise to $\frac{d}{dt} f(x, t)$ as $f^t(x) = f_x(t)$. According to the mean value theorem, there is a $\tau_{x,n}$ between t_n and t such that:

$$\frac{f^{t_n}(x) - f^t(x)}{t_n - t} = \frac{\partial}{\partial t} f(x, \tau_{x,n}) \quad (5.125)$$

Since $\left| \frac{\partial}{\partial t} f(x, t) \right| \leq g(x)$ for all $x \in X$ and $t \in I$, it is also true when $t = \tau_{x,n}$. So:

$$\left| \frac{f^{t_n} - f^t}{t_n - t} \right| \leq g(x) \quad (5.126)$$

Now, the dominated convergence theorem tells us that the differential operator $x \mapsto \frac{\partial}{\partial t} f(x, t)$ is an element of $\mathcal{L}(X, \mathbb{E}, \mu)$. We now have:

$$\frac{d}{dt} F(t) = \lim_{n \rightarrow \infty} \int_X \frac{f^{t_n}(x) - f^t(x)}{t_n - t} d\mu = \int_X \frac{\partial}{\partial t} f(x, t) \, d\mu \quad (5.127)$$

□

6 The Lebesgue measure

Given a standard box (from definition 2.6) $I = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]$ in \mathbb{R}^k , it seems reasonable to assign it a (hyper-)volume of $v(I) = (b_1 - a_1) \cdot (b_2 - a_2) \cdot \cdots \cdot (b_k - a_k)$.

Definition 6.1. Denote the collection of standard boxes in k dimensions by \mathbb{I}_k . The k -dimensional volume function $v_k : \mathbb{I}_k \rightarrow \mathbb{R}_+^*$ is defined as above.

It turns out, that requiring this volume for standard boxes, is enough to define a unique measure on \mathbb{R}^k :

Theorem 6.1. There exists a unique measure m_k on $(\mathbb{R}^k, \mathbb{B}_k)$ which satisfies $m_k(I) = v_k(I)$ for all standard boxes $I \subseteq \mathbb{R}^k$. This measure is called the Lebesgue measure in k dimensions. It has the form:

$$m_k(B) = \inf \left\{ \sum_{n=1}^{\infty} v_k(I_n) \mid \bigcup_{i=1}^{\infty} I_n \subseteq B, I_n \in \mathbb{I}_k \right\} \quad (6.1)$$

The proof of this important result is rather long, so we will devote a subsection to each of the existence and uniqueness parts.

6.1 Existence of the Lebesgue measure

Theorem 6.2. Let $I_1, I_2, \dots, I_n \in \mathbb{I}_k$. Then the union, intersection or set difference between any sets in the collection can be expressed as union of a subset of a finite set of disjoint standard boxes $J = \{J_1, J_2, \dots, J_m\}$.

Proof. Write $I_i = (a_{i1}, b_{i1}] \times (a_{i2}, b_{i2}] \times \dots \times (a_{ik}, b_{ik}]$. Then consider all the coordinates from dimension $1 \leq d \leq k$.

$$C^d = \bigcup_{i=1}^n \{a_{di}, b_{di}\} \quad (6.2)$$

Now, let C_i^d refer to the i 'th element of C^d , sorted by the usual ordering (and removing possible duplicates). Now we can define a set of standard boxes:

$$J = \{(C_{j_1}^1, C_{j_1+1}^1] \times (C_{j_2}^2, C_{j_2+1}^2] \times \dots \times (C_{j_k}^k, C_{j_k+1}^k] \mid 1 \leq j_i < |C^j|\} \quad (6.3)$$

This corresponds to dicing up the smallest standard box that contains all the I_i along cuts that corresponds to the extended sides of all the I_i 's. The result is at most $(2n - 1)^k$ disjoint standard boxes. From this it clear, that any union, intersection or difference between two I 's can be made by a disjoint union of J . Since J is finite the union is finite. From this is clear that any finite amount of set operations can still be written as a finite, disjoint union from elements of J . \square

Theorem 6.3. The map $v_k : \mathbb{I}_k \rightarrow \mathbb{R}_+^*$ has the following properties:

- (i) *Additivity*: If I_1, \dots, I_n are pairwise disjoint and $\bigcup_{i=1}^n I_i = I$, then:

$$v_k(I) = \sum_{i=1}^n v_k(I_i) \quad (6.4)$$

- (ii) *Subadditivity*: If $I \subseteq \bigcup_{i=1}^n I_i$, then:

$$v_k(I) \leq \sum_{i=1}^n v_k(I_i) \quad (6.5)$$

- (iii) *Countable subadditivity*: If $I \subseteq \bigcup_{i \in \mathbb{N}} I_i$, then:

$$v_k(I) \leq \sum_{i \in \mathbb{N}} v_k(I_i) \quad (6.6)$$

Proof. • (i) According to theorem 6.2, the union of the I_i 's can always be written as a finite unions of standard cubes J_1, \dots, J_m as described in the theorem. Since $I = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k]$ is itself a standard cube, the union must contain all the J sets. Adding up the volumes of the J 's, $\sum_{j=1}^m v_k(I_j)$, can - using the notation of theorem 6.2 - be written as:

$$\sum_{j_1=1}^{|C_1|-1} \sum_{j_2=1}^{|C_2|-1} \dots \sum_{j_k=1}^{|C_k|-1} v_k((C_{j_1}^1, C_{j_1+1}^1] \times (C_{j_2}^2, C_{j_2+1}^2] \times \dots \times (C_{j_k}^k, b_{j_k+1}]) \quad (6.7)$$

Expressing the volume as a product:

$$\sum_{j_1=1}^{|C_1|-1} \sum_{j_2=1}^{|C_2|-1} \dots \sum_{j_k=1}^{|C_k|-1} \prod_{i=1}^k (C_{j_{i+1}}^i - C_{j_i}^i) \quad (6.8)$$

Now consider only the last sum. We can now pull out all but the k 'th factor:

$$\sum_{j_k=1}^{|C_k|-1} \prod_{i=1}^k (C_{j_{i+1}}^i - C_{j_i}^i) = \sum_{j_k=1}^{|C_k|-1} (C_{j_k+1}^k - C_{j_k}^k) \prod_{i=1}^{k-1} (C_{j_{i+1}}^i - C_{j_i}^i) \quad (6.9)$$

But the sum is simply equal to $b_k - a_k$ by the construction of the J 's. We can now iterate this process for the $k-1$ 'th, $k-2$ 'th and so on all the way to 1. Then we see that we get exactly $v_k(I)$.

- (ii) By theorem 6.2 both sets can be written as a finite union of disjoint J 's. But since $I \subseteq \bigcup_{i=1}^n I_i$, all the J 's that are part of the left hand sum must also be part of the right hand sum. The desired result follows.
- (iii)

□

6.2 Uniqueness of the Lebesgue measure

7 Product measures

Given two measure spaces (X, \mathbb{E}, μ) and (Y, \mathbb{F}, ν) we would like to find a new, combined measure space $(X \times Y, \mathbb{E} \otimes \mathbb{F}, \mu \otimes \nu)$ such that:

$$\forall E \in \mathbb{E}, F \in \mathbb{F} : \mu \otimes \nu(E \times F) = \mu(E)\nu(F) \quad (7.1)$$

Such a measure is called the *product measure*. In this chapter we will investigate how and when such a measure might be constructed.

7.1 Measurability of product spaces

Let's start with tackling the following problem: Given two measurable spaces (X, \mathbb{E}) and (Y, \mathbb{F}) is there a 'natural' σ -algebra \mathbb{G} on $X \times Y$? Consider the projection operators:

$$\pi_1 : X \times Y \rightarrow X, (x, y) \mapsto x, \quad \pi_2 : X \times Y \rightarrow Y, (x, y) \mapsto y \quad (7.2)$$

We will choose to search for a σ -algebra \mathbb{G} which makes both of these functions measurable. If this is the case, then for $E \in \mathbb{E}$ and $F \in \mathbb{F}$ we must have:

$$\pi_1^{-1}(E) = E \times Y \in \mathbb{G}, \quad \pi_2^{-1}(F) = X \times F \in \mathbb{G} \quad (7.3)$$

This means that $E \times Y \cap X \times F = E \times F \in \mathbb{G}$. We will refer to products such as $E \times F$ as *rectangles*. We will use the set of rectangles as a basis of our definition of a product σ -algebra:

Definition 7.1. *The product σ -algebra between two measurable spaces (X, \mathbb{E}) and (Y, \mathbb{F}) is the smallest possible containing all rectangles. In other words:*

$$\mathbb{E} \otimes \mathbb{F} = \sigma\{E \times F | E \in \mathbb{E}, F \in \mathbb{F}\} \quad (7.4)$$

As per the construction above, another way to characterize the product algebra is as the smallest σ -algebra that makes π_1 and π_2 measurable.

Definition 7.2. *Given two functions $f : X \rightarrow \mathbb{R}$ and $g : Y \rightarrow \mathbb{R}$ we define the tensor product $f \otimes g : X \times Y \rightarrow \mathbb{R}$ by:*

$$(f \otimes g)(x, y) = f(x)g(y) \tag{7.5}$$

(Alternatively the co-domains could all be \mathbb{R}^ or \mathbb{C}).*