Measure Theory and Integration

Karl-Mikael Perfekt[*]

Reading, UK

October 29th, 2018

Contents

[*]k.perfekt@reading.ac.uk

Acknowledgement

These lecture notes are based on a previous course written by Les Bunce. I am very grateful to Les for creating and sharing his course.

1 Introduction and heuristics

The goal of this course is to construct the Lebesgue integral and a corresponding theory of integration. The Lebesgue integral is a generalisation of the Riemann integral, powerful in the sense that it is much more flexible. The set of functions which are Lebesgue integrable is vastly greater than the set of Riemann integrable functions.

Furthermore, the Lebesgue integral is much more malleable to limiting processes. With the tools in this course, we will be able to answer questions such as:

(a) If (f_n) is a sequence of functions $f_n \colon \mathbb{R} \to \mathbb{R}$ which converges **pointwise** to a function $f \colon \mathbb{R} \to \mathbb{R}$, $\lim_n f_n(x) = f(x)$, $\forall x \in \mathbb{R}$, is it true that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx?$$

(b) If (f_n) is a sequence of functions, $f_n \colon \mathbb{R} \to \mathbb{R}$, when is it true that

$$\sum_{n=1}^{\infty} \int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty} \sum_{n=1}^{\infty} f_n(x)\,dx?$$

(c) If $f \colon \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function of two variables, one of which we understand as a parameter, is true that

$$\frac{d}{da} \int_{-\infty}^{\infty} f(x,a)\,dx = \int_{-\infty}^{\infty} \frac{d}{da} f(x,a)\,dx?$$

It turns out that the right starting point is consider the integration of **characteristic functions**. Given a set $A \subset \mathbb{R}$, the characteristic function $\chi_A \colon \mathbb{R} \to \mathbb{R}$ is the function

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Note that if $A = [a, b]$ is an interval, then

$$\int_{\mathbb{R}} \chi_A(x)\, dx = \int_a^b 1\, dx = b - a$$

is the length of A. When integrable, we call the integral of χ_A the **measure** of A (generalised length of A), denoted $\mu(A)$,

$$\mu(A) = \int_{\mathbb{R}} \chi_A(x)\, dx.$$

If $A = [a_1, b_1] \cup [a_2, b_2]$ is the union of two disjoint intervals (i.e. $a_1 < b_1 < a_2 < b_2$) then it is clear that

$$\mu([a_1, b_1] \cup [a_2, b_2]) = \mu([a_1, b_1]) + \mu([a_2, b_2]) = (b_1 - a_1) + (b_2 - a_2).$$

If $A = \cup_n [a_n, b_n]$ is an infinite union of **disjoint**[1] intervals, then $\mu(A)$ is the infinite sum of the individual lengths,

$$\mu\left(\cup_n [a_n, b_n]\right) = \sum_{n=1}^{\infty} \mu([a_n, b_n]) = \sum_{n=1}^{\infty} (b_n - a_n). \tag{1}$$

This agrees with the definition $\mu(A) = \int_{\mathbb{R}} \chi_A(x)\, dx$ as an improper integral, as is usually discussed in connection with the Riemann integral. Later we will justify (1) in terms of **Lebesgue measure** and the **Lebesgue integral**. Note that the measure $\mu(A)$ could be either finite or infinite ($\mu(A) = \infty$) in this example. To describe (1) succinctly, we say that the measure μ is **countably additive** over the **disjoint** sets $[a_n, b_n]$.

If $A = \{a\}$ consists of a single point, its length is 0,

$$\mu(\{a\}) = \int_a^a 1\, dx = 0.$$

For more complicated sets it is not as clear what their measure should be.

[1]Meaning that $[a_n, b_n] \cap [a_m, b_m] = \emptyset$ whenever $n \neq m$.

1.1 question. What is the measure of the set of rational numbers $\mathbb{Q}$? That is, what is $\mu(\mathbb{Q})$?

The correct answer is that $\mu(\mathbb{Q}) = 0$. This is because $\mathbb{Q}$ is **countable** – a concept studied in the next section. This means that it can be written as a sequence (r_n) of rational numbers (without repetition),

$$\mathbb{Q} = \{r_n : n \in \mathbb{N}\}.$$

Since $\mu(\{r_n\}) = 0$, and we desire length to be additive over **countably infinite** unions of **disjoint** sets, it must be that

$$\mu(\mathbb{Q}) = \sum_{n=1}^{\infty} \mu(\{r_n\}) = \sum_{n=1}^{\infty} 0 = 0.$$

It is natural to try to assign every set $A \subset \mathbb{R}$ a measure $\mu(A)$. Unfortunately, as the following theorem shows, it is not possible to do this for every set $A \subset \mathbb{R}$ while also preserving the reasonable properties of μ associated with length. Before we can construct the **Lebesgue integral**, we must therefore first specify the **measurable sets** (and the **measurable functions**). This motivates the layout of these notes: we will begin by studying abstract **measure theory** before we construct the integral.

The following construction uses the axiom of choice and the concept of countability (covered in the next section). It is a relatively complicated construction which you do not have to learn; feel free to skip it.

1.2 theorem (Vitali). *There is a set $A \subset \mathbb{R}$ which cannot reasonably be assigned a measure.*

Proof. Define an equivalence relation R on $\mathbb{R}$ by

$$xRy \iff x - y \in \mathbb{Q}.$$

This relation divides $\mathbb{R}$ into equivalence classes $[x] = \{x + r : r \in \mathbb{Q}\}$. Let L be the set of equivalence classes. Clearly, each equivalence class $\Lambda \in L$ contains a point in $[0, 1]$. For each $\Lambda \in L$, pick exactly one point $x_\Lambda \in \Lambda \cap [0, 1]$ (axiom of choice), and let

$$A = \{x_\Lambda : \Lambda \in L\}.$$

The set $\mathbb{Q} \cap [-1, 1]$ is countable; arrange it in a sequence without repetition

$$\mathbb{Q} \cap [-1, 1] = \{r_n : n \in \mathbb{N}\}.$$

For each $n \in \mathbb{N}$, let

$$A_n = A + r_n.$$

A_n is a translation of the set A. Translations should leave length unchanged ("reasonably"), and therefore it should be that

(2) $\quad \mu(A_n) = \mu(A), \quad \forall n \in \mathbb{N}.$

Let B be the set $B = \cup_n A_n$. Since $A_n \subset [-1, 2]$ for every n, $B \subset [-1, 2]$. Suppose $x \in [0, 1]$. Then x belongs to some equivalence class $\Lambda = [x_\Lambda]$. In other words, there is $r \in \mathbb{Q}$ such that $x - x_\Lambda = r$. Since $0 \leq x, x_\Lambda \leq 1$, it follows that $r \in \mathbb{Q} \cap [-1, 1]$, and thus there is an n such that $r = r_n$. Hence $x \in A_n$, since $x = x_\Lambda + r_n$ and $x_\Lambda \in A$. Hence $x \in B$.

We have shown that

$$[0, 1] \subset B \subset [-1, 2].$$

Since $\mu([0, 1]) = 1$ and $\mu([-1, 2]) = 3$ it should ("reasonably") be that

(3) $\quad 1 \leq \mu(B) \leq 3.$

On the other hand, the family $\{A_n\}_n$ is **disjoint**. This means the following: if $n \neq m$, then

$$A_n \cap A_m = \emptyset;$$

because if $x \in A_n$ and $x \in A_m$, then there are $y, z \in A$ such that

$$x = y + r_n = z + r_m.$$

However, this means that y and z are in the same equivalence class, $[y] = [z]$. A was constructed to have exactly one point from each equivalence class, so it must be that $y = z$, contradicting that $r_n \neq r_m$.

Since $B = \cup_n A_n$ is a **countable** union of **disjoint** sets, the measure μ should ("reasonably") be additive over the sets A_n,

$$\mu(B) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A),$$

where the last equality follows from (2). If $\mu(A) = 0$, then $\mu(B) = 0$. If $\mu(A) > 0$, then $\mu(B) = \infty$. In either case we have arrived at a contradiction to (3).

The conclusion is that we cannot expect the set A to be **measurable** for a measure μ which satisfies certain properties associated with length (marked with reasonably in the argument).

□

2 Cardinality

2.1 *Definitions*

2.1 definition. Given sets X and Y we write

(a) $X \sim Y$ iff there exists a **bijection** $f\colon X \to Y$.

Note that for sets X, Y, and Z

(b) (i) $X \sim X$, if $X \neq \emptyset$; (ii) $X \sim Y \implies Y \sim X$; (iii) $X \sim Y \sim Z \implies X \sim Z$. [(i) $\mathrm{id}_X\colon X \to X$ is a bijection. (ii) if $f\colon X \to Y$ is a bijection, then so is $f^{-1}\colon Y \to X$. (iii) the composition of bijections is a bijection.]

The sets X and Y are said to have the same cardinality if $X \sim Y$, and we write

(c) $\mathrm{Card}(X) = \mathrm{Card}(Y)$ to signify this.

We also write

(d) $\mathrm{Card}(X) \leq \mathrm{Card}(Y)$ iff there is an **injection** $g\colon X \to Y$.

(e) $\mathrm{Card}(X) < \mathrm{Card}(Y)$ iff there is an injection $g\colon X \to Y$ but no bijection $f\colon X \to Y$ (iff $\mathrm{Card}(X) \leq \mathrm{Card}(Y)$ and $\mathrm{Card}(X) \neq \mathrm{Card}(Y)$).

(f) $\mathrm{Card}(\emptyset) \leq \mathrm{Card}(X)$ for every set X, and $\mathrm{Card}(X) \leq \mathrm{Card}(\emptyset)$ iff $X = \emptyset$ (these are conventions).

It is clear that

(A) $X \subset Y \implies \mathrm{Card}(X) \leq \mathrm{Card}(Y)$ [$\emptyset \neq X \subset Y \implies f\colon X \to Y, f(x) = x$, is injective.]

(B) $\mathrm{Card}(X) \leq \mathrm{Card}(Y) \leq \mathrm{Card}(Z) \implies \mathrm{Card}(X) \leq \mathrm{Card}(Z)$ [the composition of injections is injective.]

Here are two elementary observations.

2.2 proposition. *There exists an injection $f\colon X \to Y$ iff there exists a surjection $g\colon Y \to X$.*

Proof. ($\Rightarrow$) Let $f\colon X \to Y$ be injective. Consider $Y = f(X) \cup (Y \setminus f(X))$. Pick $a \in X$. Define $g\colon Y \to X$ by $g(f(x)) = x$ if $x \in X$ and $g(y) = a$ if $y \in Y \setminus f(X)$. Clearly, g is surjective.

($\Leftarrow$) Suppose $g\colon Y \to X$ is surjective. For each $x \in X$ choose $y_x \in Y$ such that $g(y_x) = x$. Then $f\colon X \to Y$, $f(x) = y_x$ is injective ($f(x_1) = f(x_2) \implies y_{x_1} = y_{x_2} \implies x_1 = g(y_{x_1}) = g(y_{x_2}) = x_2$). □

2.3 proposition. *Let X, Y be non-empty. Then* *(a)* Card$(X) \leq$ Card(Y) *iff there is a surjection* $g\colon Y \to X$ *[by Definition 2.1 (d) and Proposition 2.2].*

(b) For every function $f\colon X \to Y$,

$$\text{Card}(f(X)) \leq \text{Card}(X)$$

[since $f\colon X \to f(X)$ *is a surjection].*

We now state the **Schröder–Bernstein Theorem**.

2.4 theorem. *Let* X, Y *be sets such that* Card$(X) \leq$ Card(Y) *and* Card$(Y) \leq$ Card(X). *Then* Card$(X) =$ Card(Y).

Proof. See e.g. Algebra by S. Lang, pp. 511–512. □

2.2 *Power sets*

The **power set**, $\mathbb{P}(X)$, of a set X, is the **set of all subsets of** X. It is clear that Card$(X) \leq$ Card$(\mathbb{P}(X))$ [if $X \neq \emptyset$, then $f\colon X \to \mathbb{P}(X)$, $f(x) = \{x\}$, is an injection.] In fact;

2.5 theorem (Cantor). Card$(X) <$ Card$(\mathbb{P}(X))$.

Proof. Suppose there is a bijection $f\colon X \to \mathbb{P}(X)$ (or even just a surjection). Put $S = \{x \in X : x \notin f(x)\}$. Pick $a \in X$ such that $f(a) = S$ (f is surjective). Now, if $a \in S$ then $a \notin f(a) = S$, and if $a \notin S$ then $a \in f(a) = S$ – a contradiction. Hence, there is no such f. □

Given $\emptyset \neq A \subset X$ the characteristic function of A, $\chi_A\colon X \to \{0,1\}$ is defined by

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

By convention, $\chi_\emptyset = 0$ (the zero function).

2.6 notation. $\{0,1\}^X$ is the set of all functions $f\colon X \to \{0,1\}$.

Note that $f \in \{0,1\}^X$ iff $f = \chi_A$ where $A = \{x \in X : f(x) = 1\}$. Thus, $F\colon \mathbb{P}(X) \to \{0,1\}^X$, $F(A) = \chi_A$ is a **bijection**. Hence

2.7 proposition. Card$(\mathbb{P}(X)) =$ Card$(\{0,1\}^X)$.

2.3 *Countable sets*

2.8 DEFINITION. A set X is **countable** iff $X = \emptyset$ or there is an injection $f\colon X \to \mathbb{N}$.

Equivalently, a set X is **countable** iff $X = \emptyset$ or there is a surjection $g\colon \mathbb{N} \to X$ (by Proposition 2.2).

Sequences. Note that a sequence (x_n) in a set X is the function $f\colon \mathbb{N} \to X$, $f(n) = x_n$, and conversely every function $f\colon \mathbb{N} \to X$ may be realised as the sequence $(f(n))$ in X. Thus, the sequences in X are precisely the functions $f\colon \mathbb{N} \to X$. Thus, we have

$$\emptyset \neq \text{X is countable iff there is a sequence } (x_n) \text{ in } X \text{ such that } X = \{x_n : n \in \mathbb{N}\}.$$

2.9 PROPOSITION. *Let $S \subset \mathbb{N}$, where S is infinite. Then* $\mathrm{Card}(S) = \mathrm{Card}(\mathbb{N})$

Proof. Let x_1 be the smallest element of S. Inductively, we may choose a sequence (x_n) in S such that, for each n, x_{n+1} is the smallest element of $S \setminus \{x_1, \ldots, x_n\}$. We have $S = \{x_n : n \in \mathbb{N}\}$ [Since, by construction, given any $k \in S$ it holds that $k \leq x_k$, we must have $k = x_r$ for some $1 \leq r \leq k$.] The function $f\colon \mathbb{N} \to S$, $f(n) = x_n$ is bijective. □

For any set S such that $\mathrm{Card}(S) = \mathrm{Card}(\mathbb{N})$, we say that S is **countably infinite**.

2.10 PROPERTIES. The following is true.

(a) All finite sets are countable.

(b) Every subset of a countable set is countable.

(c) If $f\colon X \to Y$ is injective and Y is countable, then X is countable.

(d) If $f\colon X \to Y$ is any function and X is countable, then $f(X)$ is countable. (The image of a countable set is countable.)

(e) If X is infinite and countable then $\mathrm{Card}(X) = \mathrm{Card}(\mathbb{N})$.

Proof. Exercise (see Exercise 1). □

2.11 PROPOSITION. $\mathbb{N} \times \mathbb{N}$ *is countable.*

Proof. The map $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(m, n) = 2^m 3^n$ is injective. (Hence $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$). □

Next we show that **a countable union of countable sets is countable**.

2.12 proposition. *Let (X_n) be a sequence of countable sets. Then $S = \bigcup_{n=1}^{\infty} X_n$ is countable.*

Proof. We may suppose $X_m \neq \emptyset$ for all $m \in \mathbb{N}$. For each $m \in \mathbb{N}$, we may write $X_m = \{x_{m,n} : n \in \mathbb{N}\}$, so that

$$S = \{x_{m,n} : m, n \in \mathbb{N}\}.$$

Hence $f\colon \mathbb{N} \times \mathbb{N} \to S$, $f(m,n) = x_{m,n}$, is **surjective**. Therefore, since $\mathbb{N} \times \mathbb{N}$ is countable, so is S. □

2.13 proposition. *The set of rational numbers $\mathbb{Q}$ is countable. (Hence $\mathbb{Q} \sim \mathbb{N}$).*

Proof. Put $\mathbb{Q}_+ = \{x \in \mathbb{Q} : x > 0\}$, $\mathbb{Q}_- = \{x \in \mathbb{Q} : x < 0\}$. The maps $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{Q}_+$, $f(m,n) = m/n$ and $g\colon \mathbb{N} \times \mathbb{N} \to \mathbb{Q}_-$, $g(m,n) = -m/n$ are **surjective**. Hence $\mathbb{Q}_+$ and $\mathbb{Q}_-$ are countable (Proposition 2.11), and so $\mathbb{Q} = \mathbb{Q}_- \cup \{0\} \cup \mathbb{Q}_+$ is countable (Proposition 2.12). □

Here is a corollary for later use.

2.14 corollary. *If $(I_\lambda)_{\lambda \in \Lambda}$ is a disjoint[2] family of intervals of $\mathbb{R}$, then Λ is countable.*

Proof. In this case, for each $\lambda \in \Lambda$, choose a rational number $r_\lambda \in I_\lambda$. Then the map $f\colon \Lambda \to \mathbb{Q}$, $f(\lambda) = r_\lambda$ is injective. Hence, since $\mathbb{Q}$ is countable, Λ is countable. □

2.15 remark. Let X be an **infinite** set. Then

(a) $\operatorname{Card}(\mathbb{N}) \leq \operatorname{Card}(X)$ [Since X must contain a set $\{x_n : n \in \mathbb{N}\}$ where $x_m \neq x_n$ if $m \neq n$. Thus there is an injection $f\colon \mathbb{N} \to X$, $f(n) = x_n$.]

(b) $\operatorname{Card}(\mathbb{N}) < \operatorname{Card}(X)$ iff X is uncountable (i.e. X is **not** countable).

(c) $\mathbb{P}(X)$ is uncountable (by (b) and Theorem 2.5).

(d) $\mathbb{P}(\mathbb{N})$ is uncountable (in particular).

2.4 *The uncountability of* $\mathbb{R}$

We say that a decimal expansion

$$0.a_1a_2a_3\ldots, \qquad (a_n \in \{0,1,\ldots,9\}, \forall n \in \mathbb{N})$$

is **strict** if it does not have a tail of 9s (i.e. there is no N such that $a_n = 9$, $\forall n \geq N$). Every strict expansion of this form lies in $[0,1)$, and every element of $[0,1)$ has a unique strict decimal expansion.

[2] meaning $I_\lambda \cap I_\mu = \emptyset$ whenever $\lambda \neq \mu$.

2.16 PROPOSITION. $[0,1)$ *is uncountable.*

Proof. The proof is by Cantor's diagonal argument (which was also used to prove Theorem 2.5). Suppose that $[0,1)$ were countable. Then (see the discussion following Definition 2.8) there is a sequence (x_n) in $[0,1)$ with

(4) $[0,1) = \{x_n : n \in \mathbb{N}\}.$

For each n, we have the strict decimal expansion

$$x_n = 0.a_{n1}a_{n2}\ldots a_{nn}\ldots$$

But look at $b = 0.b_1b_2\ldots b_n\ldots$ given by

$$b_n = \begin{cases} 1, & \text{if } a_{nn} \neq 1 \\ 2, & \text{if } a_{nn} = 1. \end{cases}$$

This b is chosen so that $b_n \neq a_{nn}$, $\forall n$. Fix $n \in \mathbb{N}$. Since the (strict) decimal expansion of b and x_n differ at the nth place, we have $b \neq x_n$. Therefore, by (4), $b \notin [0,1)$, a contradiction. Hence, $[0,1)$ is uncountable. □

2.17 COROLLARY. *(a)* $\mathbb{R}$ *is uncountable. (b)* $\mathbb{R} \setminus \mathbb{Q}$ *is uncountable.*

Proof. (a) Since $[0,1) \subset \mathbb{R}$, and $[0,1)$ is uncountable, $\mathbb{R}$ cannot be countable. (b) If $\mathbb{R} \setminus \mathbb{Q}$ is countable, then so is $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$ (by Propositions 2.12 and 2.13). Contradiction. □

2.18 EXAMPLE (The Cantor set). The Cantor set

$$C = \left\{ \sum_{n=1}^{\infty} \frac{2a_n}{3^n} : a_n = 0 \text{ or } 1\ \forall n \right\}$$

is a famous compact subset of $[0,1]$. By definition, $f\colon \{0,1\}^{\mathbb{N}} \to C$ is a bijection,

$$f(\chi_A) = \sum_{n=1}^{\infty} \frac{2\chi_A(n)}{3^n}.$$

Hence $C \sim \{0,1\}^{\mathbb{N}}$. Since $\{0,1\}^{\mathbb{N}} \sim \mathbb{P}(\mathbb{N})$ (Proposition 2.7), it follows that $C \sim \mathbb{P}(\mathbb{N})$. In particular, C is **uncountable**.

2.19 THEOREM. $\mathbb{P}(\mathbb{N}) \sim \mathbb{R}$.

Proof. Since $\mathbb{Q} \sim \mathbb{N}$, we have $\mathbb{P}(\mathbb{Q}) \sim \mathbb{P}(\mathbb{N})$ [if $f\colon \mathbb{Q} \to \mathbb{N}$ is a bijection, so is $\bar{f}\colon \mathbb{P}(\mathbb{Q}) \to \mathbb{P}(\mathbb{N})$, $\bar{f}(S) = f(S)$]. Further, if $\mathbb{P}_b(\mathbb{Q})$ denotes the set of all

bounded subsets of $\mathbb{Q}$, then

$$g\colon \mathbb{P}_b(\mathbb{Q}) \to \mathbb{R}, \quad g(S) = \sup S$$

is **surjective** (since every real number is the supremum of a bounded set in $\mathbb{Q}$). Hence

$$\begin{aligned}\mathrm{Card}(\mathbb{R}) \leq \mathrm{Card}(\mathbb{P}_b(\mathbb{Q})) &\leq \mathrm{Card}(\mathbb{P}(\mathbb{Q})) \\ &= \mathrm{Card}(\mathbb{P}(\mathbb{N})) = \mathrm{Card}(C) \leq \mathrm{Card}(\mathbb{R}).\end{aligned}$$

Therefore, by the Schröder-Bernstein theorem (Theorem 2.4),we have equality throughout. □

3 Extended real number system

The extended real line, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, consists of $\mathbb{R}$ together with two additional symbols ∞ and $-\infty$ satisfying the following properties.

(a) $-\infty < \infty$ and $-\infty < x < \infty, \forall x \in \mathbb{R}$.

(b) x + $(\pm\infty) = \pm\infty = (\pm\infty) + x, \quad \forall x \in \mathbb{R}$.

(c) Let $x \in \overline{\mathbb{R}}$. (i) If $x > 0$, then $x \cdot (\pm\infty) = \pm\infty = (\pm\infty) \cdot x$ (including $x = \infty$). (ii) If $x < 0$, then $x \cdot (\pm\infty) = \mp\infty = (\pm\infty) \cdot x$ (including $x = -\infty$).

(d) $0 \cdot (\pm\infty) = 0 = (\pm\infty) \cdot 0$.

No meaning is attached to $\infty - \infty$, $\frac{\infty}{\infty}$, or to $\frac{x}{0}$ for any $x \in \overline{\mathbb{R}}$.

Supremum and infima. Let $\emptyset \neq A \subset \overline{\mathbb{R}}$. Note that A is bounded above by ∞ and below by $-\infty$.

- If A is not bounded above by a **real** number, we write $\sup A = \infty$.
- If A is not bounded below by a **real** number, we write $\inf A = -\infty$.
- We put $\sup\{-\infty\} = -\infty$ and $\inf\{\infty\} = \infty$.

Limits. The following definitions subsume the usual ones of real analysis.

3.1 definition. Let (x_n) be a sequence in $\overline{\mathbb{R}}$.

(a) $\lim x_n = \infty \iff \forall a \in \mathbb{R}\ \exists N$ such that $a \leq x_n, \forall n \geq N$.

(b) $\lim x_n = -\infty \iff \forall a \in \mathbb{R}\ \exists N$ such that $x_n \leq a, \forall n \geq N$.

(c) If $x \in \mathbb{R}$, then $\lim x_n = x \iff \forall \varepsilon > 0\ \exists N$ such that $x - \varepsilon < x_n < x + \varepsilon, \forall n \geq N$. [Note that in this case, (x_n) is eventually in $\mathbb{R}$ after which the definition is the usual one.]

In any case of $\lim x_n = a \in \overline{\mathbb{R}}$, we often use "$x_n \to a$" as equivalent notation.

Monotone sequences. By the above remarks together with the classical convergence theorems for monotone sequences the following are evident.

- If (x_n) is an increasing sequence in $\overline{\mathbb{R}}$, then

$$x_n \to \sup_{n \geq 1} x_n \quad (= \sup\{x_n : n \in \mathbb{N}\}).$$

- If (x_n) is a decreasing sequence in $\overline{\mathbb{R}}$, then

$$x_n \to \inf_{n \geq 1} x_n \quad (= \inf\{x_n : n \in \mathbb{N}\}).$$

3.1 *Upper and lower limits*

Let (x_n) be a sequence in $\overline{\mathbb{R}}$. For each n, put

$$u_n = \sup_{k \geq n} x_k, \quad \ell_n = \inf_{k \geq n} x_k.$$

Then (u_n) is **decreasing** and (ℓ_n) is **increasing** (in $\overline{\mathbb{R}}$), so that

$$\lim u_n = \inf_{n \geq 1} u_n = u, \quad \lim \ell_n = \sup_{n \geq 1} \ell_n = \ell.$$

- u is defined to be the **upper limit** of (x_n), written $\overline{\lim}\, x_n$ or $\limsup x_n$, i.e.

$$\overline{\lim}\, x_n = \inf_{n \geq 1} \sup_{k \geq n} x_k.$$

- ℓ is defined to be the **lower limit** of (x_n), written $\underline{\lim}\, x_n$ or $\liminf x_n$, i.e.

$$\underline{\lim}\, x_n = \sup_{n \geq 1} \inf_{k \geq n} x_k.$$

Observe that

(a) $\ell_n \leq x_n \leq u_n$ for all $n \in \mathbb{N}$. [By definition.]

(b) $\ell_m \leq u_n$, for all $m, n \in \mathbb{N}$. [$m \leq n \implies \ell_m \leq \ell_n \leq u_n$; $m \geq n \implies \ell_m \leq u_m \leq u_n$.]

(c) $\ell \leq u$. [For each m, $\ell_m \leq u_n, \forall n \implies$ for each m, $\ell_m \leq \inf_{n \geq 1} u_n = u \implies \ell = \sup_{m \geq 1} \ell_m \leq u$.]

In other words,

$$\underline{\lim}\, x_n \leq \overline{\lim}\, x_n.$$

3.2 theorem. *Let (x_n) be a sequence in $\overline{\mathbb{R}}$. Then $\lim x_n$ exists in $\overline{\mathbb{R}}$ if and only if $\overline{\lim}\, x_n = \underline{\lim}\, x_n$. In this case, $\lim x_n = \overline{\lim}\, x_n = \underline{\lim}\, x_n$.*

Proof. Exercise. □

Sequences of functions. Let (f_n) be a sequence of functions, where $f_n \colon X \to \overline{\mathbb{R}}$, for all $n \in \mathbb{N}$. For each fixed $x \in X$ we have the sequence $(f_n(x))$ in $\overline{\mathbb{R}}$. The functions $\sup_{n \geq 1} f_n$, $\inf_{n \geq 1} f_n$, $\overline{\lim}\, f_n$ and $\underline{\lim}\, f_n \colon X \to \overline{\mathbb{R}}$ are defined **pointwise** as follows.

$$(\sup_{n \geq 1} f_n)(x) = \sup_{n \geq 1} f_n(x), \quad (\inf_{n \geq 1} f_n)(x) = \inf_{n \geq 1} f_n(x), \quad \forall x \in X.$$

$$(\overline{\lim}\, f_n)(x) = \overline{\lim}\, f_n(x), \quad (\underline{\lim}\, f_n)(x) = \underline{\lim}\, f_n(x), \quad \forall x \in X.$$

We also denote $\overline{\lim}\, f_n$ by $\limsup f_n$, and $\underline{\lim}\, f_n$ by $\liminf f_n$.

3.2 *Positive and negative parts of* $f\colon X \to \overline{\mathbb{R}}$

Given $a, b \in \overline{\mathbb{R}}$, put

$$a \vee b = \max\{a, b\}, \quad a \wedge b = \min\{a, b\}.$$

Note that for $a, b \in \mathbb{R}$,

$$a \vee b = \frac{1}{2}(a + b + |a - b|), \quad a \wedge b = \frac{1}{2}(a + b - |a - b|).$$

In particular,

$$a \vee 0 = \frac{1}{2}(a + |a|), \quad (-a) \vee 0 = -(a \wedge 0) = \frac{1}{2}(|a| - a).$$

Evidently, $\infty \vee 0 = \infty$, $(-\infty) \vee 0 = -(\infty \wedge 0) = 0$. Consequently, for every **extended** real number,

$$a = a \vee 0 - (-a) \vee 0, \quad (a \vee 0) \cdot ((-a) \vee 0) = 0, \quad \forall a \in \overline{\mathbb{R}}.$$

Let $f, g\colon X \to \overline{\mathbb{R}}$. Define $f \vee g$, $f \wedge g\colon X \to \overline{\mathbb{R}}$ **pointwise** by

$$(f \vee g)(x) = f(x) \vee g(x), \quad (f \wedge g)(x) = f(x) \wedge g(x), \quad x \in X.$$

Define $f_+, f_-\colon X \to \overline{\mathbb{R}}$ by

$$f_+ = f \vee 0, \quad f_- = (-f) \vee 0.$$

Then

(a) $f_+, f_- \geq 0, \quad f_+ \cdot f_- = 0,$

(b) $f = f_+ - f_-, \quad |f| = f_+ + f_-.$

The functions f_+ and f_- are called the positive and negative parts of f. (Note, however, that $f_- \geq 0$.)

4 σ-ALGEBRAS

4.1 DEFINITION. A σ-algebra on a set X is a subset $\mathcal{F} \subset \mathbb{P}(X)$ such that

(i) $X \in \mathcal{F}$;

(ii) if $A \in \mathcal{F}$, then $X \setminus A \in \mathcal{F}$; (Hence $A \in \mathcal{F} \iff X \setminus A \in \mathcal{F}$.)

(iii) if (A_n) is a sequence in $\mathcal{F}$, then $\cup A_n \in \mathcal{F}$.

In words, a σ-algebra on a set X is a collection of subsets of X which contains X and which is closed under the taking of complements and of countable unions.

Here are some immediate consequences of the definition.

4.2 PROPERTIES. Let $\mathcal{F}$ be a σ-algebra on X. Then

(a) $\emptyset \in \mathcal{F}$ [$X \in \mathcal{F}$ by (i), so $\emptyset = X \setminus X \in \mathcal{F}$ by (ii).]

(b) If (A_n) is a sequence in $\mathcal{F}$, then $\cap A_n \in \mathcal{F}$ [$A_n \in \mathcal{F}, \forall n \overset{(ii)}{\Longrightarrow} X \setminus A_n \in \mathcal{F}, \forall n \overset{(iii)}{\Longrightarrow} X \setminus (\cap A_n) = \cup (X \setminus A_n) \in \mathcal{F} \overset{(ii)}{\Longrightarrow} \cap A_n \in \mathcal{F}$.]

(c) If $A, B \in \mathcal{F}$, then $A \setminus B \in \mathcal{F}$ [$A \setminus B = A \cap (X \setminus B) \in \mathcal{F}$ by (ii) and (b).]

4.3 EXAMPLE. $\mathbb{P}(X)$ is the largest σ-algebra on a given set X, and $\{\emptyset, X\}$ is the smallest.

The following is an **important technical lemma**.

4.4 LEMMA. *Let $\mathcal{F}$ be a σ-algebra on X. Let (A_n) be a sequence in $\mathcal{F}$. Then there is a* ***disjoint*** *sequence (B_n) in $\mathcal{F}$ such that $B_n \subset A_n$, $\forall n$, and $\cup B_n = \cup A_n$.*

Proof. Put $B_1 = A_1$, $B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$, $\forall n \geq 2$. Then $B_n \in \mathcal{F}$ and $B_n \subset A_n$, $\forall n$. Further, if $m < n$, then

$$B_m \cap B_n \subset A_m \cap [A_n \setminus (A_1 \cup \cdots \cup A_{n-1})] \subset A_m \cap (A_n \setminus A_m) = \emptyset.$$

Finally, $x \in \cup A_n \implies$ there is a **least** k with $x \in A_k \implies x \in B_k \subset \cup B_n$. Hence $\cup A_n = \cup B_n$. □

4.5 DEFINITION (**Generating σ-algebras.**). If $(F_\lambda)_{\lambda \in \Lambda}$ is a family of σ-algebras on a set X, then $\cap_{\lambda \in \Lambda} F_\lambda$ is also a σ-algebra on X. Given $S \subset \mathbb{P}(X)$, the intersection of all σ-algebras $\mathcal{F}$ on X with $S \subset \mathcal{F}$ ($\mathbb{P}(X)$ is one such $\mathcal{F}$) is the smallest σ-algebra containing S. It is called the σ-algebra on X generated by S.

4.6 DEFINITION (**Borel sets.**). Let (X, τ) be a topological space. The σ-algebra on X generated by τ (i.e. generated by the set of open subsets of X) is the **Borel algebra**, $\beta(X)$, of X. [Note that Definition 4.1 (ii) $\implies$ $\beta(X)$ is also the

σ-algebra on X generated by the closed subsets of X.] The members of $\beta(X)$ are the **Borel subsets of X**. In particular, $\beta(\mathbb{R})$ denotes the σ-algebra of Borel sets of $\mathbb{R}$, where $\mathbb{R}$ is equipped with its usual topology.

4.7 DEFINITION (**Measurable spaces.**). A measurable space is a pair $(X, \mathcal{F})$ where $\mathcal{F}$ is a σ-algebra on X. The members of $\mathcal{F}$ are referred to as the **$\mathcal{F}$-measurable subsets of** X. (Or just as the measurable subsets of X when safe.)

5 Measurable functions

Let $(X, \mathcal{F})$ be a measurable space.

5.1 definition. $f\colon X \to \overline{\mathbb{R}}$ is **measurable** iff $\{x \in X : f(x) > a\} \in \mathcal{F}$ for every $a \in \mathbb{R}$.

There are three other natural equivalences.

5.2 proposition. *For a function $f\colon X \to \overline{\mathbb{R}}$, the following are equivalent.*

(a) $\{x \in X : f(x) > a\} \in \mathcal{F},\ \forall a \in \mathbb{R}$.

(b) $\{x \in X : f(x) \le a\} \in \mathcal{F},\ \forall a \in \mathbb{R}$.

(c) $\{x \in X : f(x) \ge a\} \in \mathcal{F},\ \forall a \in \mathbb{R}$.

(d) $\{x \in X : f(x) < a\} \in \mathcal{F},\ \forall a \in \mathbb{R}$

Hence, (b), (c), and (d) are each equivalent to the measurability of f. We make ***free use*** *of this fact without mention.*

Proof. The equivalences follow from the following.

- (a)⇔(b): Take complements (and use Definition 4.1(ii)).
- (c)⇔(d): Take complements.
- (a)⇒(c): $a \in \mathbb{R} \implies \{x : f(x) \ge a\} = \cap_{n=1}^{\infty}\{x : f(x) > a - 1/n\}$.
- (c)⇒(a): $a \in \mathbb{R} \implies \{x : f(x) > a\} = \cup_{n=1}^{\infty}\{x : f(x) \ge a + 1/n\}$. □

5.3 example.

(a) Constant functions are measurable. For let $f\colon X \to \overline{\mathbb{R}}$, where $f(x) = \alpha$, $\forall x \in X$ ($\alpha \in \overline{\mathbb{R}}$). Then, for $a \in \mathbb{R}$,

$$\{x \in X : f(x) > a\} = \begin{cases} X, \text{ if } \alpha > a, \\ \emptyset, \text{ if } \alpha \le a. \end{cases}$$

Hence f is measurable, since $X, \emptyset \in \mathcal{F}$.

(b) Let $A \subset X$. Then χ_A is measurable $\iff A \in \mathcal{F}$ ($\iff A$ is measurable).

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \implies \{x : \chi_A(x) > a\} = \begin{cases} X, \text{ if } a < 0, \\ A, \text{ if } 0 \le a < 1, \\ \emptyset, \text{ if } 1 \le a, \end{cases}$$

from which the assertion follows.

(c) Consider $X = \mathbb{R}$ with the σ-algebra of Borel sets $\beta(\mathbb{R})$ – the σ-algebra generated by the open subsets of $\mathbb{R}$. Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is continuous. Then $f^{-1}(U) \subset \mathbb{R}$ is open for every open set $U \subset \mathbb{R}$. In particular, with $U = (a, \infty)$, we find that

$$\{x \in \mathbb{R} : f(x) > a\} = f^{-1}((a, \infty)) \in \beta(\mathbb{R}).$$

Thus continuous functions are Borel-measurable.

Measurability is **preserved under limits**.

5.4 proposition. *Let (f_n) be a sequence of measurable functions, $f\colon X \to \overline{\mathbb{R}}$. Then*

(a) $\sup f_n$ *and* $\inf f_n$ *are measurable, since if $a \in \mathbb{R}$*

$$\{x : \sup f_n(x) \le a\} = \cap_{n=1}^{\infty}\{x : f_n(x) \le a\} \in \mathcal{F},$$

$$\{x : \inf f_n(x) \ge a\} = \cap_{n=1}^{\infty}\{x : f_n(x) \ge a\} \in \mathcal{F}.$$

(b) $\overline{\lim} f_n$ *(*$= \inf_{k\ge 1}(\sup_{n\ge k} f_n)$*) and* $\underline{\lim} f_n$ *(*$= \sup_{k\ge 1}(\inf_{n\ge k} f_n)$*) are measurable [follows immediately from (a)].*

*(c) If (f_n) converges (**pointwise** in $\overline{\mathbb{R}}$), then* $\lim f_n$ *is measurable [in this case,* $\lim f_n = \overline{\lim} f_n$*, so (b)⇒(c)].*

The following lemma may be used to reduce measurability questions to the **real-valued** case $f\colon X \to \mathbb{R}$ (as opposed to $f\colon X \to \overline{\mathbb{R}}$).

5.5 lemma. *$f\colon X \to \overline{\mathbb{R}}$ is measurable $\iff \exists$ a sequence (f_n) of measurable functions $f_n\colon X \to \mathbb{R}$ such that $f_n \to f$.*

Furthermore, (f_n) may be chosen so that for every $x \in X$ for which $f(x) \in \mathbb{R}$, there is $N \ge 1$ such that $f_n(x) = f(x)$ whenever $n \ge N$ (N depends on x).

Proof. (⇒) See Exercise 3. (⇐) By Proposition 5.4 (c). □

5.6 proposition. *If $f, g\colon X \to \overline{\mathbb{R}}$ are measurable, then so are*

(a) $f \vee g$, $f \wedge g$, f_+ and f_-,

(b) fg,

*(c) $f + g$ **if** defined, and $|f|$. [Note that $f + g$ is certainly defined when $f, g\colon X \to [0, \infty]$.]*

Proof. Case I. Let $f, g\colon X \to \mathbb{R}$ be measurable. We shall observe that (i) cf, $c \in \mathbb{R}$, (ii) f^2, (iii) f+g, and (iv) fg are measurable. Let $a \in \mathbb{R}$ throughout.

(i) $c = 0 \implies cf = 0 \implies$ measurable (Example 5.3(a)). For $c \neq 0$,

$$\{x : cf(x) > a\} = \begin{cases} \{x : f(x) > a/c\}, \ c > 0 \\ \{x : f(x) < a/c\}, \ c < 0 \end{cases} \in \mathcal{F}.$$

(ii)

$$\{x : f(x)^2 > a\} = \begin{cases} X, \ a < 0 \\ \{x : f(x) < -\sqrt{a}\} \cup \{x : f(x) > \sqrt{a}\}, \ a \geq 0 \end{cases} \in \mathcal{F}.$$

(iii) Given $x \in X$, $f(x) + g(x) > a \Leftrightarrow f(x) > a - g(x) \Leftrightarrow \exists r \in \mathbb{Q}$ s.t. $f(x) > r > a - g(x) \Leftrightarrow \exists r \in \mathbb{Q}$ s.t. $f(x) > r$ and $g(x) > a - r$. For each $r \in \mathbb{Q}$, let

$$S_r = \{x : f(x) > r\} \cap \{x : g(x) > a - r\} \in \mathcal{F}.$$

Thus, since $\mathbb{Q}$ is **countable**,

$$\{x : f(x) + g(x) > a\} = \cup_{r \in \mathbb{Q}} S_r \in \mathcal{F}$$

(iv) Combining (i), (ii), and (iii), $fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$ is measurable.

General case Let $f, g \colon X \to \overline{\mathbb{R}}$ be measurable. The assertions concerning $f \vee g$, $f \wedge g$, and $f_+ = f \vee 0$ follow from Proposition 5.4 (a). For the other cases, choose as in Lemma 5.5 sequences (f_n) and (g_n) of measurable **real-valued** functions $f_n, g_n \colon X \to \mathbb{R}$ such that $f_n \to f$ and $g_n \to g$.

fg: By (iv), $f_n g_m$ are measurable, $\forall m, n \in \mathbb{N}$. Hence, **fixing** m, $fg_m = \lim_{n\to\infty} f_n g_m$ is measurable by Proposition 5.4 (c). As is $fg = \lim_{m\to\infty} fg_m$.

$f + g$ when defined: By (iii), $f_n + g_n$ is measurable, $\forall n$. Thus, if $f + g$ is everywhere defined, then $f + g = \lim(f_n + g_n)$ is measurable.

Finally, f measurable $\implies -f$ measurable (cf. (i)) $\implies f_- = (-f) \vee 0$ measurable $\implies |f| = f_+ + f_-$ measurable. □

6 Measures

6.1 definition. Let $(X, \mathcal{F})$ be a measurable space. A **measure** on $\mathcal{F}$ is a function $\mu\colon \mathcal{F} \to [0, \infty]$ such that

(i) $\mu(\emptyset) = 0$,

(ii) $\mu(\cup_n A_n) = \sum_{n=1}^{\infty} \mu(A_n)$, whenever (A_n) is a **disjoint sequence** in $\mathcal{F}$.

The triple $(X, \mathcal{F}, \mu)$ is a **measure space**.

6.2 example. The following are examples of measures.

(a) **Dirac measures.** Take a measurable space $(X, \mathcal{F})$ and $a \in X$. Define $\delta_a\colon \mathcal{F} \to \{0, 1\}$ by

$$\delta_a(A) = \begin{cases} 1, & \text{if } a \in A, \\ 0, & \text{if } a \notin A \end{cases} .$$

This is the Dirac measure on $\mathcal{F}$ concentrated at a.

(b) **Counting measure.** Consider $\mu\colon \mathbb{P}(\mathbb{N}) \to [0, \infty]$ given by $\mu(A) = |A|$, when A is finite, and $\mu(A) = \infty$ when A is infinite.

(c) **Lebesgue measure**. With $X = \mathbb{R}$ and $\mathcal{F} = \beta(\mathbb{R})$, let, for $A \in \beta(\mathbb{R})$,

$$\mu(A) = \inf\{\sum(b_n - a_n) \ : \ A \subset \bigcup(a_n, b_n), \\ ((a_n, b_n)) \text{ is a sequence of open intervals}\}.$$

Then μ is a measure on $\beta(\mathbb{R})$, the Lebesgue measure, studied in Section 10. Showing the countable additivity of μ is highly non-trivial, see Theorem 10.8. We will also consider μ on a larger σ-algebra than $\beta(\mathbb{R})$, namely, the σ-algebra of Lebesgue measurable sets.

Let $(X, \mathcal{F}, \mu)$ be a measure space.

6.3 properties. Let $E, F \in \mathcal{F}$ with $E \subset F$. Then

(a) $\mu(F) = \mu(E) + \mu(F \setminus E)$ [$F = E \cup (F \setminus E)$ (disjoint union),]

(b) $\mu(E) \leq \mu(F)$ [follows from (a),]

(c) $\mu(F \setminus E) = \mu(F) - \mu(E)$, **if** $\mu(E) < \infty$. [Follows from (a).]

6.4 properties. For any sequence (A_n) in $\mathcal{F}$, we have $\mu(\cup A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Proof. As in Lemma 4.4 there is a **disjoint** sequence (B_n) in $\mathcal{F}$ such that $B_n \subset A_n$, $\forall n$, and $\cup B_n = \cup A_n$. Then

$$\mu(\cup A_n) = \mu(\cup B_n) = \sum \mu(B_n) \leq \sum \mu(A_n). \quad \square$$

A sequence (A_n) in $\mathcal{F}$ is **increasing** if $A_n \subset A_{n+1}$ for every $n \in \mathbb{N}$, and **decreasing** if $A_n \supset A_{n+1}$ for every $n \in \mathbb{N}$.

6.5 PROPOSITION (Monotone convergence theorem). *If (A_n) is an **increasing** sequence in $\mathcal{F}$, then $\mu(A_n) \to \mu(\cup_{m=1}^{\infty} A_m)$.*

Proof. Let $B_1 = A_1$, and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then (B_n) is a disjoint sequence in $\mathcal{F}$ with $A_n = B_1 \cup \cdots \cup B_n$ and $\cup A_n = \cup B_n$. In fact, (B_n) is the sequence from Lemma 4.4. Then

$$\mu(A_n) = \sum_{m=1}^{n} \mu(B_n) \to \sum_{m=1}^{\infty} \mu(B_m) = \mu(\cup B_m) = \mu(\cup A_m). \quad \square$$

6.6 PROPOSITION (Monotone convergence theorem). *If (A_n) is an **decreasing** sequence in $\mathcal{F}$, with $\mu(A_1) < \infty$, then $\mu(A_n) \to \mu(\cap_{m=1}^{\infty} A_m)$.*

Proof. Apply Proposition 6.5 to the **increasing** sequence $(A_1 \setminus A_n)$:

$$\begin{aligned} \mu(A_1) - \mu(A_n) &= \mu(A_1 \setminus A_n) \to \mu[\cup(A_1 \setminus A_m)] \\ &= \mu[A_1 \setminus (\cap A_m)] = \mu(A_1) - \mu(\cap A_m). \end{aligned} \quad \square$$

6.7 REMARK. The finiteness condition on A_1 can not be dropped. For example consider the counting measure on $\mathbb{P}(\mathbb{N})$. The sets $A_n = \{k \in \mathbb{N} : k \geq n\}$, $n \geq 1$, have infinite measure but form a decreasing sequence with empty intersection.

6.1 *Almost everywhere*

Let $(X, \mathcal{F}, \mu)$ be a measure space.

6.8 DEFINITION. A set $N \in \mathcal{F}$ is said to be a **null set** if $\mu(N) = 0$.

Here are some immediate observations:

(a) If $E \in \mathcal{F}$ is contained in a null set, then E is a null set. [By Properties 6.3 (b).]

(b) **A countable union of null sets is a null set.** [If (E_n) is a sequence in $\mathcal{F}$ with $\mu(E_n) = 0$, $\forall n$, then $\mu(\cup E_n) \leq \sum \mu(E_n) = 0$ by Properties 6.4.]

6.9 REMARK. In the present generality, if $E \subset X$ is contained in a null set we cannot assert that E is a null set because it might not belong to $\mathcal{F}$. It is possible to extend $(X, \mathcal{F}, \mu)$ to its completion $(X, \overline{\mathcal{F}}, \overline{\mu})$ for which every subset of a null set is measurable, but we shall not need this.

6.10 DEFINITION (**Almost everywhere.**). A statement about points in X, which is true except possibly at the points of a set contained in a null set, is said to hold *almost everywhere*, (a.e) in abbreviation.

For example, for functions $f, g\colon X \to \overline{\mathbb{R}}$, $f = g$ (a.e) iff $\{x : f(x) \neq g(x)\}$ is contained in a null set. By confining attention to measurable phenomena the situation tends to tidier.

6.11 EXAMPLE. Let $f, g\colon X \to \overline{\mathbb{R}}$ be measurable, and let $f_n\colon X \to \overline{\mathbb{R}}$ be measurable for each $n \in \mathbb{N}$. Then (see **Exercise 4**) the sets

$$\{x : f(x) \neq g(x)\},\ \{x : f(x) = \infty\},\ \{x : f(x) = -\infty\},\ \{x : \lim_n f_n(x) \text{ exists}\}$$

are all measurable (belong to $\mathcal{F}$). As are their complements. Therefore,

(a) $f = g$ (a.e.) $\Longleftrightarrow \mu(\{x : f(x) \neq g(x)\}) = 0$,

(b) f is finite (a.e.) $\Longleftrightarrow \mu(\{x : f(x) = \pm\infty\}) = 0$,

(c) (f_n) converges (a.e.) $\Longleftrightarrow \mu(\{x : \lim_n f_n(x) \text{ does not exist}\}) = 0$.

When we think of measurable functions in the context of a measure space $(X, \mathcal{F}, \mu)$, we identify the functions which are equal almost everywhere. This is formalised in the next proposition.

6.12 PROPOSITION. *$f \sim g \Longleftrightarrow f = g$ (a.e.) defines an* ***equivalence relation*** *on the set of measurable functions on X.*

Proof. Reflexivity and symmetry are clear. Transitivity: $f \sim g$ and $g \sim h$ implies $f = g$ on $X \setminus N_1$ and $g = h$ on $X \setminus N_2$, for certain null sets N_1 and N_2. In turn this implies that $f = h$ on $(X \setminus N_1) \cap (X \setminus N_2) = X \setminus (N_1 \cup N_2)$, so that $f \sim h$, since $N_1 \cup N_2$ is a null set. □

7 Simple functions and positive measurable functions

A sequence (f_n) of functions $f_n\colon X \to \overline{\mathbb{R}}$ is **increasing** if $f_n(x) \leq f_{n+1}(x)$, $\forall x \in X$, $\forall n \in \mathbb{N}$. In this case $\lim f_n = f$ exists ($f = \sup f_n$), and we write $f_n \nearrow f$. By a **positive** function on a set X is meant a function $f\colon X \to [0,\infty]$.

Let $(X, \mathcal{F})$ be a measurable space.

7.1 definition. A simple function on X is a **measurable function**, $s\colon X \to [0,\infty)$ such that $s(X)$ **is finite**. That is, $s(X) = \{a_1, \dots, a_n\}$ for some **real** numbers $a_1, \dots, a_n \geq 0$.

It is clear that

(i) all sums and products of simple functions are simple functions,

(ii) $\sum_{i=1}^n c_i \chi_{E_i}$ is a simple function whenever $0 \leq c_i < \infty$ and $E_i \in \mathcal{F}$, $i = 1, \dots, n$. [Using Example 5.3 and Proposition 5.6.]

7.2 proposition (Canonical form of a simple function). *Let $s \neq 0$ be a simple function on X. Let $a_1, \dots, a_n$ be the **distinct positive** values taken by s. For each i, put $A_i = \{x : s(x) = a_i\}$. Then*

$$s = \sum_{i=1}^{n} a_i \chi_{A_i},$$

*the right hand side of which is the **canonical form of** s. Note that $A_1, \dots, A_n$ are measurable (that is, they belong to $\mathcal{F}$) and mutually **disjoint**.*

7.3 remark. By convention the canonical form of the **zero function** is $\chi_\emptyset$.

The following theorem is key in constructing the integral of a positive measurable function against a given measure μ.

7.4 theorem. *If $f\colon X \to [0,\infty]$ is measurable, then there is an increasing sequence (s_n) of simple functions on X such $s_n \nearrow f$.*

Proof. We will explicitly construct (s_n). For each n, partition $[0,n)$ into $n2^n$ pieces of length $1/2^n$:

$$[0,n) = \bigcup_{k=1}^{n2^n} \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right).$$

For $n \in \mathbb{N}$ and $1 \leq k \leq n2^n$, let

$$E_{n,k} = f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) = \{x : \frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n}\},$$

and let $F_n = \{x : f(x) \geq n\}$. For each n, let

$$s_n = (n-1)\chi_{F_n} + \sum_{k=1}^{n2^n} \frac{k-1}{2^n}\chi_{E_{n,k}}.$$

Note that $E_{n,1}, \ldots, E_{n,n2^n}, F_n$ are all mutually disjoint. Furthermore, $E_{n,k}$ and F_n are measurable, since f is measurable. Hence s_n is measurable, since it is a sum of measurable real-valued functions.

It remains to show that $s_n \nearrow f$ (that is, for each x, $(s_n(x))$ is increasing and $s_n(x) \to f(x)$). Fix $x \in X$. If $f(x) = \infty$, then $x \in F_n$, $\forall n$. Hence $s_n(x) = (n-1) \nearrow \infty$ for such a point x.

If $f(x) < \infty$, then there is a unique integer $m \geq 0$ such that $m \leq f(x) < m+1$. For $n \leq m$, $x \in F_n$, and hence $s_n(x) = (n-1)$. Hence $s_n(x)$ is increasing for $1 \leq n \leq m$. For $n \geq m+1$, there is a unique k such that $x \in E_{n,k}$, that is,

$$\frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n}.$$

Since

$$(5) \quad s_n(x) = \frac{k-1}{2^n},$$

we see that

$$0 \leq f(x) - s_n(x) < \frac{k}{2^n} - \frac{k-1}{2^n} = \frac{1}{2^n}.$$

Since $2^{-n} \to 0$, it follows that $\lim s_n(x) = f(x)$ also for points x with $f(x) < \infty$ (and therefore for all $x \in X$.)

Next, note that

$$\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right) = \left[\frac{2(k-1)}{2^{n+1}}, \frac{2(k-1)+1}{2^{n+1}}\right) \bigcup \left[\frac{2(k-1)+1}{2^{n+1}}, \frac{2k}{2^{n+1}}\right).$$

Hence $x \in E_{n+1,2k-2} \cup E_{n+1,2k-1}$. Therefore, either

$$s_{n+1}(x) = \frac{2k-2}{2^{n+1}} = \frac{k-1}{2^n}, \text{ or } s_{n+1}(x) = \frac{2k-1}{2^{n+1}}.$$

In both cases, it follows that $s_n(x) \leq s_{n+1}(x)$ (by (5)).

It only remains to check that $s_{n-1}(x) \leq s_n(x)$ for $n = m+1$. In this case we have that

$$\begin{aligned} s_n(x) &= \frac{k-1}{2^{m+1}} = \frac{k}{2^{m+1}} - 2^{-(m+1)} > f(x) - 2^{-(m+1)} \\ &\geq m - 2^{-(m+1)} > m - 1 = s_{n-1}(x). \end{aligned}$$ □

8 The integral for positive measurable functions

Let $(X, \mathcal{F}, \mu)$ be a measure space throughout.

8.1 definition.

(a) If s is a **simple** function on X with canonical form $\sum a_i \chi_{A_i}$, the **integral of** s **over** X with respect to μ is denoted and defined by

$$\int_X s \, d\mu = \sum_i a_i \mu(A_i).$$

(b) If $f\colon X \to [0, \infty]$ is **measurable**, then **the integral of** f **over** X with respect to μ is denoted and defined by

$$\int_X f \, d\mu = \sup \left\{ \int_X s \, d\mu \ : \ s \leq f, \, s \text{ is a simple function on } X \right\}.$$

8.2 remark. By a tedious but straightforward combinatorial argument we have:

$$c \geq 0 \text{ and } s, t \text{ simple on } X \implies \int_X (cs + t) \, d\mu = c \int_X s \, d\mu + \int_X t \, d\mu.$$

In particular, the canonical form requirement in Definition 8.1(a) can be dropped.

The following is immediate from definition.

8.3 proposition.

$$f, g\colon X \to [0, \infty] \text{ measurable and } f \leq g \implies \int_X f \, d\mu \leq \int_X g \, d\mu.$$

8.4 lemma. *If (E_n) is an increasing sequence in $\mathcal{F}$ with $\cup E_n = X$, then*

(a) $\mu(A \cap E_n) \nearrow \mu(A)$, $\forall A \in \mathcal{F}$;

(b) $\int_X s \chi_{E_n} \, d\mu \nearrow \int_X s \, d\mu$, *for every simple function s on X.*

Proof. (a): Since $(A \cap E_n)$ is an increasing sequence of measurable sets with union A, this follows from Proposition 6.5. (b): This follows from (a) and definition. Use that $\chi_E \chi_F = \chi_{E \cap F}$. □

8.5 theorem (Monotone convergence theorem (**MCT**)). *If (f_n) is an increasing* *sequence of measurable functions* $f_n \colon X \to [0,\infty]$ *such that* $f_n \nearrow f$, *then*

$$\int_X f_n \, d\mu \nearrow \int_X f \, d\mu.$$

Proof. Put $\alpha = \sup \int_X f_n \, d\mu$. Then $\alpha \leq \int_X f \, d\mu$. Take any simple $s \leq f$. Let $0 < c < 1$. Put $E_n = \{x : cs(x) \leq f_n(x)\}$, for each n. Then (E_n) is an increasing sequence of measurable sets and $X = \cup E_n$. Since, for each n, $cs\chi_{E_n} \leq f_n\chi_{E_n} \leq f_n$, we have

$$c \int_X s\chi_{E_n} \, d\mu \leq \int_X f_n \, d\mu \leq \alpha.$$

Taking limits, using Lemma 8.4, gives $c \int_X s \, d\mu \leq \alpha$. Since $c < 1$ is arbitrary, $\int_X s \, d\mu \leq \alpha$. Hence, by definition, $\int_X f \, d\mu \leq \alpha$, giving equality. □

8.6 theorem (Fatou's lemma). *If* (f_n) *is any sequence of measurable functions* $f_n \colon X \to [0,\infty]$, *then*

$$\int_X (\underline{\lim} \, f_n) \, d\mu \leq \underline{\lim} \int_X f_n \, d\mu.$$

Proof. Let $g_n = \inf_{k \geq n} f_k$, so that $g_n \nearrow \underline{\lim}_k f_k$. By MCT $\int_X g_n \, d\mu \nearrow \int_X \underline{\lim}_k f_k \, d\mu$. On the other hand,

$$\int_X g_n \, d\mu \leq \int f_k \, d\mu, \; \forall k \geq n \implies \int_X g_n \, d\mu \leq \inf_{k \geq n} \int f_k \, d\mu.$$

Taking limits implies the statement. □

Theorem 7.4 together with MCT immediately gives

8.7 theorem. *If* $f \colon X \to [0,\infty]$ *is measurable, then there exists a sequence* (s_n) *of* ***simple*** *functions such that* $s_n \nearrow f$ *and*

$$\int_X s_n \, d\mu \nearrow \int_X f \, d\mu.$$

8.8 corollary (Semi-linearity of the integral).

$$f, g \colon X \to [0,\infty] \text{ measurable, } c \geq 0 \implies \int_X (cf + g) \, d\mu = c \int_X f \, d\mu + \int_X g \, d\mu.$$

Proof. Step 1: The assertion holds when f and g are simple functions (Remark 8.2). Step 2, general case: choose increasing sequences (s_n) and (t_n) such that $s_n \nearrow f$ and $t_n \nearrow t$, giving $cs_n + t_n \nearrow f$. Now apply MCT, giving us that

$$\int_X (cf+g)\,d\mu = \lim_n \int_X (cs_n + t_n)\,d\mu = \lim_n \left(c\int_X s_n\,d\mu + \int_X t_n\,d\mu \right)$$
$$= c\int_X f\,d\mu + \int_X g\,d\mu. \qquad \square$$

8.1 *Integrating over subsets*

Let $f\colon X \to [0,\infty]$ be measurable.

Given $E \in \mathcal{F}$, we write

$$\int_E f\,d\mu = \int_X f\chi_E\,d\mu \text{ (the integral of f over } E).$$

Thus, for every $E \in \mathcal{F}$, $\int_X f\,d\mu = \int_E f\,d\mu + \int_{X\setminus E} f\,d\mu$.

8.9 proposition. *If $\mu(E) = 0$, then $\int_E f\,d\mu = 0$.*

Proof. If $f = \chi_A$, then $f\chi_E = \chi_{E\cap A}$, and so $\int_E f\,d\mu = \mu(E\cap A) = 0$. It follows that for any simple s, $\int_E s\,d\mu = 0$. For a general measurable function $f\colon X \to [0,\infty]$, take simple s_n so that $s_n \nearrow f$. Then $s_n\chi_E \nearrow f\chi_E$. Hence

$$0 = \int_E s_n\,d\mu \underset{\text{MCT}}{\nearrow} \int_E f\,d\mu \implies \int_E f\,d\mu = 0. \quad \square$$

8.10 theorem (Chebyshev's inequality). *For $0 < c < \infty$, let*

$$E = \{x : f(x) \geq c\}.$$

Then

$$\mu(E) \leq \frac{1}{c}\int_E f\,d\mu.$$

Proof. Note that $c\chi_E \leq f\chi_E$. Hence

$$c\mu(E) = \int_X c\chi_E\,d\mu \leq \int_X f\chi_E\,d\mu = \int_E f\,d\mu. \quad \square$$

8.11 proposition.

$$\int_X f\,d\mu = 0 \iff f = 0 \text{ (a.e).}$$

Proof. Let

$$E = \{x : f(x) \neq 0\} = \{x : f(x) > 0\}.$$

Then $E = \cup E_n$, where $E_n = \{x : f(x) \geq 1/n\}$, $n \geq 1$. (E_n) is increasing, so $\mu(E_n) \nearrow \mu(E)$ (MCT). By Chebyshev,

$$\mu(E_n) \leq n \int_{E_n} f \, d\mu \leq n \int_E f \, d\mu = n \int_X f \, d\mu.$$

Hence, if $\int_X f \, d\mu = 0$, then $\mu(E_n) = 0$, $\forall n$, and therefore $\mu(E) = 0$, which by definition means that $f = 0$ (a.e). Conversely , if $f = 0$ (a.e.), then $\mu(E) = 0$, and

$$\int_X f \, d\mu = \int_{X \setminus E} f \, d\mu + \int_E f \, d\mu = \int_E f \, d\mu = 0,$$

by Proposition 8.9. □

8.12 PROPOSITION. *Suppose $\int_X f \, d\mu < \infty$. Then*

*(a) f is **real-valued** (a.e) [i.e. $\mu(\{x : f(x) = \infty\}) = 0$.]*

(b) $\int_X f \, d\mu = \int_E f \, d\mu$, where $E = \{x : f(x) < \infty\}$.

Proof. (a): Let $A = \{x : f(x) = \infty\}$. Then $A \subset A_n = \{x : f(x) \geq n\}$. Hence,

$$\mu(A) \leq \mu(A_n) \leq \frac{1}{n} \int_X f \, d\mu \to 0.$$

(b): This follows by (a) and Proposition 8.9. □

9 The space of integrable functions $L(X, \mathcal{F}, \mu)$

Let $(X, \mathcal{F}, \mu)$ be a measure space. In view of Proposition 8.12 we make the following definition.

9.1 definition. A function $f\colon X \to \mathbb{R}$ is **integrable** with respect to μ if and only if f is measurable and $\int_X |f|\, d\mu < \infty$. We denote the set of all such functions by $L(X, \mathcal{F}, \mu)$.

Given a measurable $f\colon X \to \mathbb{R}$, we have $f = f_+ - f_-$, $|f| = f_+ + f_-$, and by Corollary 8.8 that

$$\int_X |f|\, d\mu = \int_X f_+\, d\mu + \int_X f_-\, d\mu.$$

The following are immediate from this and definition.

9.2 proposition. *Let $f\colon X \to \mathbb{R}$ be measurable. Then*

(a) $f \in L(X, \mathcal{F}, \mu) \iff |f| \in L(X, \mathcal{F}, \mu) \iff f_+, f_- \in L(X, \mathcal{F}, \mu)$.

(b) $|f| \leq g$ *for some* $g \in L(X, \mathcal{F}, \mu) \implies f \in L(X, \mathcal{F}, \mu)$. *[Because if* $|f| \leq g \in L(X, \mathcal{F}, \mu)$*, then* $\int_X |f|\, d\mu \leq \int_X g\, d\mu < \infty$.*]*

9.3 definition. Let $f \in L(X, \mathcal{F}, \mu)$. We **define**

$$\int_X f\, d\mu = \int_X f_+\, d\mu - \int_X f_-\, d\mu.$$

Note that

$$f \in L(X, \mathcal{F}, \mu) \implies \left| \int_X f\, d\mu \right| \leq \int_X |f|\, d\mu,$$

since

$$\left| \int_X f\, d\mu \right| = \left| \int_X f_+\, d\mu - \int_X f_-\, d\mu \right| \leq \int_X f_+\, d\mu + \int_X f_-\, d\mu = \int_X |f|\, d\mu.$$

9.4 theorem (Linearity).

(a) *$L(X, \mathcal{F}, \mu)$ is a real linear space.*

(b) *The map $I\colon L(X, \mathcal{F}, \mu) \to \mathbb{R}$ given by $If = \int_X f\, d\mu$ is a real linear map.*

Proof. (a): If $\alpha \in \mathbb{R}$ and $f, g \in L(X, \mathcal{F}, \mu)$, then, by Corollary 8.8,

$$\int_X |\alpha f + g| \, d\mu \leq \int_X |\alpha||f| + |g| \, d\mu = |\alpha| \int_X |f| \, d\mu + \int_X |g| \, d\mu < \infty,$$

and thus $\alpha f + g \in L(X, \mathcal{F}, \mu)$.

(b): (i) Suppose $f, g \in L(X, \mathcal{F}, \mu)$ with $f, g \geq 0$. Put $h = f - g$. Since $h = h_+ - h_-$, we have $h_+ + g = h_- + f$, hence, using Corollary 8.8,

$$\int_X h_+ \, d\mu + \int_X g \, d\mu = \int_X h_- \, d\mu + \int_X f \, d\mu.$$

And so, by definition,

$$\int_X (f - g) \, d\mu = \int_X h_+ \, d\mu - \int_X h_- \, d\mu = \int_X f \, d\mu - \int_X g \, d\mu.$$

(ii) Given $f, g \in L(X, \mathcal{F}, \mu)$, apply (i) to $f + g = (f_+ + g_+) - (f_- + g_-)$, to get

$$\int_X (f + g) \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu.$$

(iii) Given $\alpha \in \mathbb{R}$, apply (i) to $\alpha f = \alpha f_+ - \alpha f_-$ if $\alpha \geq 0$, and to $\alpha f = (-\alpha) f_- - (-\alpha) f_+$ if $\alpha < 0$. □

9.5 corollary (Order preservation). *If $f, g \in L(X, \mathcal{F}, \mu)$ with $f \leq g$, then $\int_X f \, d\mu \leq \int_X g \, d\mu$.*

Proof. $f \leq g \implies 0 \leq g - f \implies 0 \leq \int_X (g - f) \, d\mu = \int_X g \, d\mu - \int_X f \, d\mu.$ □

9.6 proposition (Almost everywhere). *Let $f \colon X \to \mathbb{R}$ be measurable and $g \in L(X, \mathcal{F}, \mu)$ such that $f = g$ (a.e). Then $f \in L(X, \mathcal{F}, \mu)$, $\int_X f \, d\mu = \int_X g \, d\mu$ and $\int_X |f| \, d\mu = \int_X |g| \, d\mu$.*

Proof. Note that $f = g$ (a.e) if and only if $|f - g| = 0$ (a.e). By Proposition 8.11, this is equivalent to $\int_X |f - g| \, d\mu = 0$. In particular $f - g \in L(X, \mathcal{F}, \mu)$. Since $g \in L(X, \mathcal{F}, \mu)$ and $L(X, \mathcal{F}, \mu)$ is a linear space, it follows that $f \in L(X, \mathcal{F}, \mu)$. Furthermore,

$$\left| \int_X f \, d\mu - \int_X g \, d\mu \right| = \left| \int_X (f - g) \, d\mu \right| \leq \int_X |f - g| \, d\mu = 0,$$

and so $\int_X f \, d\mu = \int_X g \, d\mu$. Since $|f| = |g|$ (a.e) it also follows that $\int_X |f| \, d\mu = \int_X |g| \, d\mu$. □

9.7 proposition. *Let $f\colon X \to \mathbb{R}$ be measurable.*

(a) If $E \in \mathcal{F}$ with $\mu(E) = 0$, then $\int_X f\chi_E \, d\mu = 0$.

(b) If $|f| \leq g$ (a.e), where $g \in L(X, \mathcal{F}, \mu)$, then $f \in L(X, \mathcal{F}, \mu)$.

Proof. See Exercise 7. □

9.1 *Integrating over subsets*

Given $f \in L(X, \mathcal{F}, \mu)$ and $E \in \mathcal{F}$ we write

$$\int_E f \, d\mu = \int_X f\chi_E \, d\mu.$$

Note that $f\chi_E \in L(X, \mathcal{F}, \mu)$, since $|f\chi_E| \leq |f|$. Thus, if $f \in L(X, \mathcal{F}, \mu)$ and $E \in \mathcal{F}$, then

$$\int_X f \, d\mu = \int_E f \, d\mu + \int_{X \setminus E} f \, d\mu.$$

Note that if $f, g \in L(X, \mathcal{F}, \mu)$ and $f = g$ (a.e), then, with $E = \{x : f(x) = g(x)\}$, we have

$$\int_E f \, d\mu = \int_X f \, d\mu = \int_X g \, d\mu = \int_E g \, d\mu.$$

9.2 *Dominated convergence theorem*

9.8 theorem (Dominated convergence theorem (**DCT**)). *Let (f_n) be a sequence in $L(X, \mathcal{F}, \mu)$ such that $f_n \to f$ (pointwise). Suppose that there exists a $g \in L(X, \mathcal{F}, \mu)$ such that $|f_n| \leq g$, $\forall n \in \mathbb{N}$. Then $f \in L(X, \mathcal{F}, \mu)$ and*

$$\lim_{n \to \infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Proof. This is essentially a corollary of Fatou's lemma. Since $|f| \leq g$ and $g \in L(X, \mathcal{F}, \mu)$ it follows by Proposition 9.2(b) that $f \in L(X, \mathcal{F}, \mu)$. Note also that $f = \lim f_n = \underline{\lim} \, f_n$. Hence

(i) $0 \leq f_n + g \implies \int_X (f + g) \, d\mu = \int_X \underline{\lim}(f_n + g) \, d\mu \leq \underline{\lim} \int_X (f_n + g) \, d\mu = \underline{\lim} \int_X f_n \, d\mu + \int_X g \, d\mu \implies \int_X f \, d\mu \leq \underline{\lim} \int_X f_n \, d\mu$.

(ii) Since $0 \leq -f_n + g$, (i) also gives that $\int_X (-f) \, d\mu \leq \underline{\lim} \int_X (-f_n) \, d\mu = -\overline{\lim} \int_X f_n \, d\mu \implies \overline{\lim} \int_X f_n \, d\mu \leq \int_X f \, d\mu$.

We have shown that

$$\int_X f \, d\mu \leq \underline{\lim} \int_X f_n \, d\mu \leq \overline{\lim} \int_X f_n \, d\mu \leq \int_X f \, d\mu.$$

Hence

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu. \quad \square$$

10 Lebesgue measure (on $\mathbb{R}$)

Intervals. The length, $L(I)$ of an interval $I \subset \mathbb{R}$ is defined as follows.

$$L(I) = \begin{cases} b-a, & \text{if } I = (a,b), [a,b], (a,b], \text{ or } [a,b), \\ \infty, & \text{if } I \text{ is an unbounded interval.} \end{cases}$$

10.1 definition (Outer measure (**not a measure!**)). Let $A \subset \mathbb{R}$. The **Lebesgue outer measure**, $\mu^*(A)$, of A is given by

$$\mu^*(A) = \inf \left\{ \sum L(I_n) \;:\; A \subset \cup I_n,\ (I_n) \text{ is a sequence of open intervals} \right\}.$$

10.2 properties.

(a) If $A \subset B \subset \mathbb{R}$, then $\mu^*(A) \leq \mu^*(B)$.

(b) $\mu^*(\{x\}) = 0$, $\forall x \in \mathbb{R}$.

(c) $\mu^*(\emptyset) = 0$.

(d) $\mu^*(I) = L(I)$, for every interval I.

Proof. (a): By definition. (b): For all n, $\{x\} \subset (x - 1/n, x + 1/n)$, whence

$$\mu^*(\{x\}) \leq L((x - 1/n, x + 1/n)) = 2/n \to 0.$$

(c): By (a) and (b), since $\emptyset \subset \{x\}$.

(d): If $I = [a,b]$, then $I \subset (a - 1/n, b + 1/n)$ for every n, so that $\mu^*(I) \leq b - a + 2/n$, giving $\mu^*(I) \leq b - a$. On the other hand, consider a cover of $[a,b]$ by a sequence of **open** intervals (I_n). Since $[a,b]$ is **compact** we may suppose, after renumbering, that $[a,b]$ is contained in the **finite union** of $I_1, \ldots, I_k$, say, where each $I_r \cap [a,b]$ is non-empty (and so is an interval, not necessarily open). This gives us that

$$b - a \leq L(I_1) + \cdots + L(I_k) \leq \sum L(I_n).$$

Hence $b - a \leq \mu^*(I)$ by definition, as required.

(d′): Suppose $I = (a,b), (a,b]$, or $[a,b)$. If $n \geq (b-a)/2$, then $[a + 1/n, b - 1/n] \subset I \subset [a,b]$. By applying (d) twice, we see that

$$(b-a) - 2/n \leq \mu^*(I) \leq b - a,\ \forall n \implies \mu^*(I) = b - a.$$

(d″): If I is an unbounded interval, we can choose $[a_n, b_n] \subset I$ with $b_n - a_n \to \infty$, and since by (d) $b_n - a_n \leq \mu^*(I)$, for all n, we have $\mu^*(I) = \infty$. □

10.3 proposition (Countable subadditivity). *For every sequence* (E_n) *of subsets*

$E_n \subset \mathbb{R}$, *we have that*

$$\mu^*\left(\cup E_n\right) \leq \sum_{n=1}^{\infty} \mu^*(E_n).$$

Proof. Let $\varepsilon > 0$. For each n choose a countable open interval cover $E_n \subset \cup I_{n,m}$ such that

$$\sum_{m=1}^{\infty} L(I_{n,m}) \leq \mu^*(E_n) + \varepsilon/2^n.$$

Since $\cup E_n \subset \cup_{n,m} I_{n,m}$ we have that

$$\mu^*(\cup E_n) \leq \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} L(I_{n,m}) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon \sum_{n=1}^{\infty} \frac{1}{2^n} = \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon.$$

Letting $\varepsilon \to 0$ finishes the proof. □

10.4 proposition. $\mu^*(E) = 0$ *for every countable set* $E \subset \mathbb{R}$.

Proof. $E = \{x_n : n \in \mathbb{N}\} = \cup\{x_n\} \implies \mu^*(E) \leq \sum \mu^*(\{x_n\}) = \sum 0 = 0.$ □

10.5 remark. It follows that every subset of $\mathbb{R}$ with strictly positive outer measure is uncountable. This gives another proof that intervals are uncountable.

For a subset E of $\mathbb{R}$, we write $E^c = \mathbb{R} \setminus E$.

10.6 definition (Lebesgue measurable sets). A set $E \subset \mathbb{R}$ is **Lebesgue measurable** if and only if

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c), \forall A \subset \mathbb{R}.$$

Let $\mathcal{M}$ denote the set of all **Lebesgue measurable subsets** of $\mathbb{R}$.

10.7 proposition. *Let* $E \subset \mathbb{R}$ *with* $\mu^*(E) = 0$. *Then* $E \in \mathcal{M}$.

Proof. $A \subset \mathbb{R} \implies \mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A \cap E^c) = \mu^*(A \cap E^c) \leq \mu^*(A).$ □

The following can be proven by straightforward, but very lengthy, manipulations of the definition.

10.8 theorem (Carathéodory's criterion).

(a) $\mathcal{M}$ *is a* σ*-algebra.*

(b) $\mu = \mu^*|_{\mathcal{M}}$ *is a measure on* $\mathcal{M}$, *called the* ***Lebesgue measure***.

(c) $\mathcal{M}$ *contains all Borel subsets of* $\mathbb{R}$.

Proof. See Theorem 1.7.3 in Tao's "An Introduction to Measure Theory" for (a) and (b).

(c) then follows by: (i) Checking that open intervals are Lebesgue measurable. (ii) Since every open set $U \subset \mathbb{R}$ is a countable union of open intervals, it follows that all open sets are in $\mathcal{M}$. (iii) Since the Borel algebra $\beta(\mathbb{R})$ is the smallest σ-algebra on $\mathbb{R}$ containing the open sets, it follows that $\beta(\mathbb{R}) \subset \mathcal{M}$. □

In particular, $\mathcal{M}$ contains all open sets, closed sets and intervals in $\mathbb{R}$, with $\mu(I) = L(I)$ for every interval I in $\mathbb{R}$. Note that every countable subset of $\mathbb{R}$ is Lebesgue measurable with measure zero, by Propositions 10.4 and 10.7.

10.9 EXAMPLE (An uncountable null set). Recall the Cantor set C of Example 2.18. It is uncountable, but also a null set, that is, $\mu(C) = 0$. [One can understand C as an intersection of closed intervals C_n of length $(2/3)^n$. Hence $C \in \mathcal{M}$ with $\mu(C) \leq (2/3)^n \to 0$, implying $\mu(C) = 0$.]

For $E \subset \mathbb{R}$ and $a \in \mathbb{R}$, let $E + a = \{x + a : x \in E\}$.

10.10 PROPOSITION (Translation invariance). *Let $E \in \mathcal{M}$ and $a \in \mathbb{R}$. Then $E + a \in \mathcal{M}$ and $\mu(E + a) = \mu(E)$.*

Proof. See Exercise 8. □

10.11 PROPOSITION (The Lebesgue measure is complete). *If $A \subset E$, and $E \in \mathcal{M}$ with $\mu(E) = 0$ then $A \in \mathcal{M}$ (and $\mu(A) = 0$).*

Proof. Since $A \subset E$ we have that $\mu^*(A) \leq \mu^*(E) = \mu(E) = 0$. Hence $A \in \mathcal{M}$ by Proposition 10.7. □

10.12 PROPOSITION (Continuous functions). *Let $f, g \colon \mathbb{R} \to \mathbb{R}$ be continuous. Then*

(a) f is Lebesgue measurable. [$a \in \mathbb{R} \implies \{x : f(x) > a\}$ is open and so belongs to $\mathcal{M}$.]

(b) $f = 0$ (a.e) $\implies f = 0$. [$f \neq 0 \implies E = \{x : f(x) \neq 0\}$ contains an interval $I \implies \mu(E) \neq 0 \implies f \neq 0$ (a.e).]

(c) $f = g$ (a.e) $\implies f = g$. [Apply (b) to $f - g$.]

10.13 EXAMPLE. The characteristic function, f, of $\mathbb{Q}$ is Lebesgue measurable with $f = 0$ (a.e). (Why?) But $f \neq 0$.

11 The Lebesgue integral (on $\mathbb{R}$)

Throughout we confine attention to the **Lebesgue measure** $\mu\colon \mathcal{M} \to [0,\infty]$, introduced in Section 10. Given $A \in \mathcal{M}$, we consider the σ-algebra of all Lebesgue measurable subsets of A,

$$\mathcal{M}_A = \{E \in \mathcal{M} \,:\, E \subset A\}, \text{ and Lebesgue measure } \mu|_{\mathcal{M}_A}\colon \mathcal{M}_A \to [0,\infty].$$

We write $L(A) = L(A, \mathcal{M}_A, \mu)$ for the space of all **Lebesgue integrable** functions $f\colon A \to \mathbb{R}$. Further, for $A \in \mathcal{M}$, we write

$$\int_A f\,d\mu = \int_A f\,dx = \int_A f(x)\,dx,$$

and in the case of intervals (for $a, b \in \mathbb{R}$ with $a < b$)

$$\int_{[a,b]} f\,d\mu = \int_a^b f\,dx, \qquad \int_{[a,\infty)} f\,d\mu = \int_a^\infty f\,dx$$

$$\int_{(-\infty,a]} f\,d\mu = \int_{-\infty}^a f\,dx, \qquad \int_{\mathbb{R}} f\,d\mu = \int_{-\infty}^\infty f\,dx.$$

11.1 remark. With respect to Lebesgue measure, μ, every subset of a null set in $\mathcal{M}$ again lies in $\mathcal{M}$ (Proposition 10.11). As a result if $A \in \mathcal{M}$ and $f, g\colon A \to \mathbb{R}$ are functions such that $f = g$ (a.e) we have

(a) f is measurable $\iff g$ is measurable [See Exercise 8],

(b) $f \in L(A) \iff g \in L(A)$ [By (a) and Proposition 9.6].

Furthermore, suppose that $B = A \cup N$, where N is a null set, $\mu(N) = 0$. Then

(c) $f \in L(B) \iff f\chi_A \in L(A)$ [By (b)].

Hence we identify the spaces $L(B)$ and $L(A)$. For example, we need not distinguish between $L([a,b])$ and $L((a,b))$, or between the integrals $\int_{[a,b]} f\,d\mu$ and $\int_{(a,b)} f\,d\mu$.

11.1 *The Lebesgue versus the Riemann integral*

For a Riemann integrable (hence bounded) function $f\colon [a,b] \to \mathbb{R}$ we (temporarily, see Theorem 11.2) denote the Riemann integral of f over $[a,b]$ by

$$R - \int_a^b f\,dx.$$

We denote the space of all Riemann integrable functions $f\colon [a,b] \to \mathbb{R}$ by $R([a,b])$.

Step functions. By a step function, $h\colon [a,b] \to \mathbb{R}$ is meant a function of the form $h = \sum a_k \chi_{J_k}$, where each $a_k \in \mathbb{R}$ and each J_k is a subinterval of $[a,b]$. In this case the Lebesgue and Riemann integrals of h clearly coincide

$$(6) \quad \int h\,dx = \sum a_k \mu(J_k) = \sum a_k L(J_k) = R - \int_a^b h\,dx.$$

11.2 theorem. *Let $f \in R([a,b])$. Then $f \in L([a,b])$ and $R - \int_a^b f\,dx = \int_a^b f\,dx$.*

Proof. By the definition of the Riemann integral we can choose sequences $(t_n) \nearrow$ and $(u_n) \searrow$ of step functions such that $-K \le t_n \le f \le u_n \le K$, where $K = \sup |f|$, and such that, using (6),

$$(7) \quad R - \int_a^b f\,dx = \lim_{n\to\infty} \int_a^b t_n\,dx = \lim_{n\to\infty} \int_a^b u_n\,dx.$$

The measurable functions $t = \lim t_n$ and $u = \lim u_n$ satisfy

$$(8) \quad t \le f \le u.$$

On the other hand, since $|t_n|, |u_n| \le K\chi_{[a,b]} \in L([a,b])$, the DCT gives us that $t, u \in L([a,b])$, $\int t_n\,dx \to \int t\,dx$, and $\int u_n\,dx \to \int u\,dx$. By (7) we conclude that

$$\int_a^b t\,dx = \int_a^b u\,dx.$$

Thus, $0 \le u - t$ with $\int (u-t)\,dx = 0$. Hence $t = u$ (a.e) so that $f = u$ (a.e) by (8). Hence $f \in L([a,b])$ (Remark 11.1) and

$$\int_a^b f\,dx = \int_a^b u\,dx = R - \int_a^b f\,dx. \quad \square$$

The converse of Theorem 11.2 is **false**. That is, $R([a,b]) \subset L([a,b])$ but $R([a,b]) \ne L([a,b])$.

11.3 example. Consider $E = [a,b] \cap \mathbb{Q}$. If f denotes the characteristic function of E, then f is Lebesgue integrable (with integral $\mu(E) = 0$.) But f is not Riemann integrable.

11.2 *Improper Riemann integrals*

Given $a \in \mathbb{R}$, recall that $f\colon [a,\infty) \to \mathbb{R}$ is said to be **improperly Riemann integrable** if f is Riemann integrable on $[a,b]$ for all real numbers $b > a$ and $\lim_{b\to\infty} \int_a^b f\,dx$ exists in $\mathbb{R}$, in which case we write

$$R - \int_a^\infty f\,dx = \lim_{b\to\infty} \int_a^b f\,dx.$$

The space of all such functions is denoted $R([a,\infty))$.

Recall from Proposition 9.2 the equivalences

$$f \in L([a,\infty)) \iff |f| \in L([a,\infty)) \iff f_+, f_- \in L([a,\infty)). \tag{9}$$

These are not valid, in general, for $R([a,\infty))$.

11.4 example. Define $f\colon [0,\infty) \to \mathbb{R}$ piecewise by: $f(x) = (-1)^n/n$, whenever $x \in [n-1, n)$ and $n \in \mathbb{N}$. Then, for every $n \in \mathbb{N}$,

$$\int_0^n f\,dx = \sum_{k=1}^n \frac{(-1)^k}{k} \to \sum_{k=1}^\infty \frac{(-1)^k}{k} \quad \text{[Alternating harmonic series]},$$

but

$$\int_0^n |f|\,dx = \sum_{k=1}^n \frac{1}{k},$$

which diverges (harmonic series). Thus, f belongs to $R([a,\infty))$, but $|f|$ does not (neither does f_+, nor f_-).

11.5 theorem. *Let $a \in \mathbb{R}$ and let $f\colon [a,\infty) \to \mathbb{R}$ be Riemann integrable on $[a,b]$ for all real $b > a$. Then $f \in L([a,\infty))$ if and only if $|f| \in R([a,\infty))$, in which case also $f \in R([a,\infty))$ and*

$$\int_a^\infty f\,dx = R - \int_a^\infty f\,dx.$$

Proof. (a): Suppose first that $f \geq 0$. Since $f\chi_{[a,a+n]} \nearrow f$, the MCT and our assumptions give

$$\int_a^{a+n} f\,dx = \int_{[a,\infty)} f\chi_{[a,a+n]}\,d\mu \to \int_{[a,\infty)} f\,d\mu.$$

Thus, since $f \in L([a,\infty)) \iff \int_{[a,\infty)} f\,d\mu < \infty$, the desired result is immediate from the definitions and Theorem 11.2.

(b): For a general f, this follows by (9) and (a). For

$$f \in L([a,\infty)) \iff |f|, f_+, f_- \in L([a,\infty)) \iff |f|, f_+, f_- \in R([a,\infty)),$$

and in this case $f = f_+ - f_- \in R([a,\infty))$ and

$$\int_a^\infty f\,dx = \int_a^\infty f_+\,dx - \int_a^\infty f_-\,dx = R - \int_a^\infty f_+\,dx - R - \int_a^\infty f_-\,dx = R - \int_a^\infty f\,dx. \quad \square$$

11.3 *Examples*

Let $\alpha > 1$ and consider $f\colon [1,\infty) \to \mathbb{R}$, $f(x) = 1/x^\alpha$. Using Theorem 11.2 we have

$$\int_1^b f\,dx = \frac{1 - 1/b^{\alpha-1}}{\alpha - 1} \to \frac{1}{\alpha - 1}, \quad b \to \infty.$$

In particular, by Theorem 11.5, $f \in L([1,\infty))$ (with $\int_1^\infty f\,dx = 1/(\alpha-1)$).

This may be exploited to derive certain limits via the dominated convergence theorem.

11.6 Example. $\lim_{n\to\infty} \int_1^\infty \frac{x^{1/2}\sin(nx)}{1+nx^4}\,dx = 0.$

For each $n \in \mathbb{N}$, define $f_n\colon [1,\infty) \to \mathbb{R}$ by

$$f_n(x) = \frac{x^{1/2}\sin(nx)}{1+nx^4}.$$

Then $f_n \to 0$ (pointwise) [for each $x \geq 1$, $|f_n(x)| \leq x^{1/2}/(1+nx^4) \to 0$, as $n \to \infty$.] Furthermore,

$$|f_n(x)| \leq \frac{1}{x^{7/2}}, \quad x \geq 1.$$

Hence, with $f\colon [1,\infty) \to \mathbb{R}$ given by $f(x) = 1/x^{7/2}$, we have

$$|f_n| \leq f,\ \forall n \in \mathbb{N},\ \text{and}\ f \in L([1,\infty))\ \text{[because } 7/2 > 1\text{]}.$$

Therefore $f_n \in L([1,\infty))$, and since $f_n \to 0$, the DCT implies that

$$\int_1^\infty f_n\,dx \to \int_1^\infty 0\,dx = 0,$$

as required.

11.7 example. $\lim_{n\to\infty} \int_1^\infty \frac{1}{(1+x/n)^n x^{1/n}}\,dx = 1/e.$

For each $n \in \mathbb{N}$, define $f_n \colon [1,\infty) \to \mathbb{R}$ by

$$f_n(x) = \frac{1}{(1+x/n)^n x^{1/n}}.$$

For each $x \geq 1$, we have

$$(1+x/n)^n \to e^x,\ x^{1/n} \to 1, \quad n \to \infty,$$

so that $f_n(x) \to 1/e^x = e^{-x}$ as $n \to \infty$ (pointwise).

Furthermore, for $x \geq 1$ and $n \geq 2$,

$$(1+x/n)^n x^{1/n} \geq (1+x/n)^n \geq 1 + \frac{x}{n} + \frac{n(n-1)}{2}\left(\frac{x}{n}\right)^2 \geq \frac{x^2}{4},$$

giving

$$|f_n(x)| \leq \frac{4}{x^2}.$$

Hence, with $f \colon [1,\infty) \to \mathbb{R}$ given by $f(x) = 4/x^2$, we have

$$|f_n| \leq f,\ \forall n \geq 2,\ \text{and } f \in L([1,\infty))\ \text{[because } 2 > 1\text{]}.$$

Therefore $f_n \in L([1,\infty))$, $n \geq 2$, and since $f_n(x) \to e^{-x}$, the DCT implies that

$$\int_1^\infty f_n\,dx \to \int_1^\infty e^{-x}\,dx = 1/e,$$

where the last equality is easy to check (calculate as an improper Riemann integral).

11.4 *Density of continuous functions*

11.8 theorem. *Let $f \in L(\mathbb{R})$. For every $\varepsilon > 0$ there is a continuous function $g \colon \mathbb{R} \to \mathbb{R}$ vanishing outside a bounded interval such that*

$$\int_{\mathbb{R}} |f - g|\,dx < \varepsilon.$$

Proof. This may be seen via a sequence of reductions.

1. Since $f = f_+ - f_-$ and $f_+, f_- \in L(\mathbb{R})$, we may suppose $f \geq 0$.

2. By Theorem 8.7 and the MCT there is a simple function $\rho \in L(\mathbb{R})$ with $0 \leq \rho \leq f$ and $\int_{\mathbb{R}} |f-\rho|\,dx = \int_{\mathbb{R}}(f-\rho)\,dx < \varepsilon$. Thus it is enough to establish the theorem for simple functions of the form $\sum_{i=1}^n \alpha_i \chi_{E_i}$, where each $\mu(E_i) < \infty$,

and hence, in turn, for integrable characteristic functions χ_E. This may be done as follows.

3. **Claim.** Let $E \in \mathcal{M}$ with $\mu(E) < \infty$. Then there is a function $h = \sum_{i=1}^m \chi_{(a_i,b_i)}$ such that $\int_{\mathbb{R}} |\chi_E - h|\, dx < \varepsilon$.

To see this, pick an open set U with $E \subset U$ and $\mu(U) < \mu(E) + \varepsilon/2 < \infty$ (see Exercise 8). Since $U = \cup(a_n b_n)$ is a **disjoint** union of a sequence (possibly finite) of bounded open intervals (why?),

$$\mu(U) = \sum_n (b_n - a_n),$$

there is an $m \in \mathbb{N}$ such that $\mu(U) - \mu(V) < \varepsilon/2$, where $V = \cup_{n=1}^m (a_n, b_n)$. This gives $\chi_V = \sum_{i=1}^m \chi_{(a_i,b_i)}$ and

$$\int_{\mathbb{R}} |\chi_E - \chi_V|\, dx \leq \int_{\mathbb{R}} |\chi_E - \chi_U|\, dx + \int_{R} |\chi_U - \chi_V|\, dx = \mu(U \setminus E) + \mu(U \setminus V) < \varepsilon,$$

proving the claim.

4. Hence it is enough to prove the theorem for a characteristic function $\chi_{(a,b)}$, where $a, b \in \mathbb{R}$ with $a < b$. To do this, for $\delta < \varepsilon/2 < (b-a)/4$, let

$$g(x) = \begin{cases} 0, & x \in \mathbb{R} \setminus (a,b), \\ \dfrac{x-a}{\delta}, & x \in [a, a+\delta], \\ 1, & x \in [a+\delta, b-\delta], \\ \dfrac{b-x}{\delta}, & x \in [b-\delta, b]. \end{cases}$$

Then $g \colon \mathbb{R} \to \mathbb{R}$ is continuous (draw a picture!), and

$$\int_{\mathbb{R}} |\chi_{(a,b)} - g|\, dx \leq \int_{[a,a+\delta]\cup[b-\delta,b]} 1\, d\mu = 2\delta < \varepsilon.$$

This completes the proof. □

11.5 *A word about the Lebesgue integral in higher dimensions*

The next natural step is to construct the σ-algebra $\mathcal{M}_d \subset \mathbb{P}(\mathbb{R}^d)$ of Lebesgue measurable subsets of $\mathbb{R}^d$, $d > 1$, and the corresponding d-dimensional Lebesgue measure μ_d. The construction closely follows the 1-dimensional setting of Section 10, guided by the fact that μ_2 should generalise the notion of area, μ_3 that of volume, and so on. There is an additional difficulty compared to Section 10 in that open sets $U \subset \mathbb{R}^d$ are significantly more complicated in higher dimensions. In particular, there is no analogue of the property that open sets $U \subset \mathbb{R}$ are a disjoint union of open intervals. Due to time constraints, we will not consider any further details, referring instead to Tao's "An Introduction to Measure Theory".

The most important result for the higher-dimensional Lebesgue measures is the Fubini–Tonelli theorem, which generalises **iterated integration** to the context of Lebesgue integration. We conclude this section by giving a statement in the special case that $d = 2$.

11.9 theorem (Fubini–Tonelli's theorem). *Let $f : \mathbb{R}^2 \to [0, \infty]$ be a non-negative μ_2-Lebesgue measurable function (measurable with respect to $\mathcal{M}_2$). Then*

(a) For μ_1-almost every $x \in \mathbb{R}$, the function $y \mapsto f(x, y)$ is μ_1-Lebesgue measurable, and in particular $\int_{\mathbb{R}} f(x, y)\, dy$ is defined (possibly infinite) almost everywhere with respect to x. Furthermore, the map $x \mapsto \int_{\mathbb{R}} f(x, y)\, dy$ is μ_1-Lebesgue measurable.

(b) For μ_1-almost every $y \in \mathbb{R}$, the function $x \mapsto f(x, y)$ is μ_1-Lebesgue measurable, and in particular $\int_{\mathbb{R}} f(x, y)\, dx$ is defined (possibly infinite) almost everywhere with respect to y. Furthermore, the map $y \mapsto \int_{\mathbb{R}} f(x, y)\, dx$ is μ_1-Lebesgue measurable.

(c) We have

$$\int_{\mathbb{R}^2} f\, d\mu_2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, dx \right) dy. \tag{10}$$

11.10 remark. Note that f is non-negative in the statement. For general functions f a similar statement can be made, if it is known beforehand that $f \in L(\mathbb{R}^2, \mathcal{M}_2, \mu_2)$. However, if $\int_{\mathbb{R}^2} |f|\, d\mu_2 = \infty$, then (10) is **not** necessarily true. A counterexample is furnished by

$$f(x, y) = \frac{x^2 - y^2}{(x^2 + y^2)^2} \chi_{(0,1)\times(0,1)}(x, y).$$

Noting that $\frac{x^2 - y^2}{(x^2+y^2)^2} = -\frac{\partial^2}{\partial x \partial y} \arctan(y/x)$, one sees that

$$\int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2}\, dy \right) dx = \frac{\pi}{4},$$

while

$$\int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2}\, dx \right) dy = -\frac{\pi}{4}.$$

Of course, in this case

$$\int_{\mathbb{R}^2} |f|\, d\mu_2 = \int_0^1 \left(\int_0^1 \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| dy \right) dx = \int_0^1 \left(\int_0^1 \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| dx \right) dy = \infty.$$