Integration Theory

Hans Reijnierse

(edited by Edwin van Dam)

January 2017

Contents

1 Riemann integration

This chapter provides an alternative introduction of the Riemann integral in such a way that we do not have to restrict ourselves to continuous functions.

1.1 Introduction

Suppose, we would like to know D; the area of the region of the plane bounded from below by the x-axis, from above by the graph of the function $x \mapsto \sqrt{x}$ and from the right by the line $x = b$, with $b > 0$.

The formulation of the problem appears to be deceivingly accurate, but it is not, since we have no formal definition of the notion *area*. This leaves some room for interpretation.

Choose a positive integer n. Take an arbitrary sequence of real numbers $a_0, a_1, \ldots, a_n$ that satisfies $a_0 = 0 < a_1 < a_2 < \ldots < a_n = b$. Consider the rectangles ($k \in \{1, \ldots, n\}$)

$$[a_{k-1}, a_k] \times [0, \sqrt{a_k}].$$

D 'obviously' cannot exceed the sum of the areas of the rectangles (draw a figure). The area of a rectangle is unambiguously equal to the product of its width and height, so

$$D \leqslant \sum_{k=1}^{n} (a_k - a_{k-1}) \sqrt{a_k}.$$

Now we choose the reals a_k in a specific manner, such that the right part of the expression above can be simplified: let a_k be $(\frac{k}{n})^2 b$. The requirement that $a_0 = 0 < a_1 < \ldots < a_n = b$ remains valid. We find

$$\begin{aligned}
\sum_{k=1}^{n} (a_k - a_{k-1}) \sqrt{a_k} &= \sum_{k=1}^{n} \left(k^2 \tfrac{b}{n^2} - (k-1)^2 \tfrac{b}{n^2} \right) \cdot \tfrac{k}{n} \sqrt{b} \\
&= \sum_{k=1}^{n} (2k-1) \tfrac{b}{n^2} \cdot \tfrac{k}{n} \cdot \sqrt{b} \\
&= \tfrac{2b\sqrt{b}}{n^3} \sum_{k=1}^{n} k^2 - \tfrac{b\sqrt{b}}{n^3} \sum_{k=1}^{n} k \\
&= \tfrac{2b\sqrt{b}}{n^3} \cdot \tfrac{1}{6} n(n+1)(2n+1) - \tfrac{b\sqrt{b}}{n^3} \cdot \tfrac{1}{2} n(n+1).
\end{aligned}$$

We have used that $\sum_{k=1}^{n} k^2 = \frac{1}{6} n(n+1)(2n+1)$ and $\sum_{k=1}^{n} k = \frac{1}{2} n(n+1)$, which both can be shown by induction. Hence,

$$D \leqslant \frac{2b\sqrt{b}}{6} \cdot \frac{n(n+1)(2n+1)}{n^3} - \frac{b\sqrt{b}}{2} \cdot \frac{n(n+1)}{n^3}$$

for all $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$. Let n tend to infinity to obtain

$$D \leqslant \frac{2b\sqrt{b}}{6} \cdot 2 = \frac{2}{3} b\sqrt{b}.$$

D exceeds (draw again a figure) the sum of the areas of the triangles $[a_{k-1}, a_k] \times [0, \sqrt{a_{k-1}}]$. Hence,

$$D \geqslant \sum_{k=1}^{n} (a_k - a_{k-1})\sqrt{a_{k-1}}.$$

Substituting $a_k = (\frac{k}{n})^2 b$, and using a little of algebra, results in

$$D \quad \geqslant \quad \frac{2b\sqrt{b}}{6} \cdot \frac{n(n+1)(2n+1)}{n^3} - \frac{3b\sqrt{b}}{2} \cdot \frac{n(n+1)}{n^3} + \frac{b\sqrt{b}}{n^2},$$

and therefore (let again n run to ∞), $\quad D \geqslant \dfrac{2b\sqrt{b}}{6} \cdot 2 = \dfrac{2}{3} b\sqrt{b}.$

D turns out to be $\frac{2}{3} b\sqrt{b}$.

Exercise 1 Determine in the same way the area of the region above the x-axis, below the graph of the function $x \mapsto 2x$ and left of the line $x = b$ with $b > 0$. $\circ$

1.2 Under- and overestimations

This section provides a general construction of which the procedure in Section 1.1 is a special case. f still represents a function and $[a, b]$ an interval in the domain of f. The previous section gives a pretty good clue how the notion *area* of the region of the plane between the x-axis and the graph of f can be formalized, at least if $f \geqslant 0$. The construction will be more general however, in the sense that functions can have negative values as well.

We start with an interval $[a, b]$. A *division* of $[a, b]$ is, by definition, a finite sequence of real numbers $(s_0, s_1, \ldots, s_n)$ with $\qquad s_0 = a < s_1 < s_2 < \cdots < s_n = b.$

The reals s_k are called *division points*. A division $(t_0, t_1, \ldots, t_m)$ is called a *refinement* of $(s_0, s_1, \ldots, s_n)$ when $\{s_0, s_1, \ldots, s_n\} \subseteq \{t_0, t_1, \ldots, t_m\}$. A simple, but useful fact is that every pair of divisions of $[a, b]$ share a common refinement. Just take two divisions $(s_0, \ldots, s_n)$ and $(t_0, \ldots, t_m)$ and consider $(u_0, \ldots, u_\ell)$ with $\{s_0, \ldots, s_n, t_0, \ldots, t_m\} = \{u_0, \ldots, u_\ell\}$. This is a division of $[a, b]$ and a refinement of both $(s_0, \ldots, s_n)$ and $(t_0, \ldots, t_m)$.

Definition 1.1 *A real number p is called an underestimation of f at $[a, b]$ if there exist a division $(s_0, \ldots, s_n)$ of $[a, b]$ and real numbers $\ell_1, \ldots, \ell_n$ such that*

- *for all $k \in \{1, \ldots, n\}$ and $x \in [s_{k-1}, s_k]$, we have $\ell_k \leqslant f(x)$,*
- $p = (s_1 - s_0)\ell_1 + (s_2 - s_1)\ell_2 + \cdots + (s_n - s_{n-1})\ell_n$.

The set of all underestimations is denoted by P.

Likewise, a real number q *overestimates* f *at* $[a,b]$ if there exist a division $(s_0,\ldots,s_n)$ of $[a,b]$ and real numbers $h_1,\ldots,h_n$ such that

- for all $k \in \{1,\ldots,n\}$ and $x \in [s_{k-1}, s_k]$, we have $f(x) \leqslant h_k$,
- $q = (s_1 - s_0)h_1 + (s_2 - s_1)h_2 + \cdots + (s_n - s_{n-1})h_n$.

The set of all overestimations is denoted by Q.

We say that underestimation p and overestimation q are (can be) made with the division $(s_0, s_1, \ldots, s_n)$ and the respective sequences $(\ell_1,\ldots,\ell_n)$ and $(h_1,\ldots,h_n)$.
Suppose, we have made an underestimation p with division $(s_0, s_1, \ldots, s_n)$ and sequence $(\ell_1,\ldots,\ell_n)$. Choose a real number s^* with $s_2 < s^* < s_3$ (assuming that $n \geqslant 3$). Then $(s_0, s_1, s_2, s^*, s_3, \ldots, s_n)$ is another division of $[a,b]$. The new division and the sequence $(\ell_1, \ell_2, \ell_3, \ell_3, \ell_4, \ldots, \ell_n)$ define an underestimation. Of course, the new underestimation is nothing but the one we already had: p. Carrying out this construction repeatedly results in

Lemma 1.2 *Let* $(s_0,\ldots,s_n)$ *be a division of* $[a,b]$ *and let* $(t_0,\ldots,t_m)$ *be a refinement of* $(s_0,\ldots,s_n)$*. If an underestimation* p *(overestimation* q*) of* f *on* $[a,b]$ *can be made with division* $(s_0,\ldots,s_n)$*, then* p *(*q*) can be made as well with division* $(t_0,\ldots,t_m)$*.*

Assume for the time being that $f(x) \geqslant 0$ for all x in $[a,b]$. Let D be the area of the region enclosed by the lines $x = a$, $x = b$, the x-axis and the graph of f. Then D is at least equal to any underestimation of f on $[a,b]$ and at most equal to any overestimation of f on $[a,b]$.
When we can find sequences $p_1, p_2, p_3, \ldots$ and $q_1, q_2, \ldots$ of under- and overestimations that both converge and have the same limit, then this limit has got to be D. Actually, this line of thoughts is nothing but the technique that we have used already in Section 1.1. We had an intuitive idea of the concept called ‘area’ on which we elaborated. We could turn things around though, by providing a formal definition by means of under- and overestimations. Something like

> *If there exists a real number that is both the limit of a sequence of underestimations and the limit of a sequence of overestimations of* f *on* $[a,b]$*, then* D *equals this number by definition.*

Two questions arise:

I Does such a number always exist?

II Can there be more of these numbers?

I Define $f : [0,1] \to \mathbb{Q}$ by $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0,1], \\ 0 & \text{if } x \in [0,1] \backslash \mathbb{Q}. \end{cases}$

All underestimations are at most 0 and all overestimations are at least 1 (why is this exactly the case?). So there can be no real number simultaneously be the limit of underestimations as well as the limit of overestimations. Our method does not bring us any further.

II The answer of the second question is, fortunately, also negative and can be given by means of the following theorem and its corollary:

Theorem 1.3 *Let $[a,b]$ be an interval inside the domain of a function f. Let p be an underestimation and q be an overestimation of f on $[a,b]$. Then $p \leqslant q$.*

Proof The underestimation p and the overestimation q both are defined by means of a division. The two divisions involved share a common refinement $(s_0, \ldots, s_n)$. Lemma 1.2 gives that p and q can both be defined by means of division $(s_0, \ldots, s_n)$, *i.e.*, there exist real numbers $\ell_1, \ldots, \ell_n$ and $h_1, \ldots, h_n$ such that

for all $k \in \{1, \ldots, n\}$ and $x \in [s_{k-1}, s_k]$, we have $\ell_k \leqslant f(x) \leqslant h_k$,

$p = \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k$ and $q = \sum_{k=1}^{n} (s_k - s_{k-1})h_k$.

In particular, for each k we have that $\ell_k \leqslant h_k$, so

$$p = \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k \leqslant \sum_{k=1}^{n} (s_k - s_{k-1})h_k = q. \qquad \square$$

There may not be a highest underestimation (lowest overestimation), but the supremum of all underestimations, *i.e.*, $\sup_{p \in P} p$ (see the Appendix) and the infimum of all overestimations, $\inf_{q \in Q} q$, are well defined. They will play such a prominent role that we provide them with the names *under-Riemann integral* and *upper-Riemann integral* respectively. They are denoted by the respective symbols

$$\underline{\int_a^b} f(x)\,\mathrm{d}x = \sup_{p \in P} p \quad \text{and} \quad \overline{\int_a^b} f(x)\,\mathrm{d}x = \inf_{q \in Q} q.$$

The previous theorem implies

Corollary 1.4 *Each bounded function obeys* $\underline{\int_a^b} f(x)\,\mathrm{d}x \leqslant \overline{\int_a^b} f(x)\,\mathrm{d}x$.

This corollary enables us to answer question II. Suppose there exist sequences of under- and overestimations $p_1, p_2, \ldots$ and $q_1, q_2, \ldots$ converging to the same limit, say α. Then

$$\alpha = \lim_{n \to \infty} p_n \leqslant \underline{\int_a^b} f(x)\,\mathrm{d}x \leqslant \overline{\int_a^b} f(x)\,\mathrm{d}x \leqslant \lim_{n \to \infty} q_n = \alpha.$$

Hence α equals $\underline{\int_a^b} f(x)\,\mathrm{d}x$, making it uniquely determined.

1.3 Integration over intervals; the Fundamental Theorem of Calculus

The preparations in the previous section enable us to provide a formal definition of the notion area, as proposed in Section 1.2. To avoid the suggestion of non-negativity, we will not use the word area, but the neutral word *integral*.

Definition 1.5 *A bounded function* $f : [a,b] \longrightarrow \mathbb{R}$ *is called Riemann integrable over* $[a,b]$ *if*

$$\underline{\int_a^b} f(x)\,\mathrm{d}x = \overline{\int_a^b} f(x)\,\mathrm{d}x.$$

This value is called the Riemann integral *of* f *over* $[a,b]$*, and is denoted by* $$\int_a^b f(x)\,\mathrm{d}x.$$

The *integration variable* x can be substituted by any other character; $\int_a^b f(t)\,\mathrm{d}t$ and $\int_a^b f(w)\,\mathrm{d}w$ mean exactly the same. Note that for each underestimation p and each overestimation q of f on $[a,b]$ we have $p \leqslant \int_a^b f(x)\,\mathrm{d}x \leqslant q$.

An *indicator function* is a 0-1-function that tells which elements of its domain are inside some given set A, so

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

Exercise 2

a) Determine $\underline{\int_0^2} \mathbb{1}_{\{1\}}(x)\,\mathrm{d}x$ and $\overline{\int_0^2} \mathbb{1}_{\{1\}}(x)\,\mathrm{d}x$.

b) Show that $P = \big(-\infty, \underline{\int_0^2} \mathbb{1}_{\{1\}}(x)\,\mathrm{d}x\big]$ and that $Q = \big(\overline{\int_0^2} \mathbb{1}_{\{1\}}(x)\,\mathrm{d}x, \infty\big)$. $\circ$

Already at secondary school Riemann integrals are computed by means of antiderivatives. Let F be an antiderivative of f. Then $\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$. Unfortunately this does not work in general. There exist Riemann integrable functions that do not have an antiderivative. The following exercise provides an example.

Exercise 3 Reconsider the function $\mathbb{1}_{\{1\}}$ of Exercise 2. Prove that this function does not have an antiderivative (you may use the results of the course *Mathematical Analysis* that any differentiable function is continuous and that all antiderivatives of the zero-function are constant). $\circ$

Exercise 4 Let $a < b$. Show that there is no segment $[a,b]$ such that the function $\mathbb{1}_{\mathbb{Q}}$ is Riemann integrable over $[a,b]$. $\circ$

The next exercise shows how one can proceed when no antiderivative is available.

Exercise 5 Let g be the indicator function of $\left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\right\}$. Show that g is Riemann integrable over $[0,1]$ and determine $\int_0^1 g(x)\,\mathrm{d}x$.

Hint. Fix n and use for an overestimation the division with 2^n subintervals of equal size (instead of n division points). Given n, how many subintervals contain real numbers x with $g(x) = 1$? ◦

The following theorem tells that if a function does have an antiderivative, then we can use it to determine integrals of f:

Theorem 1.6 [Fundamental Theorem of Calculus]

Let f be a Riemann integrable function over $[a, b]$ with antiderivative F. Then

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

Proof Let p be an underestimation of f on $[a, b]$. Choose $s_0, \ldots, s_n$ and $\ell_1, \ldots, \ell_n$ as in the definition of underestimation. For each k in $\{1, \ldots, n\}$ there exists, according to the *Mean Value Theorem* (Middelwaardestelling), a real number α_k in (s_{k-1}, s_k) with

$$F(s_k) - F(s_{k-1}) = F'(\alpha_k)(s_k - s_{k-1}) = f(\alpha_k)(s_k - s_{k-1}).$$

Hence,

$$\ell_k(s_k - s_{k-1}) \leqslant F(s_k) - F(s_{k-1}).$$

Aggregate over all k's and find that $p \leqslant F(b) - F(a)$.

In the same way you can show that $F(b) - F(a) \leqslant q$ for every overestimation q. Because f is Riemann integrable, we can choose under- and overestimations $p_1, p_2, \ldots$ and $q_1, q_2, \ldots$ such that $\lim_{i\to\infty} p_i = \lim_{i\to\infty} q_i = \int_a^b f(x)\,\mathrm{d}x$. Because $p_i \leqslant F(b) - F(a) \leqslant q_i$ for all i, $F(b) - F(a)$ has got to be equal to $\int_a^b f(x)\,\mathrm{d}x$. □

1.4 Integrability criteria

Theorem 1.7 *Every increasing (or decreasing) function with a compact domain is Riemann integrable.*

Proof Let f be increasing on $[a, b]$. Fix a natural number n in $\mathbb{N}$ and consider the division $(s_0, \ldots, s_n)$ of $[a, b]$ in n subintervals of equal length. Then for each k in $\{0, 1, \ldots, n\}$ we have $s_k = a + k \cdot \frac{b-a}{n}$ and if $x \in [s_{k-1}, s_k]$, $f(s_{k-1}) \leqslant f(x) \leqslant f(s_k)$. Define an underestimation p_n and an overestimation q_n by

$$
\begin{array}{rclcl}
p_n & = & \sum_{k=1}^{n}(s_k - s_{k-1})f(s_{k-1}) & = & \sum_{k=1}^{n}\frac{b-a}{n}f(s_{k-1}). \\
q_n & = & \sum_{k=1}^{n}(s_k - s_{k-1})f(s_k) & = & \sum_{k=1}^{n}\frac{b-a}{n}f(s_k),
\end{array}
$$

This results for every n in $\mathbb{N}$ in

$$
\begin{array}{rcl}
0 & \leqslant & q_n - p_n \\
& = & \frac{b-a}{n}\big[f(s_1) + f(s_2) + \cdots + f(s_n)\big] - \frac{b-a}{n}\big[f(s_0) + f(s_1) + \cdots + f(s_{n-1})\big] \\
& = & \frac{b-a}{n}\big[f(s_n) - f(s_0)\big] \\
& = & \frac{b-a}{n}\big[f(b) - f(a)\big].
\end{array}
$$

Hence, $\lim\limits_{n\to\infty}(q_n - p_n) = 0$. Since

$$
0 \leqslant \overline{\int_a^b} f(x)\,\mathrm{d}x - \underline{\int_a^b} f(x)\,\mathrm{d}x \leqslant q_n - p_n,
$$

we can conclude that $\underline{\int_a^b} f(x)\,\mathrm{d}x = \overline{\int_a^b} f(x)\,\mathrm{d}x.$ □

‘Combinations’ of Riemann integrable functions are integrable itself, as the following theorem shows.

Theorem 1.8 *Let f and g Riemann integrable are over $[a, b]$. Let $c \in (a, b)$.*

(i) *The sum-function $f + g$ is Riemann integrable over $[a, b]$ and*

$$
\int_a^b f(x) + g(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x.
$$

(ii) *For each c in $\mathbb{R}$, the function $c \cdot f$ is Riemann integrable over $[a, b]$ and*

$$
\int_a^b cf(x)\,\mathrm{d}x = c\int_a^b f(x)\,\mathrm{d}x.
$$

(iii) *If $f(x) \geqslant 0$ for all x in $[a, b]$, then* $\displaystyle\int_a^b f(x)\,\mathrm{d}x \geqslant 0.$

(iv) *If $f(x) \leqslant g(x)$ for all x in $[a, b]$, then* $\displaystyle\int_a^b f(x)\,\mathrm{d}x \leqslant \int_a^b g(x)\,\mathrm{d}x.$

(v) $\displaystyle\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x.$

(vi) *The product-function fg is Riemann integrable over $[a, b]$.*

Proof The first five statements are straightforward, so we focus on the sixth one. In order to prove statement (vi), first assume that $f = g$ and that $f \geqslant 0$. The function f is bounded on $[a, b]$, so there exist a number M with $0 \leqslant f(x) < M$ for $x \in [a, b]$. Let $\varepsilon > 0$ and let p and q be an underestimation and an overestimation of $\int_a^b f(x)\,\mathrm{d}x$, constructed with the same division s using, say, n subsegments such that $q - p < \varepsilon$. So there are real numbers $\ell_1, \ldots, \ell_n$ and $h_1, \ldots, h_n$ such that

for all $k \in \{1, \ldots, n\}$ and $x \in [s_{k-1}, s_k]$, we have $0 \leqslant \ell_k \leqslant f(x) \leqslant h_k \leqslant M$,
$p = \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k$ and $q = \sum_{k=1}^{n} (s_k - s_{k-1})h_k$.

Let $\bar{p}$ and $\bar{q}$ be defined as $\bar{p} = \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k^2$ and $\bar{q} = \sum_{k=1}^{n} (s_k - s_{k-1})h_k^2$ respectively. Then $\bar{p}$ is less than $\underline{\int}_a^b f^2(x)\,\mathrm{d}x$ and $\bar{q}$ exceeds $\overline{\int}_a^b f^2(x)\,\mathrm{d}x$. We are done if we can show that $\bar{q} - \bar{p}$ tends to 0 when ε tends to 0. We have

$$\begin{aligned}
\bar{q} - \bar{p} &= \sum_{k=1}^{n} (s_k - s_{k-1})(h_k^2 - \ell_k^2) \\
&= \sum_{k=1}^{n} (s_k - s_{k-1})(h_k + \ell_k)(h_k - \ell_k) \\
&\leqslant 2M \sum_{k=1}^{n} (s_k - s_{k-1})(h_k - \ell_k) \\
&= 2M(q - p) \\
&\leqslant 2M\varepsilon.
\end{aligned}$$

Indeed, when ε tends to 0, then $\bar{q} - \bar{p}$ tends to 0 as well.

Now we keep the assumption that $f = g$, but drop the assumption that $f \geqslant 0$. Still, there is a number M with $|f(x)| < M$ for $x \in [a, b]$. The function h defined by $h(x) = f(x) + M$ is nonnegative, so h^2 is Riemann integrable by the reasoning above. Assertions (i) and (ii) tell that the right hand side of the equation

$$f^2 = h^2 - 2fM - M^2$$

is Riemann integrable. So f^2 itself is Riemann integrable. What if f and g do not coincide? Then we can decompose fg as follows:

$$fg = \tfrac{1}{4}\big((f+g)^2 - (f-g)^2\big).$$

By the rules of calculus (i) and (ii) and the proof so far, the right hand side is Riemann integrable. □

We end this section by a theorem that provides sufficient conditions for switching the order of a double Riemann integral without affecting the outcome. The proof is too difficult, but we will be able to prove the discrete variant of this theorem in Section 3.8.

Theorem 1.9 [Fubini's Theorem for Riemann integrals]
Let $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$. Let the function $f : [a, b] \times [c, d] \longrightarrow \mathbb{R}$ be such that the Riemann integrals $\int_{y=c}^{d} f(x, y)\,\mathrm{d}y$ exist for all $x \in [a, b]$ and $\int_{x=a}^{b} f(x, y)\,\mathrm{d}x$ exist for all $y \in [c, d]$. If $f \geqslant 0$ or if $\int_a^b \int_c^d |f(x, y)|\,\mathrm{d}y\,\mathrm{d}x < \infty$, then

$$\int_{x=a}^{b} \int_{y=c}^{d} f(x, y)\,\mathrm{d}y\,\mathrm{d}x = \int_{y=c}^{d} \int_{x=a}^{b} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

2 Stieltjes integration

2.1 Introduction

Let X be a random variable with a uniform distribution function F over the interval $[1,4]$, or $X \sim \mathrm{Un}([1,4])$ for short. This is its distribution function:

$$F(x) = \mathbb{P}[X \leqslant x] = \begin{cases} 0 & \text{if } x \leqslant 1, \\ \frac{1}{3}(x-1) & \text{if } 1 < x \leqslant 4, \\ 1 & \text{if } x > 4. \end{cases}$$

Recall that $F(x)$ denotes the probability that X realizes somewhere in $(-\infty, x]$, which gives in particular that $F(b) - F(a)$ is equal to the probability that X realizes in $(a, b]$.

Suppose that we would like to determine $\mathbb{E}[X^2]$, *i.e.*, the expectation of random variable X^2. We could proceed as follows. Choose a division $(s_0, \ldots, s_n)$ of $[1,4]$ and real numbers $\ell_1, \ldots, \ell_n$ and $h_1, \ldots, h_n$ in such a way that for each k in $\{1, \ldots, n\}$ and every x in $[s_{k-1}, s_k]$ we have $\ell_k \leqslant x^2 \leqslant h_k$. If X realizes in the interval $(s_{k-1}, s_k]$, which happens with probability $F(s_k) - F(s_{k-1})$, then the realization of X^2 is at least ℓ_k. This leads to the observation that

$$\begin{aligned} p &:= \ell_1\big(F(s_1) - F(s_0)\big) + \cdots + \ell_n\big(F(s_n) - F(s_{n-1})\big) \\ &\leqslant \mathbb{E}[X^2]. \end{aligned}$$

A similar argument gives

$$\begin{aligned} q &:= h_1\big(F(s_1) - F(s_0)\big) + \cdots + h_n\big(F(s_n) - F(s_{n-1})\big) \\ &\geqslant \mathbb{E}[X^2]. \end{aligned}$$

If we can find a division $(s_0, \ldots, s_n)$ and accompanying lower bounds $\ell_1, \ldots, \ell_n$ and upper bounds $h_1, \ldots, h_n$ such that p and q do not differ too much, then p and q are pretty good estimations of $\mathbb{E}[X^2]$.

The numbers p and q resemble under- and overestimations of Riemann integrals. There is a striking difference however: the length $s_k - s_{k-1}$ of interval $[s_{k-1}, s_k]$ has been replaced by $F(s_k) - F(s_{k-1})$; the probability that X realizes in $(s_{k-1}, s_k]$. This adaptation leads to Stieltjes integrals.

2.2 Integration over intervals with respect to an increasing function

Let g be an *increasing*[1] function on interval $[a,b]$. Let f be a *bounded* function on the same interval. A real number p is called an *underestimation of f with respect to g on* $[a,b]$ if

$$p = \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big)\ell_k$$

for some division $(s_0, s_1, \ldots, s_n)$ of $[a,b]$ and sequence $\ell_1, \ldots, \ell_n$ with $\qquad (k \in \{1,\ldots,n\})$

$$x \in [s_{k-1}, s_k] \implies \ell_k \leqslant f(x).$$

Because f is bounded, there exists a natural number M with $-M \leqslant f(x) \leqslant M$ for all $x \in [a,b]$. Hence, underestimations exist (choose, *e.g.*, $\ell_k = -M$ for all k) and underestimations are bounded from above by $M\big(g(b) - g(a)\big)$. Let P be the set of all underestimation of f with respect to g. We define

$$\underline{\int_a^b} f \,\mathrm{d}g = \sup_{p \in P} p.$$

This number is called the *under-Stieltjes integral* of f with respect to g. Similarly we define *overestimations* q, the set of all overestimations Q, and the *upper-Stieltjes integral*

$$\overline{\int_a^b} f \,\mathrm{d}g = \inf_{q \in Q} q.$$

The proofs of the following statements follow the ones in analogs in the Riemann chapter:

(i) Any under- or overestimation, defined with division s, can also be defined with any refinement of s (cf. Lemma 1.2).

(ii) $p \leqslant q$ for all underestimations p and overestimations q (cf. Theorem 1.3).

(iii) $\underline{\int_a^b} f \,\mathrm{d}g \leqslant \overline{\int_a^b} f \,\mathrm{d}g$ (cf. Corollary 1.4).

Definition 2.1 *A function f over $[a,b]$ is called Stieltjes integrable with respect to function g if $\underline{\int_a^b} f \,\mathrm{d}g = \overline{\int_a^b} f \,\mathrm{d}g$. If so, the real number $\overline{\int_a^b} f \,\mathrm{d}g$ is called the Stieltjes integral of f over $[a,b]$ and is denoted by*

$$\int_a^b f \,\mathrm{d}g \quad \text{or} \quad \int_a^b f(x) \,\mathrm{d}g(x).$$

The two functions f and g are called the integrand and the integrator respectively.

Exercise 6 Let f be a bounded function on interval $[a,b]$. The Stieltjes integral is a generalization of the Riemann integral. Show this by inferring that the Stieltjes integral $\int_a^b f \,\mathrm{d}g$ coincides with the Riemann integral $\int_a^b f(x) \,\mathrm{d}x$ in case g is the function $x \mapsto x$. ○

[1]A function g is called *increasing* if $x \leqslant y$ implies $g(x) \leqslant g(y)$. It is called *strictly increasing* if $x < y$ implies $g(x) < g(y)$.

Exercise 7 Consider a constant function g, *e.g.*, $g(x) = 37$ for every x in $[a, b]$. Show that every bounded function $f : [a, b] \longrightarrow \mathbb{R}$ is Stieltjes integrable with respect to g. Which number is $\int_a^b f \, \mathrm{d}g$? $\circ$

The following exercise shows that the expectation $\mathbb{E}[X]$ of a discrete random variable X can be considered to be a Stieltjes integral in which the distribution function F performs as the integrator.

Exercise 8 Let $X \sim \mathrm{Un}(\{1, 2, 3\})$. The distribution function F of X is thereby

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 1, \\ \frac{1}{3} & \text{if} \quad 1 \leqslant x < 2, \\ \frac{2}{3} & \text{if} \quad 2 \leqslant x < 3, \\ 1 & \text{if} \quad 3 \leqslant x. \end{cases}$$

We would like to compute the Stieltjes integral of the function $f(x) = x$ with respect to F over the interval $[0, 4]$. Let $n \in \mathbb{N}$ be a multiple of 4 and consider the division $(s_0, s_1, \ldots, s_n)$ of $[0, 4]$, given by $s_k = \frac{4k}{n}, k \in \{0, 1, 2, \ldots, n\}$.

a) Show that the greatest underestimation to be found with this division is $2 - \frac{4}{n}$.

b) Show that the least overestimation to be found with this division is 2.

c) Show that $\int_0^4 f \, \mathrm{d}F = 2$.

Note that $\mathbb{E}[X] = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 = 2$. The customary notation $\mathbb{E}[X] = \int x \, \mathrm{d}F(x)$ is justified here. $\circ$

It is possible to visualize Stieltjes integrals with areas. Drawing a graph of a function in order to visualize a Riemann integral can be done by choosing a number of values x in its domain, plotting the pairs $(x, f(x))$, and drawing a curve that connects the drawn points. In the case of a Stieltjes integral of f with respect to g not the choices of x, but their g-values matter. Hence, the pairs $\big(g(x), f(x)\big)$ should be plotted. If f is non-negative, the area of the region between the horizontal axis, the graph of f with respect to g, and the vertical lines $f(\cdot) = g(a)$ and $f(\cdot) = g(b)$ equals $\int f \, \mathrm{d}g$.

Example 1 Reconsider the functions f and F of the previous exercise. Take the division $s_0, s_1, \ldots, s_{16}$ of $[0, 4]$ into 16 equally sized subintervals. The range of f is $[0, 4]$ and the range of F is $[0, 1]$. The horizontal axis corresponds to F (rather than to x) and the vertical axis corresponds, as usual, to f. Plot the 16 pairs $\big(F(s_k), f(s_k)\big)$ (see Figure 1, it is quite possible that points are situated on the same vertical line). Connect the points by a curve and the result is the graph of f with respect to F. For each of the subintervals $[s_{k-1}, s_k]$ *with a positive weight* a column can be drawn with width $F(s_k) - F(s_{k-1})$ and

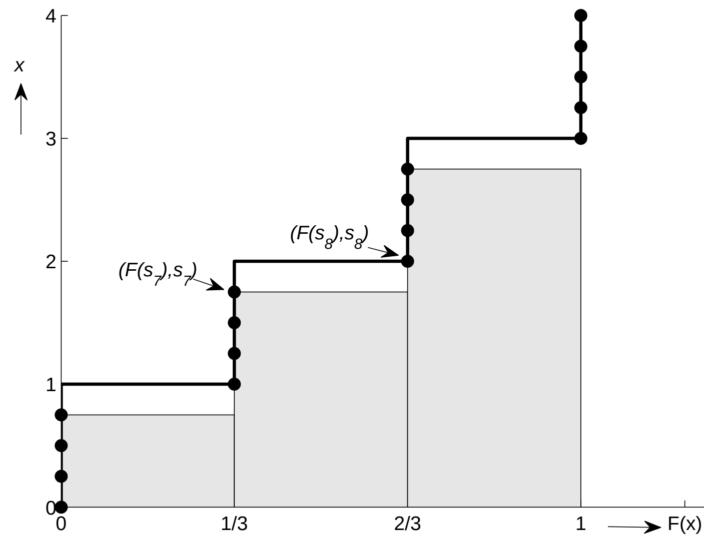


Figure 1: *The function $f(x) = x$ is drawn with respect to the distribution function F of Un$(\{1, 2, 3\})$. The light-grey area corresponds to an underestimation that uses the division into 16 equally sized subintervals of domain $[0, 4]$.*

height ℓ_k. Subintervals with nihil weight do not contribute to the integral and are not displayed. The sum of the areas of the drawn columns make an underestimation of $\int_0^4 x \, \mathrm{d}F$. $\triangle$

Stieltjes integrals inherit many properties of Riemann integrals. *E.g.*, Theorem 1.8 can easily be modified.

Theorem 2.2 *Let $g : [a, b] \longrightarrow \mathbb{R}$ be an increasing function and let f, f_1 and f_2 be Stieltjes integrable functions with respect to g.*

(i) *The sum-function $f_1 + f_2$ is Stieltjes integrable with respect to g over $[a, b]$ and*
$$\int_a^b (f_1 + f_2) \, \mathrm{d}g = \int_a^b f_1 \, \mathrm{d}g + \int_a^b f_2 \, \mathrm{d}g.$$

(ii) *For each c in $\mathbb{R}$, $c \cdot f$ is Stieltjes integrable with respect to g over $[a, b]$ and*
$$\int_a^b cf \, \mathrm{d}g = c \int_a^b f \, \mathrm{d}g,$$

(iii) *If f is nonnegative on $[a, b]$, then* $\displaystyle\int_a^b f \, \mathrm{d}g \geqslant 0.$

(iv) *If $f_1 \leqslant f_2$, then* $\displaystyle\int_a^b f_1 \, \mathrm{d}g \leqslant \int_a^b f_2 \, \mathrm{d}g.$

(v) $\displaystyle\int_a^b f(x) \, \mathrm{d}g = \int_a^c f(x) \, \mathrm{d}g + \int_c^b f(x) \, \mathrm{d}g.$

(vi) *The product-function* $f_1 \cdot f_2$ *is Stieltjes integrable with respect to* g *over* $[a, b]$.

The proof this theorem follows exactly the lines of the proof of Theorem 1.8. Two integrands can be aggregated as well (the proof is again straightforward):

Theorem 2.3 *Let* $g_1, g_2 : [a, b] \longrightarrow \mathbb{R}$ *be increasing functions and let* f *be Stieltjes integrable functions with respect to both* g_1 *and* g_2. *Then* f *is Stieltjes integrable with respect to* $g_1 + g_2$ *over* $[a, b]$ *and*

$$\int_a^b f \, \mathrm{d}(g_1 + g_2) = \int_a^b f \, \mathrm{d}g_1 + \int_a^b f \, \mathrm{d}g_2.$$

Another way to obtain Stieltjes integrable functions is considering continuous functions. It turns out that every continuous function is Stieltjes integrable with respect to any increasing function. To prove this, we introduce the notion of 'uniform continuity'. A function f is called uniformly continuous if, roughly speaking, small changes in the input x effect small changes in the output $f(x)$ ('continuity'), and furthermore the size of the changes in $f(x)$ depends only on the size of the changes in x but not on x itself ('uniformity').

Definition 2.4 *Let* I *be an interval and* f *a function on* I. *Then* f *is called uniformly continuous on* I *if for all* $\varepsilon > 0$ *there exists a* $\delta > 0$, *such that for all* x, y *in* I *with* $|x - y| \leqslant \delta$ *we have:* $|f(x) - f(y)| \leqslant \varepsilon$.

Any uniformly continuous function on I is continuous on I. For compact intervals I the converse statement is valid as well: a continuous function on a compact interval I is uniformly continuous on I. If I is not compact, it does not have to be the case. Consider for instance the function $x \mapsto 1/x$ with domain $(0, 1]$. This function is continuous, but not uniformly continuous, since as x approaches 0, the changes in $f(x)$ grow beyond any bound.

Exercise 9 Show that $x \mapsto x^2$ is uniformly continuous on $[0, 1]$ but not on $\mathbb{R}_+$. $\circ$

Theorem 2.5 *Let* $g : [a, b] \longrightarrow \mathbb{R}$ *be an increasing function. Then every (uniformly) continuous function* $f : [a, b] \longrightarrow \mathbb{R}$ *is Stieltjes integrable over* $[a, b]$ *with respect to* g *(and thereby Riemann integrable over* $[a, b]$ *as well).*

Proof Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous. Let $\varepsilon > 0$. We make an underestimation p and an overestimation q with $q - p = 2\varepsilon[g(b) - g(a)]$ (why is this sufficient to prove the theorem?). Because

f is uniformly continuous on $[a,b]$, there exists a $\delta > 0$ such that for all x, y in $[a,b]$ with $|x-y| \leqslant \delta$ we have: $|f(x) - f(y)| \leqslant \varepsilon$. Take a division $(s_0, \ldots, s_n)$ with $s_k \leqslant s_{k-1} + \delta$ ($k \in \{1, \ldots, n\}$). Then $p = \sum_{k=1}^n \big(g(s_k) - g(s_{k-1})\big)\big(f(s_k) - \varepsilon\big)$ and $q = \sum_{k=1}^n \big(g(s_k) - g(s_{k-1})\big)\big(f(s_k) + \varepsilon\big)$ will do. $\square$

The following example expresses that even increasing functions are not necessarily Stieltjes integrable with respect to every function:

Example 2 Take a closed interval $[a,b] \subset \mathbb{R}$ and let g be the indicator function $\mathbb{1}_{(a,b]}$ on this domain. We will show that this increasing function fails to be Stieltjes integrable with respect to itself. Take an underestimation p and an overestimation q of g with respect to g:

$$p = \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big)\ell_k$$
$$q = \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big)h_k,$$

(with s, ℓ_k, h_k as usual). Then $p = \ell_1 \leqslant 0$, and $q = h_1 \geqslant 1$! $\triangle$

Exercise 10 Let the functions f_1, f_2, g_1, and g_2 on domain $[0,2]$ be given by

$$f_1(x) = x^2, \qquad f_2 = \mathbb{1}_{\{1\}}, \qquad g_1(x) = 3x, \qquad g_2 = \mathbb{1}_{[1,2]}.$$

Determine whether the following Stieltjes integrals are well defined and compute the ones that are.

a) $\displaystyle\int_0^2 f_1 \, \mathrm{d}g_1$, **b)** $\displaystyle\int_0^2 f_1 \, \mathrm{d}g_2$, **c)** $\displaystyle\int_0^2 f_2 \, \mathrm{d}g_1$, **d)** $\displaystyle\int_0^2 f_2 \, \mathrm{d}g_2$.

Hint for part **a)**. $\displaystyle\sum_{k=1}^{n} k^2 = \tfrac{1}{6}(n)(n+1)(2n+1)$. $\circ$

There is also a positive result concerning increasing functions and Stieltjes integration. If integrand and integrator both are increasing, Stieltjes integration admits integration by parts.

Theorem 2.6 *Let f and g be increasing functions on interval $[a,b]$ such that $\int_a^b g \, \mathrm{d}f$ exists. Then $\int_a^b f \, \mathrm{d}g$ exists as well and*

$$\int_a^b f \, \mathrm{d}g = f(b)g(b) - f(a)g(a) - \int_a^b g \, \mathrm{d}f.$$

Proof Let s be any division of interval $[a,b]$ consisting of, say, n subintervals. The best underestimation p_{fg} and overestimation q_{fg} of integrand f with respect to integrator g that can be made by means of this division are obtained by defining the lower bounds $\ell_1, \ldots, \ell_n$ to be $\ell_k = \min\big\{f(x) \bigm| x \in [s_{k-1}, s_k]\big\} = f(s_{k-1})$ and the upper bound to be $h_k = \max\big\{f(x) \bigm| x \in [s_{k-1}, s_k]\big\} = f(s_k)$.

Then $$p_{fg} = \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big) f(s_{k-1})$$
and $$q_{fg} = \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big) f(s_k).$$
Similarly, if we switch the roles of f and g and define p_{gf} and q_{gf} by means of the same division s to be the least underestimation and overestimation of integrand g with respect to integrator f, we find
$$p_{gf} = \sum_{k=1}^{n} \big(f(s_k) - f(s_{k-1})\big) g(s_{k-1})$$
and $$q_{gf} = \sum_{k=1}^{n} \big(f(s_k) - f(s_{k-1})\big) g(s_k).$$
Adding p_{fg} and q_{gf} yields
$$p_{fg} + q_{gf} \;=\; f(s_n)g(s_n) - f(s_0)g(s_0) \;=\; f(b)g(b) - f(a)g(a).$$
When we let the number of subintervals tend to infinity, we find that p_{fg} tends to $\underline{\int}_a^b f \, \mathrm{d}g$ and, because by assumption $\int_a^b g \, \mathrm{d}f$ exists, q_{gf} tends to $\int_a^b g \, \mathrm{d}f$. Hence,
$$\underline{\int_a^b} f \, \mathrm{d}g \;+\; \int_a^b g \, \mathrm{d}f \;=\; f(b)g(b) - f(a)g(a).$$
Similarly, adding q_{fg} and p_{gf} and then taking the limits yields
$$\overline{\int_a^b} f \, \mathrm{d}g \;+\; \int_a^b g \, \mathrm{d}f \;=\; f(b)g(b) - f(a)g(a).$$

Hence, $$\underline{\int_a^b} f \, \mathrm{d}g = \overline{\int_a^b} f \, \mathrm{d}g = f(b)g(b) - f(a)g(a) - \int_a^b g \, \mathrm{d}f.$$ □

Exercise 11 Let $[a, b] \subset \mathbb{R}$ and let X be a continuous random variable with distribution function F with the property that $\mathbb{P}[a \leqslant X \leqslant b] = 1$. Find $\mathbb{E}[F(X)]$, which equals $\int_a^b F \, \mathrm{d}F$. Verify your answer by means of a figure (draw the graph of F with respect to F and determine the area below the graph). ○

2.3 From Stieltjes back to Riemann

If f and g are sufficiently smooth, the Stieltjes integral $\int_a^b f \, \mathrm{d}g$ can be replaced by a Riemann integral.

Theorem 2.7 *Let $g : [a, b] \longrightarrow \mathbb{R}$ be a function with a continuous derivative. For each Riemann integrable function $f : [a, b] \longrightarrow \mathbb{R}$ we have*
$$\int_a^b f \, \mathrm{d}g = \int_a^b f(x)g'(x) \, \mathrm{d}x.$$

Proof For ease, we assume that $f \geqslant 1$ and $g' \geqslant 1$. Let $\varepsilon > 0$. Because f and g' are both Riemann integrable, they are by definition bounded by, say, $M > 0$. Furthermore, their product-function is Riemann integrable as well (see Theorem 1.8(v)). Let $\varepsilon \in (0, 1)$ and let p be an underestimation of $\int_a^b f(x)g'(x)\,\mathrm{d}x$ with $\int_a^b f(x)g'(x)\,\mathrm{d}x - p < \varepsilon$. Let s be a division of $[a, b]$ into, say, n subsegments that can be used to define p. As usual, we call the lower bounds used ℓ_k. Moreover, s must be sufficiently fine to ensure that $|g'(x) - g'(y)| < \varepsilon$ for all x, y inside the same subsegment of s. The last requirement can be fulfilled because g' is uniformly continuous. Let $k \in \{1, \ldots, n\}$. Define $\ell_k^f = \inf_{x\in[s_{k-1},s_k]} f(x)$ and $\ell_k^{g'} = \inf_{x\in[s_{k-1},s_k]} g'(x)$. Then there exists an $\bar{x} \in [s_{k-1}, s_k]$ with $\ell_k^f \geqslant f(\bar{x}) - \varepsilon$. We have that $\ell_k^{g'} \geqslant g'(x) - \varepsilon$ for all $x \in [s_{k-1}, s_k]$, so in particular $\ell_k^{g'} \geqslant g'(\bar{x}) - \varepsilon$. Therefore,

$$\ell_k^f \cdot \ell_k^{g'} \geqslant (f(\bar{x}) - \varepsilon) \cdot (g'(\bar{x}) - \varepsilon) \geqslant \ell_k - 2\varepsilon M.$$

The first inequality needs the assumptions that $f \geqslant 1 > \varepsilon$ and $g' \geqslant 1 > \varepsilon$.

Furthermore, the Mean Value Theorem guarantees the existence of a number $\alpha_k \in [s_{k-1}, s_k]$ with $g'(\alpha_k) = \frac{g(s_k)-g(s_{k-1})}{s_k-s_{k-1}}$, i.e., $g(s_k) - g(s_{k-1}) = g'(\alpha_k)(s_k - s_{k-1})$.

We have

$$\begin{aligned}
p &= \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k \\
&\leqslant \sum_{k=1}^{n} (s_k - s_{k-1})(\ell_k^f \cdot \ell_k^{g'} + 2\varepsilon M) \\
&= 2\varepsilon M(b-a) + \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k^f \cdot \ell_k^{g'} \\
&\leqslant 2\varepsilon M(b-a) + \sum_{k=1}^{n} (s_k - s_{k-1})\ell_k^f \cdot g'(\alpha_k) \\
&= 2\varepsilon M(b-a) + \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big)\ell_k^f \\
&\leqslant 2\varepsilon M(b-a) + \underline{\int_a^b} f\,\mathrm{d}g.
\end{aligned}$$

If we let ε tend to 0, we find that $\int_a^b f(x)g'(x)\,\mathrm{d}x \leqslant \underline{\int_a^b} f\,\mathrm{d}g$. We can repeat the argument, but now with overestimations, to establish that $\int_a^b f(x)g'(x)\,\mathrm{d}x \geqslant \overline{\int_a^b} f\,\mathrm{d}g$. This proves the theorem for functions that are larger than 1. To prove the statement in general, we should first consider functions $\hat{f} = f + K$ and $\hat{g} = g + Kx$ with $K > 0$ such that $\hat{f} \geqslant 1$ and $\hat{g}' \geqslant 1$. This gives that $\int_a^b \hat{f}\hat{g}'\,\mathrm{d}x = \int_a^b \hat{f}\,\mathrm{d}\hat{g}$. We leave it to the reader to verify that the last statement ensures that $\int_a^b f(x)g'(x)\,\mathrm{d}x = \int_a^b f\,\mathrm{d}g$ as well. □

The standard way of integrating by parts is a byproduct of Theorems 2.6 and 2.7. If f and g are differentiable, the equation of Theorem 2.6 rewrites to the well-known formula

$$\int_a^b f(x)g'(x)\,\mathrm{d}x = \Big[f(x)g(x)\Big]_a^b - \int_a^b f'(x)g(x)\,\mathrm{d}x.$$

Example 3 The distribution function F of random variable $X \sim \mathrm{Un}\big([1,3]\big)$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leqslant 1, \\ \frac{1}{2}(x-1) & \text{if } 1 \leqslant x \leqslant 3, \\ 1 & \text{if } 3 \leqslant x. \end{cases}$$

Then
$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x\,\mathrm{d}F(x) = \int_{-\infty}^{1} x\,\mathrm{d}F(x) + \int_1^3 x\,\mathrm{d}F(x) + \int_3^{\infty} x\,\mathrm{d}F(x) \\ &= \int_{-\infty}^{1} xF'(x)\,\mathrm{d}x + \int_1^3 xF'(x)\,\mathrm{d}x + \int_3^{\infty} xF'(x)\,\mathrm{d}x = 0 + \int_1^3 \tfrac{1}{2}x\,\mathrm{d}x + 0 \\ &= \Big[\tfrac{1}{4}x^2\Big]_1^3 = \tfrac{9}{4} - \tfrac{1}{4} = 2. \end{aligned}$$
△

Exercise 12 Verify Exercises 10**a)** and **c)** by means of Theorem 2.7. ○

Exercise 13 Find $\mathbb{E}[X^2]$, in which $X \sim \mathrm{Un}\big([1,4]\big)$ (cf. Section 2.1). ○

2.4 Stieltjes integration and convergence

Even more Stieltjes integrable functions can be obtained by taking limits of sequences of Stieltjes integrable functions. However, to do so the meaning of convergence of a sequence of functions must be clear. There are two types of convergence we would like to introduce.

Definition 2.8 *Suppose $f, f_1, f_2, \ldots$ are functions on interval $[a,b]$. The sequence $f_1, f_2, \ldots$* converges pointwise *to f on $[a,b]$ if for each x in $[a,b]$: $\lim_{k\to\infty} f_k(x) = f(x)$. $f_1, f_2, \ldots$* converges uniformly *to f if for every $\varepsilon > 0$ there exists a natural number N such that for all $k \geqslant N$ and all x in $[a,b]$ we have: $|f_k(x) - f(x)| \leqslant \varepsilon$.*

Pointwise convergence means that for each x in $[a,b]$, the sequence of numbers $f_1(x), f_2(x), \ldots$ converges. Uniform convergence means that the convergence takes place globally at a same pace. The following exercise shows that uniform convergence is a (much) stronger property than pointwise convergence.

Exercise 14 Consider the sequence of functions $f_1, f_2, \ldots$ on $[0,1]$ defined by $f_k(x) = x^k$ for all x in $[0,1]$ and k in $\mathbb{N}$. Show that this sequence converges pointwise (to which function?), but does not converge uniformly. ○

Without a proof we postulate the following theorem.

Theorem 2.9 *Let $g : [a,b] \longrightarrow \mathbb{R}$ be an increasing function. If $f_1, f_2, \ldots$ are Stieltjes integrable functions with respect to g and $f = \lim_{k\to\infty} f_k$ uniformly, then f is Stieltjes integrable with respect to g and*

$$\int_a^b f \,\mathrm{d}g = \lim_{k\to\infty} \int_a^b f_k \,\mathrm{d}g.$$

The assumption of *uniform* convergence is necessary, but it is not easy to present an example to show this. First, we have to introduce the notion of *countability*. Two sets A and B are called of equal *cardinality* if there exists a bijection from A to B. A set is called *countable* if it is finite or of equal cardinality with the set of positive integers. A set is countably infinite if it is countable and infinite, just like the set of positive integers, *i.e.*, with $\mathbb{N}$. 'A and B are of equal cardinality' is the mathematical translation of the informal phrase 'A and B are just as large'. Obviously, two finite sets can only be of equal size if they have an equal number of elements. It is more complicated in the case of infinite sets. The set of nonnegative integers ($\mathbb{N} \cup \{0\}$) is countable, as shown by the bijection $k \mapsto k+1$. The set of even numbers is countable; map k to $k/2$. Thus an infinite set can be just as large as a proper subset of itself. The real line $\mathbb{R}$ has the same cardinality as $(-\frac{\pi}{2}, \frac{\pi}{2})$ according to the bijection $x \mapsto \arctan(x)$. We will see that $\mathbb{R}$ is *not* countable.

Exercise 15 Show that $(0,1]$ and $[-3,2)$ are of equal cardinality. Do the same for $(0,\infty)$ and $\mathbb{R}$. $\circ$

Any set that can be listed in order, *i.e.*, enumerated, is countable. *E.g.*, any subset of $\mathbb{N}$ is countable. Take the smallest number in the set and place it first on the list. Take the next smallest and place it second on the list. Continue this process until the entire set has been 'counted' (this might take an infinite amount of time...), *i.e.*, mapped into $\mathbb{N}$ (and if the subset is not finite, also onto). Thus the set of prime numbers is countable, the set of squares is, any set of natural numbers. Another example of a countable set is $\mathbb{Q} \cap [0,1]$. How can we enumerate it? A possibility is to start with all elements with denominator equal to 1, then all elements (not yet listed) with denominator equal to 2, then all new elements with denominator 3, and so on. This results in

$$\frac{0}{1}, \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \ldots$$

(so $f(1) = 0, f(7) = \frac{3}{4}, \ldots$) and clearly each element of $\mathbb{Q} \cap [0,1]$ gets a position in the sequence.

This might be counterintuitive, since rational numbers are densely packed in the segment $[0,1]$. There are infinitely many rationals packed into the tiniest of intervals, yet there are just as many rationals as integers.

Exercise 16 Show that the sets $\mathbb{Z}$ and $\mathbb{Q}$ are countable. Show that every subset of a countable set is countable. $\circ$

Exercise 17 Let $A_1, A_2, A_3, \ldots$ be a sequence of infinitely countable sets. Each A_i can be enumerated:

$$\begin{array}{rcl} A_1 &=& \{a_{11}, a_{12}, a_{13}, \ldots\}, \\ A_2 &=& \{a_{21}, a_{22}, a_{23}, \ldots\}, \\ A_3 &=& \{a_{31}, a_{32}, a_{33}, \ldots\}, \\ \vdots & & \quad\quad\vdots \end{array}$$

Prove that $\bigcup_{i=1}^{\infty} A_i$ is countable as well by enumerating its elements. $\circ$

An example of a non-countable set is $\mathbb{R}$. Why? Suppose, on the contrary, that $\mathbb{R}$ is countable, say $\mathbb{R} = \{f(1), f(2) \ldots\}$ for some bijection $f : \mathbb{N} \longrightarrow \mathbb{R}$. Choose an interval I_1 of length 1 containing $f(1)$, a second interval I_2 of length $\frac{1}{2}$ containing $f(2)$, a third interval I_3 of length $\frac{1}{4}$ containing $f(3)$, and so on. Then $\mathbb{R}$ is a subset of $I_1 \cup I_2 \cup \ldots$; an infinite union of intervals of which the total length does not exceed 2. This is impossible, which makes $\mathbb{R}$ uncountable.

Exercise 18 Show that $[0,1]$ is uncountable. $\circ$

Finally we have reached the example we announced.

Example 4 Because $[0,1] \cap \mathbb{Q}$ is countable, we can write $[0,1] \cap \mathbb{Q} = \{q_1, q_2, q_3, \ldots\}$. Define for all $k \in \mathbb{N}$, f_k by $f_k = \mathbb{1}_{\{q_1,\ldots,q_k\}}$. Then f_k is Riemann integrable over $[0,1]$ (cf. Exercise 2, combined with Theorem 1.8(i)) and $\int_0^1 f_k(x)\,\mathrm{d}x = 0$. The pointwise limit of $f_1, f_2, \ldots$ is $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$. This function is *not* Riemann integrable (Exercise 4). $\triangle$

Exercise 19 Verify that in the example above f_k does not converge to f uniformly. $\circ$

(for the interested reader) Usually, Stieltjes integration is applied by means of an increasing integrator g, but it is not necessary to restrict ourselves to this type of integrators. Stieltjes integrals can be defined with respect to *functions of bounded variation.* Such functions are the difference of two increasing functions. If g is a function of bounded variation, then a function f is called Stieltjes integrable with respect to g if f there can be found two functions g_1 and g_2 on $[a, b]$ such that

g_1 and g_2 are both increasing,

$g = g_1 - g_2$,

$\int_a^b f \, \mathrm{d}g_1$ and $\int_a^b f \, \mathrm{d}g_2$ are well defined.

Its Stieltjes integral is defined to be $\int_a^b f \, \mathrm{d}g = \int_a^b f \, \mathrm{d}g_1 - \int_a^b f \, \mathrm{d}g_2$. It can be shown that the choices of g_1 and g_2 do not matter, as long as they obey the three conditions above. The results of this section can all be generalized to this setting, but a generalized Stieltjes integral cannot be visualized by an area.

3 Lebesgue integration

3.1 Introduction

Although Stieltjes integration can handle more situations than Riemann integration, like the expectation of a discrete random variable, the theory of Stieltjes integration still leaves some questions open. Let us present some of them.

(i) The function $\mathbb{1}_{\mathbb{Q}}$ is not Stieltjes integrable with respect to non-constant integrators (cf. Exercise 4). The function has value 1 on only a small set, because $\mathbb{Q}$ is countable. It is the pointwise limit of functions of which the Stieltjes integrals all equal zero. It would be convenient if zero were the integral of $\mathbb{1}_{\mathbb{Q}}$ over any bounded segment $[a, b]$ as well.

(ii) Stieltjes integrals are defined for functions on a compact interval. How should we define integrals of functions with (semi-)infinite domains?

(iii) The functions that we have integrated so far all have domains in $\mathbb{R}$. Is it possible to build a type of integration theory for functions $f : S \longrightarrow \mathbb{R}$ in which S is another type of set? In particular, can S be the sample space of some experiment?

(iv) Summing sequences and integrating functions have several properties in common. Can both mathematical operations be brought under one umbrella? In other words, is it possible to regard summation as a special type of integration?

(v) Theorem 2.9 shows that under certain conditions a limit and an integral can switch positions without affecting the outcome, *i.e.*, $\lim_{k\to\infty}\left(\int f_k\right) = \int\left(\lim_{k\to\infty} f_k\right)$. Is it possible to find milder (or other) conditions such that this is still the case?

It will turn out that this Lebesgue integration theory can answer the questions mentioned above affirmatively. We have seen that in order to find an appropriate definition of the integral of $\mathbb{1}_{\mathbb{Q}}$, the division of the domain into subintervals does not work. If we make the domain an abstract set S, like in the third question, we cannot speak of subintervals of the domain at all. It turns out to be a good idea to subdivide the *range* into finitely many levels, or strips, or band-widths. If B is one of the band-widths, we can define the set $A \subseteq S$ to be the set of all elements x of S with $f(x) \in B$. A is called the *inverse image* of (het *volledig origineel* van) B and is denoted by $f^{-1}(B)$:

$$f^{-1}(B) = \{x \in S : f(x) \in B\}.$$

If $S \subseteq \mathbb{R}$ and f is monotone, A is again a subinterval, but in general it can be any subset of S. If we take $f = \mathbb{1}_{\mathbb{Q}}$ on domain $[0,1]$, and subdivide its range into two levels, $\{0\}$ and $\{1\}$, we find that $f^{-1}(\{0\}) = [0,1]\backslash\mathbb{Q}$ and $f^{-1}(\{1\}) = [0,1] \cap \mathbb{Q}$. If we can somehow measure the importance (or size, weight) of $f^{-1}(\{0\})$ and $f^{-1}(\{1\})$, then the integral could be something like

‘$\int f \approx 0 \cdot$ [the importance of $f^{-1}(\{0\})$] $+$ $1 \cdot$ [the importance of $f^{-1}(\{1\})$]’,

i.e., ‘$\int f \approx 0 \cdot$ [the size of $[0,1]\backslash\mathbb{Q}$] $+$ $1 \cdot$ [the size of $[0,1] \cap \mathbb{Q}$]’.

Exercise 20 We are going to give an underestimation of $\int_0^\pi \sin(x)\,\mathrm{d}x$ by means of the idea above. Firstly, we divide the range $[-1,1]$ into subintervals:

$B_1 = [-1,0]$, $B_2 = [0,\frac{1}{2}]$, $B_3 = [\frac{1}{2},\frac{1}{2}\sqrt{3}]$, and $B_4 = [\frac{1}{2}\sqrt{3},1]$.

Define $A_i = \sin^{-1}(B_i)$ for each i in $\{1,\ldots,4\}$.

a) Give for each i the set A_i explicitly.

b) How should you define the ‘importance’ (size) of A_i for each i?

Let $a_i = \min\{y : y \in B_i\}$. For each x in A_i, we have that $\sin(x) \in B_i$, so $\sin(x) \geqslant a_i$. Then $\sum_{i=1}^{4} a_i \cdot$ [the size of A_i] should be a lower bound for $\int_0^\pi \sin(x)\,\mathrm{d}x$ (= 2).

c) Draw a figure to illustrate this inequality (this can only be done if you have answered part **b** reasonably).

d) Compute your lower bound. ○

We have seen that questions lead to ideas, and now the ideas lead to new questions. Can we always determine the inverse image of a subset of the range of a function? And if so, can we always measure its importance?

We will not brainstorm any further, but provide a solid mathematical basis to formalize the ideas and answer the questions. Section 3.3 provides a way to measure subsets of the domain of a function. In fact, a *measure* will turn out to be a special type of function, not working on real numbers, but on subsets of the domain S. It is often not necessary (or even possible) to measure all subsets. Before we can define a measure, we must first define which collection of subsets it associates a value to. The next section defines appropriate collections that can be used for this purpose. These collections are called *σ-fields* or *σ-algebras*.

3.2 σ-Fields

Let S be a set. The collection of all subsets of S is called the *power set* of S, and is denoted by $\mathcal{P}(S)$.

For example, $\quad \mathcal{P}(\{1,2,3\}) = \Big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\Big\}.$

If S has finite cardinality, then $|\mathcal{P}(S)| = 2^{|S|}$, which explains the name *power set*.

Definition 3.1 *Let $\mathcal{A} \subseteq \mathcal{P}(S)$, so $\mathcal{A}$ is a collection of subsets of S. $\mathcal{A}$ is called a σ-field (over S) if*

(i) $\emptyset \in \mathcal{A}$,

(ii) *if $A \in \mathcal{A}$, then $S \backslash A \in \mathcal{A}$,*

(iii) *if $A_1, A_2, \ldots$ is a sequence of elements of $\mathcal{A}$, then $\bigcup\limits_{k=1}^{\infty} A_k$ is also an element of $\mathcal{A}$.*

These three properties qualify a collection of subsets of S for being used as the domain of a measure. σ-Fields are closed under taking complements (property (ii)) and under uniting a countably infinite number of sets (property (iii)). The following exercise shows that they are also closed under uniting finitely many sets or intersecting a countable number of sets.

Exercise 21 Let $\mathcal{A}$ be a σ-field over S and let $A_1, A_2, \ldots$ be a sequence of elements of A. Show that

a) $A_1 \cup A_2 \in \mathcal{A}$, **b)** $A_1 \backslash A_2 \in \mathcal{A}$, **c)** $A_1 \cap A_2 \in \mathcal{A}$, **d)** $\bigcap\limits_{k=1}^{\infty} A_k \in \mathcal{A}$. ○

Exercise 22 How many σ-fields can be formed if the set S has three elements? ○

Exercise 23 A subset A of S is called *cofinite* if $S \backslash A$ is finite and *cocountable* if $S \backslash A$ is countable.

a) Take $S = \mathbb{N}$ and show that $\mathcal{G} = \{A \in \mathcal{P}(\mathbb{N}) \mid A \text{ is finite or cofinite}\}$ is *not* a σ-field.

b) Take $S = \mathbb{R}$ and prove that $\mathcal{Z} = \{A \in \mathcal{P}(\mathbb{R}) \mid A \text{ is countable or cocountable}\}$ is a σ-field. ∘

A σ-field over S is a subcollection of $\mathcal{P}(S)$. $\mathcal{P}(S)$ itself is the largest of all σ-fields in S. The smallest one is $\{\emptyset, S\}$. For two σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$, $\mathcal{A}_1 \cap \mathcal{A}_2$ is (obviously) the collection of all sets that are members of both $\mathcal{A}_1$ and $\mathcal{A}_2$. The union $\mathcal{A}_1 \cup \mathcal{A}_2$ is the collection of all sets that are member of at least one of the collections $\mathcal{A}_1$ and $\mathcal{A}_2$.

Exercise 24 Show: if $\mathcal{A}_1$ and $\mathcal{A}_2$ are σ-fields over S, then $\mathcal{A}_1 \cap \mathcal{A}_2$ is another one. ∘

Exercise 25 Show by means of an example that the union of two σ-fields over $\{1, 2, 3\}$ need not be a σ-field itself. ∘

Exercise 26 Let $S = \{1, 2, 3, 4\}$ and let $\mathcal{G}$ be the collection $\{\{1\}, \{1, 2\}\}$. Obviously, $\mathcal{G}$ is not a σ-field. Try to add as few as possible subsets of S, such that the result is a σ-field. In other words, find a σ-field $\mathcal{A}$ with $\mathcal{G} \subset \mathcal{A}$ and $|\mathcal{A}|$ as small as possible (it can be done by adding 6 subsets of S to $\mathcal{G}$). ∘

As discussed in the introduction section of this chapter, we are going to define measures. For the time being, it suffices to know about measures that they always have σ-fields as their domains. Suppose we would like to give some measure on some collection $\mathcal{G}$ of subsets of S. Then we first must extend the collection $\mathcal{G}$ to a σ-field $\mathcal{A}$, just like in Exercise 26. Of course we can choose the complete power set of S to be $\mathcal{A}$, but the larger the collection $\mathcal{A}$ is chosen, the more effort it takes to define the measure (if possible at all). Hence, we would like to find a σ-field $\mathcal{A}$ containing $\mathcal{G}$ as small as possible.

Example 5 Let $S = \mathbb{R}$ and let $\mathcal{G}$ be the collection of all finite subsets of $\mathbb{R}$. Suppose there exists a smallest σ-field that contains $\mathcal{G}$. Can we find it?
Any countable subset $\{x_1, x_2, \ldots\}$ of $\mathbb{R}$ is the countable union of finite subsets, because $\{x_1, x_2, \ldots\} = \bigcup_{i=1}^{\infty} \{x_i\}$. Hence, by property (iii) of Definition 3.1, every countable subset of $\mathbb{R}$ must be added to $\mathcal{G}$. By property (ii), every cocountable subset of $\mathbb{R}$ must be added as well. Hence, any σ-field containing $\mathcal{G}$ contains the collection $\mathcal{Z}$ of Exercise 23b). On the other hand, this collection $\mathcal{Z}$ is a σ-field. Apparently $\mathcal{Z}$ is the smallest σ-field containing $\mathcal{G}$. △

The problem is much more difficult if we take an arbitrary collection $\mathcal{A}$. Actually, there are many collection of which the smallest σ-field cannot explicitly be given. The best we can do is to prove that:

(i) There always exists a σ-field containing $\mathcal{G}$ (this is obvious, just take $\mathcal{P}(S)$).

(ii) Of all σ-fields containing $\mathcal{G}$, one is the smallest.

The trick to find the smallest one is to consider *all* σ-fields containing $\mathcal{G}$ and form their intersection.

Theorem 3.2 *Let S be a set and let $\mathcal{G} \subset \mathcal{P}(S)$. Define*

$$\sigma(\mathcal{G}) = \Big\{ B \in \mathcal{P}(S) \mid B \in \mathcal{A} \text{ for every } \sigma\text{-field } \mathcal{A} \text{ containing } \mathcal{G} \Big\}.$$

Then $\sigma(\mathcal{G})$ is a σ-field, and it is the smallest one that contains $\mathcal{G}$.

$\sigma(\mathcal{G})$ is called 'the σ-field *generated* by $\mathcal{G}$' and $\mathcal{G}$ is called a '*generating collection* of $\sigma(\mathcal{G})$'.

Proof Surely $\sigma(\mathcal{G})$ is a collection that contains $\mathcal{G}$. Every σ-field that contains $\mathcal{G}$, also contains $\sigma(\mathcal{G})$. Hence, *if* $\sigma(\mathcal{G})$ is a σ-field, we are done. Let us show property (ii) of Definition 3.1.
Let $A \in \sigma(\mathcal{G})$. For every σ-field $\mathcal{A}$ that contains $\mathcal{G}$, we have $A \in \sigma(\mathcal{G}) \subset \mathcal{A}$, so $S \backslash A \in \mathcal{A}$ as well. The definition of $\sigma(\mathcal{G})$ gives that $S \backslash A \in \sigma(\mathcal{G})$.

Exercise 27 Complete the proof of Theorem 3.2 by showing that $\sigma(\mathcal{G})$ also obeys properties (i) and (iii) of a σ-field (cf. Exercise 24). □○

Proposition 3.3 *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two collections of subsets of S, so $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{P}(S)$.*

If $\mathcal{G}_1 \subseteq \sigma(\mathcal{G}_2)$, then
$$\sigma(\mathcal{G}_1) \subseteq \sigma(\mathcal{G}_2).$$

The proof is straightforward and omitted. The proposition can be applied to prove that two collections $\mathcal{G}_1$ and $\mathcal{G}_2$ generate the same σ-field. An example:

Example 6 Let $S = \mathbb{R}$ and

$$\mathcal{G}_1 = \Big\{ (a,b] \ \Big| \ a, b \in \mathbb{R}, a < b \Big\},$$
$$\mathcal{G}_2 = \Big\{ (a,b) \ \Big| \ a, b \in \mathbb{R}, a < b \Big\}.$$

For every element $(a,b]$ of $\mathcal{G}_1$ we have that $(a,b] = \bigcap_{n \in \mathbb{N}} (a, b+\frac{1}{n})$. For every natural number n, $(a, b+\frac{1}{n}) \in \mathcal{G}_2 \subseteq \sigma(\mathcal{G}_2)$. Hence, $(a,b] \in \sigma(\mathcal{G}_2)$. This makes $\mathcal{G}_1$ a subset of $\sigma(\mathcal{G}_2)$, so $\sigma(\mathcal{G}_1) \subseteq \sigma(\mathcal{G}_2)$. Conversely, every element (a,b) of $\mathcal{G}_2$ is the countably infinite union of elements of $\mathcal{G}_1$ (verify this) and thereby an element of $\sigma(\mathcal{G}_1)$. This gives $\mathcal{G}_2 \subseteq \sigma(\mathcal{G}_1)$, and thus $\sigma(\mathcal{G}_2) \subseteq \sigma(\mathcal{G}_1)$. Conclusion: $\sigma(\mathcal{G}_1)$ equals $\sigma(\mathcal{G}_2)$.

△

The σ-field $\sigma(\mathcal{G}_1)$ plays such a key role in set theory and Lebesgue theory that it has been given a name: the *Borel-σ-field* (of $\mathbb{R}$).

Definition 3.4 *Let $S \subseteq \mathbb{R}$ be an interval and let $\mathcal{G}_1 = \Big\{(a,b] \Big| \, a, b \in S, a < b\Big\}$. Then $\sigma(\mathcal{G}_1)$ is called the Borel-σ-field of S. It is denoted by $\mathcal{B}_S$, or simply $\mathcal{B}$ if confusion about which set S is referred to is not likely to occur.*

Exercise 28 Let $S = \mathbb{R}$. Show that the collection $\mathcal{G}_3 = \big\{(-\infty, b] \,\big|\, b \in \mathbb{R}\big\}$ also generates $\mathcal{B}$. $\circ$

3.3 Measures

The next notion that has to be defined to set up Lebesgue theory is the notion of *measure*; a function that assigns to every element of a σ-field $\mathcal{A}$ a size.

Definition 3.5 *Let $\mathcal{A}$ be a σ-field over S. A function $\mu : \mathcal{A} \longrightarrow [0, \infty]$ is called a measure if*

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

(ii) μ *is σ-additive , i.e.,* $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$ *if all A_k's are elements of $\mathcal{A}$ and are pairwise disjoint.*

A pair $(S, \mathcal{A})$ in which S is a set and $\mathcal{A}$ is a σ-field on S is called *measurable space*. A triple $(S, \mathcal{A}, \mu)$ in which $(S, \mathcal{A})$ is a measurable space and μ is a measure over $\mathcal{A}$, is called a *measured space*. Elements of $\mathcal{A}$ are called *$\mathcal{A}$-measurable sets* (or simply measurable sets if no confusion can occur).
According to this definition, the measure $\mu(A)$ of a set A in $\mathcal{A}$ is an element of $[0, \infty]$. Apparently it is possible that the size of a set is infinite. The interpretation of $\mu(A)$, the measure of A, can be its cardinality, length, area, volume, weight or any other notion of size. $\mu(A)$ can represent more abstract properties of A as well. For instance, if S is the sample space of an experiment, then $\mu(A)$ can denote the probability that the outcome of the experiment is somewhere in A. This interpretation requires that

$$\mu(S) = \mathbb{P}[\text{the outcome is an element of } S] = 1.$$

It motivates the convention to call any measure μ with $\mu(S) = 1$ a *probability measure*.

Example 7 [Counting measure]

In this context it is customary to use the character τ instead of μ. Let S be any set, let $\mathcal{A}$ be its power set $\mathcal{P}(S)$ and define for all subsets A of S,

$$\tau(A) = \text{ the number of elements of } A.$$

Verify that the properties of a measure are valid, even if $\tau(S) = \infty$. $\triangle$

Example 8 [Probability measure, product measure]

Take S to be the set of all possible outcomes of the following experiment: a coin and a dice are thrown. An element of S is a pair (c, d), in which c denotes the result of the throw of the coin, which is either 'head' or 'tail', and d denotes the result of the dice roll. Hence $S = \{(\text{head}, 1), \ldots, (\text{head}, 6), (\text{tail}, 1), \ldots, (\text{tail}, 6)\}$. Let $\mathcal{A} = \mathcal{P}(S)$ and define the probability measure μ by $\mu(A) = \mathbb{P}[A \text{ contains the realized outcome}] = \frac{|A|}{12}$ for each subset of A of S.

Another way to look at this situation is to consider it as two independent experiments. This can be modeled as follows: $S_c = \{\text{head}, \text{tail}\}$ and $\mu_c(A) = \frac{|A|}{2}$ for all $A \subseteq S_c$, and $S_d = \{1, \ldots, 6\}$ and $\mu_d(A) = \frac{|A|}{6}$ for all $A \subseteq S_d$. The sample space S is called the *product space* of S_c and S_d, *i.e.*, $S = S_c \times S_d$. Similarly, we call μ the *product measure* of μ_c and μ_d, *i.e.*, $\mu = \mu_c \times \mu_d$. μ is the unique measure on $\mathcal{P}(S)$ satisfying $\mu(A_c \times A_d) = \mu_c(A_c) \cdot \mu_d(A_d)$ for all subsets A_c, A_d of S_c and S_d respectively, *e.g.*, $\mu(\{\text{head}\} \times \{1, 3\}) = \mu_c(\{\text{head}\}) \cdot \mu_d(\{1, 3\}) = \frac{1}{2} \cdot \frac{2}{6} = \frac{1}{6}$. $\triangle$

Exercise 29 Let $S = \mathbb{R}$ and $\mathcal{A} = \{A \in \mathcal{P}(\mathbb{R}) : A \text{ is countable or cocountable}\}$.

a. Provide a subset of $\mathbb{R}$ that is countable, one that is cocountable, and one that is neither countable nor cocountable. Are there subsets of $\mathbb{R}$ that are both countable and cocountable?

b. Prove that $\mu_c : \mathcal{A} \longrightarrow [0, \infty]$ with $c \geqslant 0$ is a measure, if we define

$$\mu_c(A) = \begin{cases} 0 & \text{if } A \text{ is countable,} \\ c & \text{if } A \text{ is cocountable.} \end{cases}$$

$\circ$

The definition of a measure shows how to find the measure of the union of a pairwise disjoint sequence of sets. The following theorem provides the measure of the limit set of sequences of nested sets.

Theorem 3.6 *Let $(S, \mathcal{A}, \mu)$ be a measured space.*

(i) *Let $A, B \in \mathcal{A}$ with $A \subseteq B$. Then $\mu(B) = \mu(A) + \mu(B \backslash A)$ and thus $\mu(A) \leqslant \mu(B)$.*

(ii) *If $A_1 \subseteq A_2 \subseteq \ldots$ is an increasing sequence of sets in $\mathcal{A}$, then we have*

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \lim_{k \to \infty} \mu(A_k).$$

(iii) *If $B_1 \supseteq B_2 \supseteq \ldots$ is a decreasing sequence of sets in $\mathcal{A}$, then we have*

$$\mu\Big(\bigcap_{n=1}^{\infty} B_n\Big) = \lim_{n\to\infty} \mu(B_n)$$

<u>*provided that*</u> $\mu(B_t) < \infty$ *for some* t.

Proof (i) $\mu(B) = \mu((B\backslash A) \cup A) = \mu(B\backslash A) + \mu(A) \geqslant \mu(A)$ because of σ-additivity and the fact that $\mu \geqslant 0$.

(ii) Define $C_1 := A_1, C_2 := A_2\backslash A_1, \ldots, C_k := A_k\backslash A_{k-1}, \ldots$.

We have that $C_k \in \mathcal{A}$ for each k and $C_i \cap C_j = \emptyset$ if $i \neq j$. Moreover, $\bigcup_{k=1}^{n} C_k = A_n$ for each n and $\bigcup_{k=1}^{\infty} C_k = \bigcup_{k=1}^{\infty} A_k$. Hence,

$$\begin{aligned} \mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) &= \mu\Big(\bigcup_{k=1}^{\infty} C_k\Big) = \sum_{k=1}^{\infty} \mu(C_k) \\ = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(C_k) &= \lim_{n\to\infty} \mu\Big(\bigcup_{k=1}^{n} C_k\Big) = \lim_{n\to\infty} \mu(A_n). \end{aligned}$$

□

Exercise 30 Prove Theorem 3.6(iii). *Hint.* Choose a t in $\mathbb{N}$ with $\mu(B_t) < \infty$ and apply (ii) on the sequence of sets $A_k = B_t\backslash B_k$ $(k \in \mathbb{N})$. Pinpoint at which place(s) the condition that $\mu(B_t) < \infty$ is used. ○

Example 9 The condition at Theorem 3.6(iii) is not redundant. Let $S = \mathbb{R}$,

$\mathcal{A} = \mathcal{P}(S)$, and
$$\mu(A) = \begin{cases} 0 & \text{if } A \text{ is countable,} \\ \infty & \text{if } A \text{ is uncountable.} \end{cases}$$

This is a measure (verify this). Consider $B_1 = \mathbb{R}, B_2 = \mathbb{R}\backslash\{(0,1) \cup (-1,0)\}, B_3 = \mathbb{R}\backslash\{(0,1) \cup (-1,0) \cup (1,2) \cup (-2,-1)\}, \ldots$. Then $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$. However, $\bigcap_{k=1}^{\infty} B_k = \mathbb{Z}$ is countable, so $\mu\left(\bigcap_{k=1}^{\infty} B_k\right) = 0$ and $\lim_{n\to\infty} \mu(B_n) = \infty$. △

To conclude this section we concentrate on the measurable space $(\mathbb{R}, \mathcal{B})$. This σ-field is generated by the collection $\mathcal{G}_1$ of left-open, right-closed intervals, *i.e.*, intervals of the form $(a, b]$. If we would like to construct a measure λ on $\mathcal{B}$, we could start by defining that λ assigns to elements of $\mathcal{G}_1$ their natural sizes:
$$\lambda\big((a, b]\big) := b - a \text{ for each } a, b \text{ in } \mathbb{R} \text{ with } a < b.$$

Notice that σ-additivity is satisfied: let $A_1, A_2, \ldots$ be a sequence of pairwise disjoint intervals in $\mathcal{G}_1$, such that $\bigcup_{k=1}^{\infty} A_k \in \mathcal{G}_1$.[2] We have: $\lambda\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \lambda(A_k)$. It can be shown (albeit with a lot of effort) that

[2]Take *e.g.* $A_k = (\frac{1}{k+1}, \frac{1}{k}]$. Here the at first sight peculiar choice of half-open intervals is convenient.

there exists a unique way to extend λ to $\mathcal{B}$ such that it becomes a measure. This extension is denoted as well by λ. It is called the *Borel-Lebesgue-measure* .
It is not possible to provide a procedure to compute the λ-measure of an arbitrary set A in $\mathcal{B}$. Theorem 3.6 does help however in a lot of cases.

Exercise 31 Calculate $\lambda\big((-\infty, 9]\big)$, $\lambda\big((3, 5)\big)$, $\lambda\big(\{3\}\big)$, and $\lambda\big(\mathbb{Z}\big)$. $\circ$

3.4 Measurable functions

Let $\overline{\mathbb{R}}$ be the set of real numbers extended with $-\infty$ and $+\infty$:

$$\overline{\mathbb{R}} = [-\infty, \infty] = \mathbb{R} \cup \{\infty, -\infty\}.$$

For some mathematical background on $\overline{\mathbb{R}}$ we refer to the Appendix. We would like to define the Lebesgue integrals for functions $f : S \longrightarrow \overline{\mathbb{R}}$, *i.e.*, functions that can assign ∞ and $-\infty$ to elements of S. We have to restrict to functions that behave nicely with respect to the underlying σ-field, so called *measurable functions.*

Definition 3.7 *Let $(S, \mathcal{A})$ be a measurable space. A function $f : S \longrightarrow \overline{\mathbb{R}}$ is called $\mathcal{A}$-measurable (or just measurable if no confusion can occur) if $\{s \in S \mid f(s) \leqslant a\} \in \mathcal{A}$ for each $a \in \overline{\mathbb{R}}$.*

The following exercise shows that a σ-field $\mathcal{A}$ must be relatively large in order to obtain a reasonable collection of $\mathcal{A}$-measurable functions.

Exercise 32 Determine all $\mathcal{A}$-measurable functions in the case:

a) $\mathcal{A} = \{\emptyset, S\}$,
Hint. The answer should be something like: ‘all functions on S with the property that ...’. Just try an example, *e.g.*, , $S = \mathbb{R}$ and $f(x) = x^2$. What feature makes it (not) $\mathcal{A}$-measurable?

b) $S = \{1, 2, 3\}$ and $\mathcal{A} = \big\{\emptyset, \{1\}, \{2, 3\}, S\big\}$,

c) $\mathcal{A} = \mathcal{P}(S)$. $\circ$

Exercise 33 Let $(S, \mathcal{A})$ be a measurable space and let $A \subseteq S$. Show that the indicator function $\mathbb{1}_A$ is $\mathcal{A}$-measurable if and only if $A \in \mathcal{A}$. $\circ$

The following theorem shows that the property of measurability survives several operations. Firstly, we provide some notations:

$$f \vee g := \max\{f, g\}, \quad f^+ := \max\{f, 0\},$$
$$f \wedge g := \min\{f, g\}, \quad f^- := -\min\{f, 0\}.$$

We have: $(f \vee g) + (f \wedge g) = f + g,\ f^+ - f^- = f$ and $f^+ + f^- = |f|$.

Proposition 3.8 *Let $(S, \mathcal{A})$ be a measurable space.*

(i) *If f is $\mathcal{A}$-measurable and $a \in \overline{\mathbb{R}}$, then $a \cdot f$ is $\mathcal{A}$-measurable.*

(ii) *If f and g are $\mathcal{A}$-measurable, then $f \vee g$, $f \wedge g$, f^+, f^- and $|f|$ are all $\mathcal{A}$-measurable.*

(iii) *If $f_1, f_2, \ldots$ are $\mathcal{A}$-measurable and $\lim\limits_{k\to\infty} f_k = f$ pointwise, then f is $\mathcal{A}$-measurable.*

(iv) *If f and g are $\mathcal{A}$-measurable and for each s in S, $f(s)+g(s)$ is defined,[3] then $f+g$ is an $\mathcal{A}$-measurable function.*

(v) *If f and g are $\mathcal{A}$-measurable, then $f \cdot g$ is $\mathcal{A}$-measurable.*

Exercise 34 Prove that, as Proposition 3.8(ii) postulates, if f and g are $\mathcal{A}$-measurable, then so is $f \wedge g$. ∘

Exercise 35 Let $S \subseteq \mathbb{R}$ be an interval, possibly unbounded. Show that every increasing function $f : S \longrightarrow \mathbb{R}$ is $\mathcal{B}$-measurable. ∘

3.5 Integrals of simple functions

This section contains the first definition of a Lebesgue integral. We start to define integrals for functions that are as basic as possible and build our way to more and more complex functions.

Probably the most basic type of functions in this context will be indicator functions. Let $(S, \mathcal{A})$ be a measurable space and let A be an $\mathcal{A}$-measurable subset of S. What should be the integral of the indicator function $\mathbb{1}_A$? Let us recall the area-interpretation of an integral. The area below the graph of $\mathbb{1}_A$ and above the horizontal axis has a block type of structure with height 1 and its 'width' is the size of A. But what is the size of A? That depends on the measure we choose. If $\mu : \mathcal{A} \longrightarrow [0, \infty]$ is chosen, then the only reasonable definition of the integral of $\mathbb{1}_A$ is $\mu(A)$. This is the way it is denoted:

$$\int \mathbb{1}_A \, \mathrm{d}\mu \;=\; \mu(A).$$

From functions that assign only the values 0 and 1, we will go to nonnegative functions that assign finitely many values.

[3] *i.e.*, $\{f(s), g(s)\} \neq \{-\infty, \infty\}$, see the Appendix.

Definition 3.9 *A function $f : S \to \mathbb{R}_+$ is called simple if*

(i) $f \geqslant 0$,

(ii) f *is $\mathcal{A}$-measurable,*

(iii) f *assigns only finitely many different values.*

The class of simple functions is denoted by $\mathcal{T}^+$.

Let $f \in \mathcal{T}^+$ and suppose that f attains n different positive values, say $a_1, a_2, \dots, a_n$. Define for all i in $\{1, \dots, n\}$, A_i to be the inverse image of a_i:

$$A_i = f^{-1}(a_i) = \{s \in S \mid f(s) = a_i\}.$$

Then $A_i \in \mathcal{A}$ (because f is $\mathcal{A}$-measurable), the sets A_i are pairwise disjoint and for all $x \in S$, we have $f(x) = a_1 \mathbb{1}_{A_1}(x) + \cdots + a_n \mathbb{1}_{A_n}(x)$, or shortly

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}.$$

The right hand sided part of this equation is called the *standard form* of f. Integrals should commute with additions and multiplications in the sense that $\int (f+g) = \int f + \int g$ and $\int cf = c \cdot \int f$, just like we are used to. Taking this into account, there is only one way to define the Lebesgue integral of a simple function.

Definition 3.10 *Let $(S, \mathcal{A}, \mu)$ be a measured space. If $f \in \mathcal{T}^+$ has standard form $\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$, then its Lebesgue-integral $\int f \, \mathrm{d}\mu$ is defined by*

$$\int f \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$

Notice that the integral is well defined and can be infinite.

Example 10 Let $S = [0, 8]$, take the Borel-σ-field $\mathcal{B}$ and the Borel-Lebesgue-measure λ. Let $f = \mathbb{1}_{[0,4]} + 2 \cdot \mathbb{1}_{[3,8]}$. There are three possible f-values for elements of S, *i.e.*,

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, 3), \\ 2 & \text{if } x \in (4, 8], \\ 3 & \text{if } x \in [3, 4]. \end{cases}$$

Therefore the standard form of f is $f = 1 \cdot \mathbb{1}_{[0,3)} + 2 \cdot \mathbb{1}_{(4,8]} + 3 \cdot \mathbb{1}_{[3,4]}$. Why is this a $\mathcal{B}$-measurable function? It can be proved by the definition of a $\mathcal{B}$-measurable function directly:

$$\left\{s \in S \mid f(s) \leqslant a\right\} = \begin{cases} \emptyset & \text{if} \quad a < 1 \\ [0,3) & \text{if } 1 \leqslant a < 2, \\ [0,3) \cup (4,8] & \text{if } 2 \leqslant a < 3, \\ S & \text{if } 3 \leqslant a. \end{cases}$$

For all choices of a, the set $\left\{s \in S \mid f(s) \leqslant a\right\}$ is $\mathcal{B}$-measurable. Another way to prove that f is a $\mathcal{B}$-measurable function is to apply Exercise 33 and Proposition 3.8(i) and (iv). The Lebesgue integral of f is thereby

$$\int f \, \mathrm{d}\lambda = 1 \cdot \lambda\big([0,3)\big) + 2 \cdot \lambda\big((4,8]\big) + 3 \cdot \lambda\big([3,4]\big) = 3 + 8 + 3 = 14.$$

Verify that the Riemann integral $\int_0^8 f(x) \, \mathrm{d}x$ equals 14 as well. △

Any notion that deserves to be called 'integral' should obey the following properties.

Proposition 3.11 *Properties of the Lebesgue integral on $\mathcal{T}^+$:*

(i) $0 \leqslant \int f \, \mathrm{d}\mu \leqslant \infty$ *for each* f *in* $\mathcal{T}^+$*,*

(ii) [Monotonicity] *if* $f, g \in \mathcal{T}^+$ *and* $f \leqslant g$*, then* $\int f \, \mathrm{d}\mu \leqslant \int g \, \mathrm{d}\mu$*,*

(iii) $\int c \cdot f \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu$ *for all* $c \in [0, \infty]$ *and* $f \in \mathcal{T}^+$*,*

(iv) [Additivity] *if* $f, g \in \mathcal{T}^+$*, then* $f + g \in \mathcal{T}^+$ *and* $\int (f+g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$*.*

Proof We only provide a proof for additivity. It is clear that $f + g \in \mathcal{T}^+$. Let f and g have the respective standard forms $\sum_{i=1}^n a_i \mathbb{1}_{A_i}$ and $\sum_{j=1}^m b_j \mathbb{1}_{B_j}$. Define $a_0 = 0$ and $A_0 = \{x \in S : \; f(x) = 0\}$. Similarly, let $b_0 = 0$ and $B_0 = \{x \in S : \; g(x) = 0\}$. Then

$$f + g = \sum_{i=0}^n \sum_{j=0}^m (a_i + b_j) \mathbb{1}_{A_i \cap B_j}.$$

This is not the standard form of $f + g$, because $a_i + b_j$ can equal $a_k + b_\ell$ for different choices of indices, $a_0 + b_0 = 0$, and $A_i \cap B_j$ can be the empty set.

Nevertheless,

$$\begin{aligned} \int (f+g) \, \mathrm{d}\mu &= \sum_{i=0}^n \sum_{j=0}^m (a_i + b_j) \mu(A_i \cap B_j) \\ &= \sum_{i=0}^n \sum_{j=0}^m a_i \mu(A_i \cap B_j) + \sum_{j=0}^m \sum_{i=0}^n b_j \mu(A_i \cap B_j) \\ &= \sum_{i=0}^n a_i \mu(A_i) + \sum_{j=0}^m b_j \mu(B_j) \\ &= \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu. \end{aligned}$$

The third equality uses that for each i the sets $A_i \cap B_0,\ A_i \cap B_1,\ A_i \cap B_2,\ \ldots,\ A_i \cap B_m$ are pairwise disjoint and their union is A_i. $\square$

Exercise 36 We have shown that the Stieltjes integral $\int_0^2 \mathbb{1}_{\{1\}}\, \mathrm{d}\mathbb{1}_{[1,2]}$ does not exist (Exercise 10d). Consider the measured space $([0,2], \mathcal{B}, \mu)$, where the measure $\mu : \mathcal{B} \longrightarrow \{0,1\}$ is the probability measure satisfying for all $c, d \in [0,2]$ with $c < d$:

$$\mu\big((c,d]\big) = \begin{cases} 1 & \text{if } c < 1 \leqslant d, \\ 0 & \text{otherwise.} \end{cases}$$

It is not asked to prove that μ is uniquely determined by this property. Show that the Lebesgue integral $\int \mathbb{1}_{\{1\}}\, \mathrm{d}\mu$ does exist and compute its value. $\circ$

Proposition 3.8(iii) implies that each function $S \longrightarrow \overline{\mathbb{R}}$, that is the pointwise limit of a sequence of simple functions, is $\mathcal{A}$-measurable. The converse statement is valid as well. Even stronger, we can always find an increasing sequence of functions. Let $f_1, f_2, f_3, \ldots$ be a sequence of $\mathcal{A}$-measurable functions with shared domain S. Then $f_n \nearrow f$ denotes: the sequence of functions is increasing, *i.e.*, $f_1(x) \leqslant f_2(x) \leqslant f_3(x) \ldots$ for all x in S, and has pointwise limit f.

Theorem 3.12 *If f is $\mathcal{A}$-measurable and non-negative, then $f_n \nearrow f$ for some sequence of simple functions $f_1, f_2, \ldots$.*

We will not prove the theorem, but the following example gives an idea how a proof could look like.

Example 11 Consider the function $f(x) = x^2 \quad (x \in [0,1))$. Define the sequence of functions $f_1, f_2, \ldots$ by

$$f_n(x) = \frac{k}{2^n} \text{ in which } k \in \{0,1,2,3,\ldots,2^n - 1\} \text{ obeys } \frac{k}{2^n} \leqslant f(x) < \frac{k+1}{2^n}.$$

The function f_1 rounds off all values of f downward to 0 or $\frac{1}{2}$. The function f_2 rounds off downward to quarters, the function f_3 to multiples of $\frac{1}{8}$ and so on. It might help to sketch the graphs of f_1, f_2 and f_3 before continuing reading.

We will show that $f_n \nearrow f$. Let $n \in \mathbb{N}$ and take x in $[0,1)$. To prove that the sequence of functions increases, it suffices to show that $f_n(x) \leqslant f_{n+1}(x)$. Let k be such that $\frac{k}{2^n} \leqslant f(x) < \frac{k+1}{2^n}$. This is equivalent with $\frac{2k}{2^{n+1}} \leqslant f(x) < \frac{2k+2}{2^{n+1}}$. Define

$$\ell = \begin{cases} 2k & \text{if } f(x) < \frac{2k+1}{2^{n+1}}, \\ 2k+1 & \text{if } \frac{2k+1}{2^{n+1}} \leqslant f(x). \end{cases}$$

In both cases $\frac{\ell}{2^{n+1}} \leqslant f(x) < \frac{\ell+1}{2^{n+1}}$. Hence,

$$f_n(x) = \frac{k}{2^n} = \frac{2k}{2^{n+1}} \leqslant \frac{\ell}{2^{n+1}} = f_{n+1}(x).$$

It remains to show that $\lim_{n\to\infty} f_n(x) = f(x)$. For fixed x in $[0,1)$, the choice of k only depends on n. For all n in $\mathbb{N}$, define k_n such that $\frac{k_n}{2^n} \leqslant f(x) < \frac{k_n+1}{2^n}$. Then we have

$$|f(x) - f_n(x)| = f(x) - f_n(x) = f(x) - \frac{k_n}{2^n} < \frac{k_n+1}{2^n} - \frac{k_n}{2^n} = \frac{1}{2^n} \underset{n\to\infty}{\longrightarrow} 0.$$

Notice that we have hardly used that $f(x) = x^2$; the method applies quite general. △

The example illustrates the ideas of introduction section 3.1. f_n is a simple function with 2^n values. These values divide the range $[0,1)$ of f into 2^n levels. For each $x \in [0,1)$, $f_n(x)$ is the highest level below $f(x)$. The Lebesgue integrals $\int f_n \,\mathrm{d}\mu$ can be interpreted to be underestimations of a, to be defined, Lebesgue integral $\int f \,\mathrm{d}\mu$ of f. This idea will be elaborated in the next section.

3.6 The integrals of non-negative measurable functions

Our next class of functions of which we define the Lebesgue integrals will be the class of non-negative $\mathcal{A}$-measurable functions. Because of the previous theorem, this is the class of all functions that are the pointwise limit of an increasing sequence of simple functions. The collection of non-negative $\mathcal{A}$-measurable functions is denoted by $\mathcal{I}^+$.

Definition 3.13 *Let $(S,\mathcal{A},\mu)$ be a measured space. If $f \in \mathcal{I}^+$, then*

$$\int f \,\mathrm{d}\mu = \sup\left\{ \int g \,\mathrm{d}\mu \;\middle|\; g \in \mathcal{T}^+, g \leqslant f \right\}.$$

This definition resembles the one of a Stieltjes integral. If $g \leqslant f$ for some simple function g, then $\int g \,\mathrm{d}\mu$ is an underestimation of the Lebesgue integral of f and $\sup\left\{ \int g \,\mathrm{d}\mu \;\middle|\; g \in \mathcal{T}^+, g \leqslant f \right\}$ is the least upper bound of all such underestimations. It might be striking that Lebesgue does not need overestimations. In the special case that f is a simple function, this definition coincides with the old definition (due to Proposition 3.11, Monotonicity). It is immediately clear that

(i) $\int af \,\mathrm{d}\mu = a\int f \,\mathrm{d}\mu$ if $a \in [0,\infty]$ and $f \in \mathcal{I}^+$.

(ii) $\int f \,\mathrm{d}\mu \leqslant \int g \,\mathrm{d}\mu$ if $f, g \in \mathcal{I}^+$ and $f \leqslant g$.

Example 12 In the introduction section 3.1, we wondered whether it is possible to regard summation as a special type of integration. Now we have the tools to provide an example how. Let $a_1, a_2, a_3, \ldots$ a sequence of nonnegative real numbers. Take $S = \mathbb{N}$; the index set of the sequence. Let $\mathcal{A} = \mathcal{P}(\mathbb{N})$, so every function on $\mathbb{N}$ is $\mathcal{A}$-measurable, and choose the counting measure τ in order to assure that every number in the sequence counts for one. Define $f : \mathbb{N} \longrightarrow \mathbb{R}_+$ by $f(i) = a_i$ for all $i \in \mathbb{N}$. Verify that

$$\int f \, \mathrm{d}\tau = \sum_{i \in \mathbb{N}} a_i. \qquad \triangle$$

The following theorem is the basis for several theorems concerning convergence.

Theorem 3.14 [Monotone Convergence Theorem]

Let $(S, \mathcal{A}, \mu)$ be a measured space. Let $f_1, f_2, \ldots \in \mathcal{I}^+$ and $f_k \nearrow f$. Then $f \in \mathcal{I}^+$ and

$$\int f_k \, \mathrm{d}\mu \underset{k\to\infty}{\longrightarrow} \int f \, \mathrm{d}\mu.$$

The proof of this theorem is facultative.

Proof Because of Proposition 3.8(iii) we know that $f \in \mathcal{I}^+$. Since $f_1 \leqslant f_2 \leqslant \ldots$ and (thereby) $f_k \leqslant f$ for every k, we have: $\int f_1 \, \mathrm{d}\mu \leqslant \int f_2 \, \mathrm{d}\mu \leqslant \ldots$ and $\int f_k \, \mathrm{d}\mu \leqslant \int f \, \mathrm{d}\mu$ for every k. Hence, limit $\lim\limits_{k\to\infty} \int f_k \, \mathrm{d}\mu$ exists and is bounded from above by $\int f \, \mathrm{d}\mu$. In order to show that $\lim\limits_{k\to\infty} \int f_k \, \mathrm{d}\mu \geqslant \int f \, \mathrm{d}\mu$ it is, because of the definition of $\int f \, \mathrm{d}\mu$, sufficient to show that $\lim\limits_{k\to\infty} \int f_k \, \mathrm{d}\mu \geqslant \int g \, \mathrm{d}\mu$ for every g in $\mathcal{T}^+$ with $g \leqslant f$. Take such a simple function g. Let $\sum_{i=1}^n a_i \mathbb{1}_{A_i}$ be the standard form of g. Let $\varepsilon \in (0, 1)$.

For each $k \in \mathbb{N}$, let $$E_k = \{s \in S \mid f_k(s) \geqslant (1-\varepsilon)g(s)\}.$$

Each E_k is $\mathcal{A}$-measurable and $f_k \geqslant (1-\varepsilon)g\mathbb{1}_{E_k}$ (why?). Furthermore, $E_1 \subseteq E_2 \subseteq \cdots$ and $\bigcup\limits_{k=1}^{\infty} E_k = S$ (verify this). Therefore, for each $i \in \{1, \ldots, n\}$, the sequence $A_i \cap E_1, A_i \cap E_2, A_i \cap E_3, \ldots$ is increasing and has union $\bigcup\limits_k (A_i \cap E_k) = A_i \cap \left(\bigcup\limits_k E_k\right) = A_i \cap S = A_i$. By Theorem 3.6(ii) , we obtain $\mu(A_i) = \lim\limits_{k\to\infty} \mu(A_i \cap E_k)$.

For each k, the function $g\mathbb{1}_{E_k}$ is simple and has standard form $\sum_{i=1}^n a_i \mathbb{1}_{(A_i \cap E_k)}$. Hence,

$$\int g\mathbb{1}_{E_k} \, \mathrm{d}\mu = \sum_{i=1}^n a_i \mu(A_i \cap E_k) \underset{k\to\infty}{\longrightarrow} \sum_{i=1}^n a_i \mu(A_i) = \int g \, \mathrm{d}\mu.$$

Therefore, $$\lim_{k\to\infty} \int f_k d\mu \geqslant \lim_{k\to\infty} \int (1-\varepsilon) g\mathbb{1}_{E_k} \, \mathrm{d}\mu = (1-\varepsilon) \int g \, \mathrm{d}\mu.$$

This inequation is valid for all ε in $(0, 1)$, so $\lim\limits_{k\to\infty} \int f_k \, \mathrm{d}\mu \geqslant \int g \, \mathrm{d}\mu$. $\square$

A corollary of the Monotone Convergence Theorem is that if we construct an increasing sequence of simple functions $f_1, f_2, \ldots$ with pointwise limit f, then $\lim\limits_{k\to\infty} \int f_k \, \mathrm{d}\mu$ equals $\int f \, \mathrm{d}\mu$. Example 11 shows such a construction. The following example shows why the sequence of functions must be increasing.

Example 13 Let $S = \mathbb{R}_+, \mathcal{A} = \mathcal{B}$, and $\mu = \lambda$. Define for each k in $\mathbb{N}$, $f_k = \frac{1}{k}\mathbb{1}_{[0,k]}$. Then f_k is a simple function and $\int f_k \,\mathrm{d}\lambda = \frac{1}{k}\lambda\big([0,k]\big) = \frac{1}{k} \cdot k = 1$. The pointwise limit of the sequence $f_1, f_2, \ldots$ is the zero function. We see that in this example limit and integral do not commute: $\lim\limits_{k\to\infty} \int f_k \,\mathrm{d}\lambda = 1 \neq 0 = \int \big(\lim\limits_{k\to\infty} f_k\big) \,\mathrm{d}\lambda$ $\triangle$

Exercise 37 Let $S = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ and let τ be the counting measure at $\mathcal{A}$. Define $f : \mathbb{N} \longrightarrow \overline{\mathbb{R}}$ by $f(n) = 2^{-n}$.

a) Motivate why f is $\mathcal{A}$-measurable, but not simple.

b) Find a sequence of simple functions $f_1 : S \longrightarrow \mathbb{R}$, $f_2 : S \longrightarrow \mathbb{R}$, $f_3 : S \longrightarrow \mathbb{R}$, ... of simple functions that converge pointwise and monotone to f.

c) Give the standard notation of f_k. $(k \in \mathbb{N})$

d) Give $\int f_k \,\mathrm{d}\tau$, $(k \in \mathbb{N})$

e) Apply the Monotone Convergence Theorem to find $\int f \,\mathrm{d}\tau$. $\circ$

Exercise 38 Let $(\mathbb{R}_+, \mathcal{B}, \lambda)$ be the measured space in which $\mathcal{B}$ the Borel-σ-field over $\mathbb{R}_+$ and λ is the Borel-Lebesgue measure. Define $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{R}_+ \backslash \mathbb{Q}. \end{cases}$$

Show that $f \in \mathcal{I}^+$ and determine $\int f \,\mathrm{d}\lambda$. $\circ$

Exercise 39 Let $S = \mathbb{N} \cup \{0\}$, $\mathcal{A} = \mathcal{P}(\mathbb{N} \cup \{0\})$ and let c, r and β be strictly positive real numbers. Let μ be the measure $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}_+$ defined by $\mu(A) = \sum_{i \in A} \beta^i$ for all $A \in \mathcal{A}$. Define $f : \mathbb{N} \cup \{0\} \longrightarrow \mathbb{R}$ by $f(i) = c(1+r)^i$. Show that $f \in \mathcal{I}^+$ and $\int f \,\mathrm{d}\mu = \frac{c}{1-\beta(1+r)}$ if $\beta(1+r) < 1$ and $\int f \,\mathrm{d}\mu = \infty$ if $\beta(1+r) \geqslant 1$. $\circ$

The previous exercise models the following situation. c denotes some contribution that is donated to some organization on a yearly basis. r denotes the interest rate. As an inflation correction each year the donation is adjusted by a factor $1 + r$. β can be interpreted as the yearly factor with which capital decreases in value because of inflation. $\mu(\{k\})$, *i.e.*, the measure of year k, equals β^k and represents the factor with which an amount must by multiplied to find its utility in case the amount will be donated k years in the future. The integral represents the utility of the complete cash flow.

Proposition 3.15 *Let $(S, \mathcal{A}, \mu)$ be a measured space. Then for every $f, g \in \mathcal{I}^+$:*

$$\int (f+g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu.$$

Proof In Section 3.5 we have seen that there exist increasing sequences of simple functions $f_1, f_2, \dots$ and $g_1, g_2, \dots$, such that $f_k \nearrow f$ and $g_k \nearrow g$ and that for every $k : \int (f_k + g_k)\,\mathrm{d}\mu = \int f_k\,\mathrm{d}\mu + \int g_k\,\mathrm{d}\mu$. The Monotone Convergence Theorem implies

$$\begin{aligned}\int (f+g)\,\mathrm{d}\mu &= \lim_{n\to\infty} \int (f_n + g_n)\,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \Big(\int f_n\,\mathrm{d}\mu + \int g_n\,\mathrm{d}\mu \Big) \\ &= \lim_{n\to\infty} \int f_n\,\mathrm{d}\mu + \lim_{n\to\infty} \int g_n\,\mathrm{d}\mu \\ &= \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu.\end{aligned}$$

□

Exercise 40 Apply the previous proposition, induction, and the Monotone Convergence Theorem to show that if $g_1, g_2, g_3, \dots$ is a sequence of functions in $\mathcal{I}^+$, then

$$\sum_{i=1}^{\infty} g_i \in \mathcal{I}^+ \qquad \text{and} \qquad \int \sum_{i=1}^{\infty} g_i\,\mathrm{d}\mu = \sum_{i=1}^{\infty} \int g_i\,\mathrm{d}\mu.$$

∘

3.7 The integral of measurable functions

The largest class of functions that we will discuss is the one of measurable functions of which the range is $\overline{\mathbb{R}}$. With the help of the integrals on functions in $\mathcal{I}^+$, we can extend the notion of Lebesgue integral to this class. A natural way to do this is by means of the functions f^+ and f^- (see Section 3.4): $\int f\,\mathrm{d}\mu = \int f^+\,\mathrm{d}\mu - \int f^-\,\mathrm{d}\mu$. The only problem that can occur is that this may result in $\infty - \infty$. If that is the case, the function f is not integrable.

Definition 3.16 *Let $(S, \mathcal{A}, \mu)$ be a measured space. An $\mathcal{A}$-measurable function $f : S \longrightarrow \overline{\mathbb{R}}$ is called Lebesgue integrable if at least one of the functions f^+ and f^- in $\mathcal{I}^+$ has a finite Lebesgue integral:*

$$\int f^+\,\mathrm{d}\mu < \infty \ \text{ or } \ \int f^-\,\mathrm{d}\mu < \infty \ \text{ (or both).}$$

Its Lebesgue integral is defined by

$$\int f\,\mathrm{d}\mu = \int f^+\,\mathrm{d}\mu - \int f^- d\mu.$$

The class of Lebesgue integrable functions is denoted by $\mathcal{I}$.

Notice that the new definition coincides with the old one for f in $\mathcal{I}^+$, because in that case f equals f^+ and f^- is the zero function. To obtain some feeling for the definition, let us consider an easy example.

Example 14 Suppose a lottery ticket costs € 4,-. It gives the right to throw a dice and the outcome will be payed back in euros. What will be the expected loss?

In order to model this situation in the Lebesgue setting, choose $S = \{1, \dots, 6\}$, $\mathcal{A} = \mathcal{P}(S)$, and μ the uniform probability measure, *i.e.*, $\mu(A) = \frac{1}{6}|A|$ $(A \in \mathcal{A})$. Because the loss will be $4 - s$ for each outcome s of the roll of the dice, choose $f : S \longrightarrow \mathbb{R}$ to be $f(s) = 4 - s$.

Then $\quad f^+ = \max(f, 0) = 3 \cdot \mathbb{1}_{\{1\}} + 2 \cdot \mathbb{1}_{\{2\}} + \mathbb{1}_{\{3\}}$

and $\quad f^- = -\min(f, 0) = \mathbb{1}_{\{5\}} + 2 \cdot \mathbb{1}_{\{6\}}$.

Therefore $\quad \int f^+ \,\mathrm{d}\mu = 3\mu(\{1\}) + 2\mu(\{2\}) + \mu(\{3\}) = 1$

and $\quad \int f^- \,\mathrm{d}\mu = \mu(\{5\}) + 2\mu(\{6\}) = \frac{1}{2}$.

This results in $\quad \int f \,\mathrm{d}\mu = 1 - \frac{1}{2} = \frac{1}{2}$, which indeed equals the gamblers expected loss. $\triangle$

Proposition 3.17 *Let f and g be functions in $\mathcal{I}$.*

(i) *If $c \in \mathbb{R}$, then $f + g \in \mathcal{I}, cf \in \mathcal{I}, \int (f + g) \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\mu$ and $\int cf \,\mathrm{d}\mu = c \int f \,\mathrm{d}\mu$,*

(ii) *if $f \geqslant 0$, then $\int f \,\mathrm{d}\mu \geqslant 0$,*

(iii) *if $f \geqslant g$, then $\int f \,\mathrm{d}\mu \geqslant \int g \,\mathrm{d}\mu$,*

(iv) *$|f| \in \mathcal{I}$ and $|\int f \,\mathrm{d}\mu| \leqslant \int |f| \,\mathrm{d}\mu$.*

The Monotone Convergence Theorem provides conditions under which limit and integral can switch positions. The following theorem uses another set of criteria. It will be applied in the final two sections.

Theorem 3.18 [Lebesgue's Dominated Convergence Theorem]

Let $(S, \mathcal{A}, \mu)$ be a measured space. Let $f_1, f_2, \ldots : S \longrightarrow \overline{\mathbb{R}}$ be a sequence of $\mathcal{A}$-measurable functions that pointwise converges to a function $f : S \longrightarrow \overline{\mathbb{R}}$. Assume the existence of a function g in $\mathcal{I}^+$ such that $\int g \,\mathrm{d}\mu < \infty$ and $|f_n| \leqslant g$ for every n.

Then $f, f_1, f_2, \ldots \in \mathcal{I}$ and

$$\int f_n \,\mathrm{d}\mu \underset{n\to\infty}{\longrightarrow} \int f \,\mathrm{d}\mu.$$

Here is an example to show that sometimes the Dominated Convergence Theorem can be applied much easier than the Monotone Convergence Theorem:

Example 15 Let $S = (-\pi, \pi)$ and define $f_k : S \longrightarrow [0, 1]$ by

$$f_k(x) = \mathbb{1}_{(-\pi+\frac{1}{k},\pi-\frac{1}{k})}(x) \cdot \sin(x).$$

Such functions are $\mathcal{B}$-measurable, since for every $a \in \overline{\mathbb{R}}$, $k \in \mathbb{N}$, the set $\{x \in (-\pi, \pi) : f_k(x) \leqslant a\}$ is empty, an interval, or the union of two intervals. The sequence $f_1, f_2, \ldots$ converges pointwise (even uniformly) to the sine function. The functions are not nonnegative and the sequence does not converge monotonically, so the Monotone Convergence Theorem cannot be applied right away. The Dominated Convergence Theorem can: define $g := \mathbb{1}_{(-\pi,\pi)}$. Then $|f_k(x)| \leqslant g(x)$ for all $x \in S$ and $\int g \, \mathrm{d}\lambda = 2\pi < \infty$. Hence, $\lim\limits_{k\to\infty} \int f_k \, \mathrm{d}\lambda = \int \sin(x) \, \mathrm{d}\lambda$, which will be shown to be equal to $\int_{-\pi}^{\pi} \sin(x) \, \mathrm{d}x = 0$ in Section 3.9. $\triangle$

3.8 Fubini's Theorem

Probably the most frequently applied result of this course is Fubini's theorem. It gives conditions under which it is possible to switch the order of two integrals or of two summations. We will not discuss the general version concerning arbitrary and abstract measured spaces $(S, \mathcal{A}, \mu)$, but only two more down to earth versions. The first one concerns Riemann integrals and has already been given in Section 1.4. The second concerns double summations. Let us first consider an exercise that shows that is necessary to be careful when switching summations.

Exercise 41 Consider the $\infty \times \infty$-matrix $A = \begin{pmatrix} 3 & -3 & 0 & 0 & \cdots \\ 0 & 3 & -3 & 0 & \cdots \\ 0 & 0 & 3 & -3 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}$.

a) Calculate $\sum\limits_{i=1}^{\infty} \sum\limits_{j=1}^{\infty} A_{ij}$ and $\sum\limits_{j=1}^{\infty} \sum\limits_{i=1}^{\infty} A_{ij}$.

Let $S = \mathbb{N} \times \mathbb{N}$, $\mathcal{A} = \mathcal{P}(S)$ and let τ be the counting measure on $\mathcal{A}$. Let $f : S \longrightarrow \mathbb{R}$ be defined by $f(i, j) = A_{ij}$ for each $(i, j) \in S$.

b) Show that the function f is $\mathcal{A}$-measurable.

c) Determine $\int f^+ \, \mathrm{d}\tau$ and $\int f^- \, \mathrm{d}\tau$.

d) Determine whether f is an element of $\mathcal{I}$ or not.

Now, consider the $\infty \times \infty$-matrix B defined by $B = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \cdots \\ 0 & \frac{1}{4} & -\frac{1}{8} & 0 & \cdots \\ 0 & 0 & \frac{1}{8} & -\frac{1}{16} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}$.

e) Compute $\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty} B_{ij}$ and $\sum\limits_{j=1}^{\infty}\sum\limits_{i=1}^{\infty} B_{ij}$.

Let $g : S \longrightarrow \mathbb{R}$ be defined by $g(i,j) = B_{ij}$ for each $(i,j) \in S$.

f) Determine $\int g^+ \, \mathrm{d}\tau$, $\int g^- \, \mathrm{d}\tau$ and $\int g \, \mathrm{d}\tau$. $\circ$

The following theorem generalizes the exercise above:

Theorem 3.19 [Fubini's Theorem for summations]

Let A_{ij} be real numbers ($i, j \in \mathbb{N}$). If all numbers are nonnegative or if $\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty}|A_{ij}| < \infty$, then

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}A_{ij} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}A_{ij}.$$

Proof Let $S = \mathbb{N} \times \mathbb{N}$, $\mathcal{A} = \mathcal{P}(S)$ and let τ be the counting measure on $\mathcal{A}$. Let $f : S \longrightarrow \mathbb{R}$ be defined by $f(i,j) = A_{ij}$ for each $(i,j) \in S$.

First assume that $A_{ij} \geqslant 0$ for all $i, j \in \mathbb{N}$. Then $f \in \mathcal{I}^+$. Let $R_i : S \longrightarrow \mathbb{R}$ be the function that is identical to f on row i and to the zero function elsewhere, so

$$R_i(k,j) = \begin{cases} A_{ij} & \text{if } k = i, \\ 0 & \text{else.} \end{cases}$$

Similarly, let $C_j : S \longrightarrow \mathbb{R}$ be the function that is identical to f on column j and to the zero function elsewhere. Applying the Monotone Convergence Theorem gives that $\int R_i \, \mathrm{d}\tau = \sum\limits_{j=1}^{\infty}A_{ij}$ and that $\int C_j \, \mathrm{d}\tau = \sum\limits_{i=1}^{\infty}A_{ij}$ (why?). Since $\sum\limits_{i=1}^{\infty}R_i = \sum\limits_{j=1}^{\infty}C_j = f$, Exercise 40 tells that $\sum\limits_{i=1}^{\infty}\int R_i \, \mathrm{d}\tau = \sum\limits_{j=1}^{\infty}\int C_j \, \mathrm{d}\tau = \int f \, \mathrm{d}\tau$. This all combined yields

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}A_{ij} = \int f \, \mathrm{d}\tau = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}A_{ij}.$$

Now assume that $\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty}|A_{ij}| < \infty$. Then $f^+ \in \mathcal{I}^+$, $f^- \in \mathcal{I}^+$, $\int f^+ \, \mathrm{d}\tau < \infty$, and $\int f^- \, \mathrm{d}\tau < \infty$. The first part of this theorem gives that $\int f^+ \, \mathrm{d}\tau = \sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty}A_{ij}^+ = \sum\limits_{j=1}^{\infty}\sum\limits_{i=1}^{\infty}A_{ij}^+$ as well as $\int f^- \, \mathrm{d}\tau = \sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty}A_{ij}^- = \sum\limits_{j=1}^{\infty}\sum\limits_{i=1}^{\infty}A_{ij}^-$, in which A^+ and A^- are defined as usual. Therefore,

$$\begin{aligned}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}A_{ij} &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}A_{ij}^{+} - \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}A_{ij}^{-}\\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}A_{ij}^{+} - \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}A_{ij}^{-}\\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}A_{ij}.\end{aligned}$$

□

3.9 Riemann and Stieltjes integration vs. Lebesgue integration

This section shows that Riemann and Stieltjes integrability each imply Lebesgue integrability with respect to an appropriate measure.

Consider a Riemann (Stieltjes) integrable function f over an interval $[a, b]$. Let us assume for ease that f is non-negative (otherwise we should consider f^+ and f^-). The idea is that any subset D of $\mathbb{R}^2$ that is the finite union of disjoint rectangles and that is situated above the x-axis, below the graph of f and between the vertical lines $x = a$ and $x = b$ can be used to provide both an underestimation for the Riemann (Stieltjes) integral and a simple function that is smaller than f.

Theorem 3.20 *Let $f : [a, b] \longrightarrow \mathbb{R}$ be Riemann integrable. Then f is Lebesgue integrable as well with respect to the Borel-Lebesgue measure λ and*

$$\int f \,\mathrm{d}\lambda = \int_a^b f(x)\,\mathrm{d}x.$$

Proof Assume for ease that $f \geqslant 0$ (otherwise we can consider first f^+ and f^-). Let $p_1, p_2, p_3, \ldots$ be a sequence of underestimations that converges to the Riemann integral $\int f(x)\,\mathrm{d}x$.

Fix a natural number i. The underestimation p_i is constructed using some division $(s_0, \ldots, s_n)$ and non-negative reals $\ell_1, \ldots, \ell_n$ such that $p_i = \sum\limits_{k=1}^{n}(s_k - s_{k-1})\ell_k$. Define the simple function $f_i : [a, b] \longrightarrow \mathbb{R}$ by

$$f_i = \ell_1 \mathbb{1}_{[s_0,s_1]} + \ell_2 \mathbb{1}_{(s_1,s_2]} + \cdots + \ell_n \mathbb{1}_{(s_{n-1},s_n]}.$$

We have

$$\int f_i \,\mathrm{d}\lambda = \sum_{k=1}^{n} \ell_k \lambda([s_{k-1}, s_k]) = \sum_{k=1}^{n}(s_k - s_{k-1})\ell_k = p_i.$$

Because of the assumption that $D_1 \subset D_2 \subset \cdots$, we have that $f_i(x) \leqslant f_{i+1}(x)$ for all $i \in \mathbb{N}$ and $x \in [a, b]$. Therefore $f_i \nearrow \ell \leqslant f$ for some function ℓ. The Monotone Convergence Theorem states that ℓ is Lebesgue integrable and

$$\int \ell \,\mathrm{d}\lambda = \lim_{i\to\infty}\int f_i \,\mathrm{d}\lambda = \lim_{i\to\infty} p_i = \int_a^b f(x)\,\mathrm{d}x.$$

Similarly, by using overestimations, a Lebesgue integrable function $h \geqslant f$ can be found with $\int h \, \mathrm{d}\lambda = \int_a^b f(x) \, \mathrm{d}x$. It can be shown that if ℓ and h are functions with the same Lebesgue integral and $\ell \leqslant f \leqslant h$, then f is also Lebesgue integrable. Because of Proposition 3.17(iii), f has this integral as well. $\square$

Theorem 3.21 *Let $f : [a,b] \longrightarrow \mathbb{R}$ be Stieltjes integrable with respect to integrator g. Then f is Lebesgue integrable with respect to the unique measure $\mu_g : \mathcal{B} \longrightarrow \mathbb{R}_+$ that obeys $\mu_g\big((c,d]\big) = g(d) - g(c)$ for every interval $(c,d] \subset [a,b]$. Furthermore,*

$$\int f \, \mathrm{d}\mu_g = \int f \, \mathrm{d}g.$$

The proof follows the lines of the previous one.

We conclude this section by considering (semi-)infinite Riemann integrals. Let $f : [0,\infty) \longrightarrow \mathbb{R}$ be a function. A natural choice for defining $\int_0^\infty f(x) \, \mathrm{d}x$ would be $\lim\limits_{k\to\infty} \int_0^k f(x) \, \mathrm{d}x$, or, in terms of λ, $\lim\limits_{k\to\infty} \int f \mathbb{1}_{[0,k]} \, \mathrm{d}\lambda$. Given that $\int_0^k f(x) \, \mathrm{d}x$ exists for all $k > 0$, does this limit always exist? The answer is no, just think of the sine function. However, there are two (by now familiar) sufficient conditions.

- Lebesgue's Dominated Convergence Theorem provides a sufficient condition: the functions $|f\mathbb{1}_{[0,1]}|$, $|f\mathbb{1}_{[0,2]}|$, $|f\mathbb{1}_{[0,3]}|, \ldots$ all must be smaller than some function with a finite Lebesgue integral. The smallest function that dominates all of these functions is $|f|$. Hence, if $|f| \in \mathcal{I}$ and $\int |f| \, \mathrm{d}\lambda < \infty$, then $\lim\limits_{k\to\infty} \int_0^k f(x) \, \mathrm{d}x = \lim\limits_{k\to\infty} \int f\mathbb{1}_{[0,k]} \, \mathrm{d}\lambda = \int \lim\limits_{k\to\infty} f\mathbb{1}_{[0,k]} \, \mathrm{d}\lambda = \int f \, \mathrm{d}\lambda$, so $\lim\limits_{k\to\infty} \int_0^k f(x) \, \mathrm{d}x$ exists. Hence, it is justified to speak of the semi-infinite integral $\int_0^\infty f(x) \, \mathrm{d}x$. According to Theorem 3.20 and the Dominated Convergence Theorem it equals $\int f \, \mathrm{d}\lambda$.

- The Monotone Convergence Theorem provides another sufficient condition: f must be non-negative and Lebesgue measurable, i.e., $f \in \mathcal{I}^+$. In that case, the sequence of functions $f\mathbb{1}_{[0,1]}, f\mathbb{1}_{[0,2]}, f\mathbb{1}_{[0,3]}, \ldots$ converges monotonically to f. Again, $\lim\limits_{k\to\infty} \int f\mathbb{1}_{[0,k]} \, \mathrm{d}\lambda = \int f \, \mathrm{d}\lambda$ and we can copy the previous argument to infer that it is justified to speak of the semi-infinite integral $\int_0^\infty f(x) \, \mathrm{d}x$.

Similar statements can be formulated in case the domain is $(-\infty, 0]$ or $\mathbb{R}$.

3.10 Two applications in the theory of probability

Well written articles in probability theory usually contain the phrase 'Let $\langle \Omega, \mathcal{A}, \mathbb{P} \rangle$ be a probability space …'. Here, $\langle \Omega, \mathcal{A}, \mathbb{P} \rangle$ is a measured space in which $\mathbb{P}$ is a probability measure. Ω plays the role of S. It is often called the *sample space*, each of whose members is thought of as a potential outcome of a random experiment. $\mathcal{A}$ is a σ-field of subsets of Ω. The members of $\mathcal{A}$ are called *events*. E.g., if we go back to Example 8 in which a coin and a dice are thrown, the set $\{(\text{head}, 1), (\text{tail}, 1)\}$ models the event that the dice roll results '1'. A random variable is formally a measurable function $X : \Omega \longrightarrow \mathbb{R}$, so random variables are $\mathcal{A}$-measurable by definition. Statements like $\mathbb{P}\big(\{\omega \in \Omega \mid X(\omega) \leqslant 3\}\big) = \frac{1}{2}$ are usually denoted by the much more transparent notation $\mathbb{P}[X \leqslant 3] = \frac{1}{2}$. The Lebesgue integral $\int X \, \mathrm{d}\mathbb{P}$ can be used as the definition of $\mathbb{E}[X]$. It covers the old definitions for discrete and continuous random variables, but it is far more general.

Exercise 42 Let $\langle \Omega, \mathcal{A}, \mathbb{P} \rangle$ be a probability space and let X be a random variable with values in $\mathbb{N}$. Define for all $k \in \mathbb{N}$ the function X_k by ($\omega \in \Omega$)

$$X_k(\omega) = \begin{cases} X(\omega) & \text{if } X(\omega) \leqslant k, \\ 0 & \text{if } X(\omega) > k. \end{cases}$$

a) Why are the functions X_k $\mathcal{A}$-measurable (and thereby random variables)?

b) Use the monotone convergence theorem to show $\displaystyle\int X \, \mathrm{d}\mathbb{P} = \sum_{i=1}^{\infty} i \mathbb{P}[X = i]$. ◦

This section translates the monotone convergence theorem and the Dominated Convergence Theorem in terms of probabilities and expectations. It has the advantage that they can be applied without referring to Lebesgue theory at all. Nevertheless, one should keep in mind that a solid mathematical basis is lying underneath such statements.

Theorem 3.22 *Let $X, X_1, X_2, \ldots$ be a sequence of non-negative random variables such that $\mathbb{P}[X_k \leqslant X_{k+1}$ for all $k \in \mathbb{N}$ and $\lim_{k \to \infty} X_k = X] = 1$. Then*

$$\lim_{k \to \infty} \mathbb{E}[X_k] = \mathbb{E}[X].$$

In order to prove the theorem, we only need one assertion:

> *Let X and Y be random variables such that $\mathbb{P}[X = Y] = 1$ and $\mathbb{E}[X]$ exists. Then $\mathbb{E}[Y]$ exists as well and equals $\mathbb{E}[X]$.*

The proof of this trivial looking statement is omitted.

Proof of the theorem Let $A \in \mathcal{A}$ be the event

$$\{\omega \in \Omega \mid X_k(\omega) \leqslant X_{k+1}(\omega) \text{ for all } k \in \mathbb{N} \text{ and } \lim_{k\to\infty} X_k(\omega) = X(\omega)\}.$$

Then $\mathbb{P}(A) = 1$. Define $\bar{X}_k$ by $\bar{X}_k(\omega) = \begin{cases} X_k(\omega) & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$

Define $\bar{X}$ similarly. Because of the statement above we have that $\mathbb{E}[X_k] = \mathbb{E}[\bar{X}_k]$ and $\mathbb{E}[X] = \mathbb{E}[\bar{X}]$ (the expectations exist because the random variables are non-negative). For all ω in Ω and k in $\mathbb{N}$, we have that $\bar{X}_k(\omega) \leqslant \bar{X}_{k+1}(\omega)$ and $\lim_{i\to\infty} \bar{X}_i(\omega) = \bar{X}(\omega)$. In other words, $\bar{X}_k \nearrow \bar{X}$. Hence, we can apply the monotone convergence theorem by choosing $f_k = \bar{X}_k$ and $f = \bar{X}$. We find that

$$\lim_{k\to\infty} \mathbb{E}[X_k] = \lim_{k\to\infty} \mathbb{E}[\bar{X}_k] = \lim_{k\to\infty} \int \bar{X}_k \, \mathrm{d}\mathbb{P} = \int \bar{X} \, \mathrm{d}\mathbb{P} = \mathbb{E}[\bar{X}] = \mathbb{E}[X]. \qquad \square$$

Theorem 3.23 *Let $X, X_1, X_2, X_3, \ldots$ be a sequence of random variables satisfying $\mathbb{P}[\lim_{k\to\infty} X_k = X] = 1$. If there exists a random variable Y with finite mean such that $\mathbb{P}[|X_k| \leqslant Y] = 1$ for all $k \in \mathbb{N}$, then the expectations $\mathbb{E}[X]$ and $\mathbb{E}[X_k]$ exist, are all finite, and*

$$\lim_{k\to\infty} \mathbb{E}[X_k] = \mathbb{E}[X].$$

The proof of Theorem 3.23 follows the lines of the proof above. It is a translation of the Dominated Convergence Theorem into the language of probability theory.

Exercise 43 Let $\Omega = (0, 1)$. Then $\langle \Omega, \mathcal{B}, \lambda \rangle$ is a probability space. Define the sequence of random variables $X_1, X_2, \cdots$ by

$$X_k(\omega) = \begin{cases} k & \text{if } \omega \in (0, \frac{1}{k}], \\ 0 & \text{if } \omega \in (\frac{1}{k}, 1). \end{cases}$$

a) Determine the random variable X to which this sequence converges pointwise. Determine $\lim_{k\to\infty} \mathbb{E}[X_k]$ and $\mathbb{E}[X]$.

b) Let $Y := \sup_{k\in\mathbb{N}} X_k$. Show that $\mathbb{E}[Y] = \infty$. $\circ$

3.11 Extra exercises

X1. Let $S = [1, 9]$ and define, for all $n \in \mathbb{N}$, $f_n : S \longrightarrow \mathbb{R}$ by

$$f_n = 3 \cdot \mathbb{1}_{[2\ ,\ 6-\frac{1}{n}]} - 5 \cdot \mathbb{1}_{[4\ ,\ 7]}.$$

Define $\mathcal{G} = \{(a, b] \mid 1 \leqslant a < b \leqslant 9\}$ and $\mathcal{A} = \mathcal{A}(\mathcal{G})$, *i.e.*, the σ-field generated by $\mathcal{G}$. Let λ be the Borel-Lebesgue-measure restricted to S.

a) Show that $[2, 4) \in \mathcal{A}$ and that $\lambda([2, 4)) = 2$.

b) Provide the standard forms of f_n^+ and f_n^- $(n \in \mathbb{N})$.

We note that f_n^+ and f_n^- are $\mathcal{B}$-measurable (a proof is not asked for) for all $n \in \mathbb{N}$.

c) Determine, for all $n \in \mathbb{N}$, $\int f_n^+ \, \mathrm{d}\lambda$, $\int f_n^- \, \mathrm{d}\lambda$, and $\int f_n \, \mathrm{d}\lambda$.

d) Let f be the pointwise limit of the sequence $f_1, f_2, \ldots$. Determine by means of the Dominated Convergence Theorem the Lebesgue integral $\int f \, \mathrm{d}\lambda$ (the Monotone Convergence Theorem works as well, but is more elaborative).

X2. Consider the measured space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \tau)$. Let $a_1, a_2, a_3, \ldots$ be a sequence of real numbers. This sequence corresponds to the function $f_a : \mathbb{N} \longrightarrow \mathbb{R}$, defined by $f_a(n) = a_n$ for all $n \in \mathbb{N}$. Show that if $\displaystyle\int f_a \, \mathrm{d}\tau$ is well defined, then $\displaystyle\lim_{n\to\infty} \sum_{i=1}^{n} a_i$ either converges or diverges to $\pm\infty$, and $\displaystyle\int f_a \, \mathrm{d}\tau = \sum_{i=1}^{\infty} a_i$.

Hint: First apply the Monotone Convergence Theorem or Exercise 40 to show that if $f \geqslant 0$, then $\displaystyle\int f \, \mathrm{d}\tau = \lim_{n\to\infty} \sum_{i=1}^{n} a_i$.

X3. Consider the measured space $\langle S, \mathcal{A}, \mu \rangle$ with $S = \mathbb{N} = \{1, 2, 3, \ldots\}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$, and μ is defined by

$$\mu(A) = \sum_{i \in A} 2^{-i}. \qquad A \in \mathcal{P}(\mathbb{N})$$

Define $f : \mathbb{N} \longrightarrow \mathbb{R}$ by $f(n) = (-1)^n \cdot n^2$ for all $n \in \mathbb{N}$. Furthermore, define for each $k \in \mathbb{N}$ the function $f_k : \mathbb{N} \longrightarrow \mathbb{R}$ by

$$f_k(n) = \begin{cases} f(n) & \text{if } 1 \leqslant n \leqslant k, \\ 0 & \text{if } n > k. \end{cases}$$

a. Does the sequence of functions $f_1, f_2, f_3, \ldots.$ converge pointwise to f?

b. Does the sequence of functions $f_1, f_2, f_3, \ldots.$ converge uniformly to f?

c. Does the sequence of functions $f_1, f_2, f_3, \ldots.$ converge monotonically to f?

d. Provide the standard forms of f_k^+ and f_k^-.

e. Determine $\int f_k^+ \, \mathrm{d}\mu$ and $\int f_k^- \, \mathrm{d}\mu$. $(k \in \mathbb{N})$

f. Determine $\int f^+ \, \mathrm{d}\mu$ and $\int f^- \, \mathrm{d}\mu$. It is not necessary to compute occurring limits (just leave these limits in the answer), but indicate whether (you think that) they are finite or not.

g. Determine $\int f \, \mathrm{d}\mu$ or show that this integral is not defined. If you are not able to answer **3f** (fully), describe how an answer of **3f** could be used to answer this question.

A Appendix; The extended set of real numbers $\overline{\mathbb{R}}$

In order to define functions that can have ∞ and $-\infty$ as values, we have defined, in Section 3.4, $\overline{\mathbb{R}}$ to be $\mathbb{R} \cup \{-\infty, \infty\}$. This appendix provides the necessary mathematical background. The *ordering* of $\mathbb{R}$ is y extended to an ordering of $\overline{\mathbb{R}}$ b $\quad -\infty \leqslant a,\ a \leqslant \infty$ for all a in $\overline{\mathbb{R}}$.

Because of this ordering we can speak of intervals as subsets of $\overline{\mathbb{R}}$. Moreover we can use notations like $[0, \infty](= \overline{\mathbb{R}}_+)$ and $(-\infty, \infty]$.

The *sum*-operator of $\mathbb{R}$ is a map of $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$. We extend it to a map of (almost) $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$ to $\overline{\mathbb{R}}$ as follows:

$$\begin{array}{ccccccccl} a & + & \infty & := & \infty & + & a & := & \infty \quad \text{for } a \in (-\infty, \infty], \\ a & + & (-\infty) & := & (-\infty) & + & a & := & -\infty \quad \text{for } a \in [-\infty, \infty). \end{array}$$

$(-\infty) + \infty$ and $\infty + (-\infty)$ remain *undefined.* Multiplication is extended as well:

$$\begin{array}{ccccccccl} a & \cdot & \infty & := & \infty & \cdot & a & := & \infty \quad \text{for } a \in (0, \infty], \\ a & \cdot & (-\infty) & := & (-\infty) & \cdot & a & := & -\infty \quad \text{for } a \in (0, \infty], \\ a & \cdot & \infty & := & \infty & \cdot & a & := & -\infty \quad \text{for } a \in [-\infty, 0), \\ a & \cdot & (-\infty) & := & (-\infty) & \cdot & a & := & \infty \quad \text{for } a \in [-\infty, 0) \end{array}$$

and, less obviously,

$$\begin{array}{ccccccccl} 0 & \cdot & \infty & := & \infty & \cdot & 0 & := & 0, \\ 0 & \cdot & (-\infty) & := & (-\infty) & \cdot & 0 & := & 0. \end{array} .$$

Let $V \subseteq \overline{\mathbb{R}}$. A real number a in $\overline{\mathbb{R}}$ is an upper bound (*majorant*) of V if $V \subseteq [-\infty, a]$ and a lower bound (*minorant*) if $[a, \infty] \supseteq V$. An upper bound is called the *supremum* of V if all other upper bounds are larger. A lower bound is called the *infimum* of V if all other lower bounds are smaller.

Proposition A.1 *Every non-empty subset of $\overline{\mathbb{R}}$ has a supremum and an infimum in $\overline{\mathbb{R}}$. Every increasing sequence in $\overline{\mathbb{R}}$ converges to its supremum. Every decreasing sequence converges to its infimum.*

The difference with $\mathbb{R}$ is that boundedness does not have to be assumed anymore.

B Answers to exercises

Exercise 1: b^2.

Exercise 2: **a)** 0 and 0. **b)** The main point is to understand that 0 is an underestimation, but not an overestimation.

Exercise 4: $\underline{\int_a^b} \mathbb{1}_{\mathbb{Q}} \, \mathrm{d}x = 0$ while $\overline{\int_a^b} \mathbb{1}_{\mathbb{Q}} \, \mathrm{d}x = b - a$.

Exercise 5: You'll have to find sequences of underestimations and overestimations having the same limit. If you can guess that this limit is zero, you can work your way to it.

The underestimation 0 is readily given. Take the division $(s_0, s_1) = (0, 1)$ and let $\ell_1 = 0$. This leads to $p = (s_1 - s_0) \cdot 0 = 0$. Hence, the constant sequence $0, 0, \ldots$ will do.

Finding the sequence of overestimations is less trivial. Let $n \in \mathbb{N}$ and let s be the division with 2^n subsegments of equal size. The first subsegment $[0, \frac{1}{2^n}]$ catches all but $n+1$ real numbers that have g-value 1. If you set $h_1 = 1$, only the numbers $\frac{1}{2^n}, \frac{1}{2^{n-1}}, \ldots, \frac{1}{2}$, and 1 still need attention. Set $h_2 = h_3 = 1$ to take care of $\frac{1}{2^n}$ and $\frac{1}{2^{n-1}}$; the division points s_1 and s_2 (draw a sketch). Set $h_{2^n} = 1$ because $s_{2^n} = 1 \in g$. The remaining numbers, $\frac{1}{2^{n-2}}, \ldots, \frac{1}{2}$, are all division points and are each element of 2 subsegments. So, each of them causes two subsegments to have an overestimation h_k of at least 1. In total $2n$ bars, all of length $\frac{1}{2^n}$, must have height 1, all other upper bars h_k can be chosen to be 0. This leads to the overestimation $q_n = 2n \cdot \frac{1}{2^n}$, which tends to 0 when n tends to infinity.

Exercise 7: $\int_a^b f \, \mathrm{d}37 = 0$.

Exercise 8: **a)** Since integrand f is increasing, we get the best underestimations by choosing ℓ_k equal to the function value of the left hand side border of the k^{th} interval: $\ell_k = f(s_{k-1}) = s_{k-1} = \frac{4k-4}{n}$. The importance of a subinterval is 0 if F is constant on that subinterval and $\frac{1}{3}$ if the subinterval contains a discontinuity-point of F:

$$g(s_k) - g(s_{k-1}) = \begin{cases} \frac{1}{3} & \text{if } s_k \in \{1, 2, 3\}, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} \frac{1}{3} & \text{if } k \in \{\frac{1}{4}n, \frac{1}{2}n, \frac{3}{4}n\}, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the greatest underestimation with this division equals

$$\begin{aligned} p_n &= \sum_{k=1}^{n} \big(g(s_k) - g(s_{k-1})\big)\ell_k \\ &= \sum_{k \in \{\frac{1}{4}n, \frac{1}{2}n, \frac{3}{4}n\}} \tfrac{1}{3} \cdot \tfrac{4k-4}{n} \\ &= 2 - \tfrac{4}{n}. \end{aligned}$$

b) Likewise, choose $h_k = f(s_k) = s_k$ for each k to obtain $q_n = 2$.

Exercise 9: If the domain I equals $[0,1]$, we can estimate $|f(x) - f(y)|$ for every pair x, y in I by

$$|f(x) - f(y)| = |x^2 - y^2| = |x + y| \cdot |x - y| \leqslant 2|x - y|.$$

Hence, given ε, we should choose δ smaller than $\frac{1}{2}\varepsilon$.

If the domain is $\mathbb{R}_+$ however, $|x + y|$ can be arbitrarily large. In order to show that the function is not uniformly continuous on $\mathbb{R}_+$, we may choose an $\varepsilon > 0$, must let δ be arbitrary, and must choose x and y such that $|x - y| \leqslant \delta$ and $|f(x) - f(y)| > \varepsilon$. The choices $\varepsilon = 1$, $x = \delta + \frac{1}{\delta}$, and $y = \frac{1}{\delta}$ will do:

$$|f(x) - f(y)| = |(\tfrac{1}{\delta} + \delta)^2 - (\tfrac{1}{\delta})^2| = \delta^2 + 2 > \varepsilon.$$

Exercise 10: **a)** 8, **b)** 1, **c)** 0, **d)** undefined.

Exercise 11: Integration by parts leads to $\int_a^b F \,\mathrm{d}F = F(b)F(b) - F(a)F(a) - \int_a^b F \,\mathrm{d}F$. Therefore $\int_a^b F \,\mathrm{d}F = \frac{1}{2}$. Drawing the graph of F with respect to F leads to a line from $\big(F(a), F(a)\big)$ to $\big(F(b), F(b)\big)$, *i.e.*, a line from $(0,0)$ to $(1,1)$. The area under this line equals $\frac{1}{2}$ indeed.

Exercise 13: $\displaystyle\int_1^4 x^2 \,\mathrm{d}\tfrac{1}{3}x = 7.$

Exercise 14: The sequence of functions converges to $f = \mathbb{1}_{\{1\}}$. Let $\varepsilon = \frac{1}{4}$ and define, for each $k \in \mathbb{N}$, $x_k = \sqrt[k]{\frac{1}{2}}$. Then $|f(x_k) - f_k(x_k)| > \varepsilon$ for all $k \in \mathbb{N}$.

Exercise 15: Use, *e.g.*, the bijections $x \mapsto -5x + 2$ and $x \mapsto \ln(x)$.

Exercise 16: An enumeration of $\mathbb{Z}$ is $z_1 = 0, z_2 = 1, z_3 = -1, z_4 = 2, z_5 = -2, \ldots$. $\mathbb{Q}$ can be enumerated just like $\mathbb{Q} \cap [0,1]$. If S is an infinite subset of $T = \{t_1, t_2, t_3, \cdots\}$, let n_1 be the smallest natural number with $t_{n_1} \in S$, n_2 be the second smallest natural number with $t_{n_2} \in S$, and so on. Then $s_1 = t_{n_1}$, $s_2 = t_{n_2} \cdots$ is an enumeration of S.

Exercise 17: *E.g.*, order by increasing index-sum, choose a tie-breaking rule.

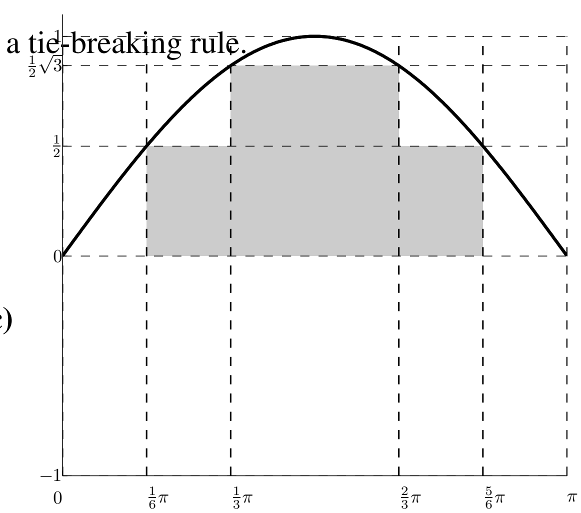
Exercise 18: Apply Exercise 16.

Exercise 19: $\forall k \in \mathbb{N} : |f_k(q_{k+1}) - f(q_{k+1})| > \frac{1}{2} =: \varepsilon.$

Exercise 20: **a)**

$$\begin{aligned} A_1 &= \{0, \pi\}, \\ A_2 &= [0, \tfrac{1}{6}\pi] \cup [\tfrac{5}{6}\pi, \pi], \\ A_3 &= [\tfrac{1}{6}\pi, \tfrac{1}{3}\pi] \cup [\tfrac{2}{3}\pi, \tfrac{5}{6}\pi], \\ A_4 &= [\tfrac{1}{3}\pi, \tfrac{2}{3}\pi]. \end{aligned}$$

c)



Exercise 21:

a) Define $B_1 = B_3 = B_5 = \cdots = A_1$ and $B_2 = B_4 = B_6 = \cdots = A_2$. Then $A_1 \cup A_2 = \bigcup_{k=1}^{\infty} B_k \in \mathcal{A}$,

b) $A_1 \backslash A_2 = S \backslash [(S \backslash A_1) \cup A_2] \in \mathcal{A}$,

c) $A_1 \cap A_2 = S \backslash [(S \backslash A_1) \cup (S \backslash A_2)] \in \mathcal{A}$,

d) $\bigcap\limits_{k=1}^{\infty} A_k = S \backslash [\bigcup\limits_{k=1}^{\infty} (S \backslash A_k)] \in \mathcal{A}$.

Exercise 22: 5.

Exercise 23: **a)** Define $A_k = \{2k\}$ for all $k \in \mathbb{N}$. Then $\bigcup\limits_{k=1}^{\infty} A_k \notin \mathcal{G}$.

Exercise 25: Choose $\mathcal{A}_1 = \big\{\emptyset, \{1,2,3\}, \{1\}, \{2,3\}\big\}$ and $\mathcal{A}_2 = \big\{\emptyset, \{1,2,3\}, \{3\}, \{1,2\}\big\}$.

Exercise 26: $\mathcal{A} = \big\{\emptyset, S, \{1\}, \{2,3,4\}, \{1,2\}, \{3,4\}, \{2\}, \{1,3,4\}\big\}$.

Exercise 28:

Because $\mathcal{B} = \sigma(\mathcal{G}_1)$, it is requested to show that $\sigma(\mathcal{G}_1) = \sigma(\mathcal{G}_3)$. Proposition 3.3 gives that it is sufficient to show that $\mathcal{G}_3 \subset \sigma(\mathcal{G}_1)$ and $\mathcal{G}_1 \subset \sigma(\mathcal{G}_3)$.

Take an arbitrary element $(-\infty, a]$ of $\mathcal{G}_3$. Then for every $n \in \mathbb{N}$, the interval $(a-n, a]$ is an element of $\mathcal{G}_1$ and therefore also of $\sigma(\mathcal{G}_1)$. Since $\sigma(\mathcal{G}_1)$ is a σ-field by definition, we get

$$(-\infty, a] = \bigcup_{n=1}^{\infty} (a-n, a] \in \sigma(\mathcal{G}_1).$$

Now take an arbitrary element $(a, b]$ of $\mathcal{G}_1$. We have

$$(a, b] = (-\infty, b] \backslash (-\infty, a] \in \sigma(\mathcal{G}_3).$$

Exercise 29:

$\mu_c(\emptyset) = 0$, because the empty set is countable. In order to prove the second requirement, take a sequence $A_1, A_2, \ldots$ of pairwise disjoint elements of $\mathcal{A}$.

If all A_i's are countable, then $\sum_i^{\infty} \mu_c(A_i) = 0 = \mu_c\big(\bigcup\limits_{i=1}^{\infty} A_i\big)$.

If all A_i's *but one*, say A_k, are countable, then $\sum_{i=1}^{\infty} \mu_c(A_i) = c = \mu_c\big(\bigcup\limits_{i=1}^{\infty} A_i\big)$.

Finally, suppose that two elements of the sequence are cocountable, say A_k and A_ℓ. Then $S \backslash A_k$ is countable. Since $A_k \cap A_\ell = \emptyset$, we have that $A_\ell \subset S \backslash A_k$. So, a cocountable set were a subset of a countable set. Why is this impossible?

Exercise 30:

Let t in $\mathbb{N}$ be such that $\mu(B_t) < \infty$ and define $A_n := B_t \backslash B_n$ for all $n \geqslant t$.

We have $\emptyset = A_t \subseteq A_{t+1} \subseteq A_{t+2} \subseteq \ldots$, so (ii) is valid for this sequence, yielding

$$\mu(\bigcup_{n>t} A_n) = \lim_{n\to\infty} \mu(A_n) \quad \text{and} \quad \mu(\bigcup_{n>t} A_n) = \mu(\bigcup_{n>t} B_t \backslash B_n)$$

$$= \lim_{n\to\infty} \mu(B_t \backslash B_n) \qquad = \mu(B_t \backslash \bigcap_{n>t} B_n)$$

$$\doteq \mu(B_t) - \lim_{n\to\infty} \mu(B_n) \qquad \doteq \mu(B_t) - \mu(\bigcap_{n>t} B_n).$$

The assumption that $\mu(B_t) < \infty$ is used at the $\doteq$-signs.

Exercise 31: ∞, 2, 0, and 0.

Exercise 32:

a) Only constant functions are $\mathcal{A}$-measurable.

b) f is $\mathcal{A}$-measurable if and only if $f(2) = f(3)$.

c) All functions on S are $\mathcal{A}$-measurable.

Exercise 34:

Let $h := f \vee g$. Then

$\{s \in S \mid h(s) \leqslant a\} = \{s \in S \mid f(s) \leqslant a \text{ and } g(s) \leqslant a\} = \{s \in S \mid f(s) \leqslant a\} \cap \{s \in S \mid g(s) \leqslant a\}$.

Since intersections of measurable sets are measurable, $\{s \in S \mid h(s) \leqslant a\}$ is measurable.

Exercise 36:

$\{1\} \in \mathcal{B}$, since $\{1\} = \bigcap_{i=1}^{\infty}(1 - \frac{1}{n}, 1]$. Theorem 3.6(ii) gives that

$$\int f_2 \,\mathrm{d}\mu = \int \mathbb{1}_{\{1\}} \,\mathrm{d}\mu = \mu(\{1\}) = \lim_{n\to\infty} \mu((1 - \tfrac{1}{n}, 1]) = \lim_{n\to\infty} g_2(1) - g_2(1 - \tfrac{1}{n}) = 1 - 0 = 1.$$

This is the desirable outcome, because if X is a random variable with distribution function g_2, then $\mathbb{E}[f_2(X)] = \sum_{x:\mathbb{P}[X=x]>0} f_2(x)\mathbb{P}[X = x] = f_2(1) = 1$.

Exercise 37: **e)** 1.

Exercise 38: 0.

Exercise 40:

Define $f_n = \sum_{i=1}^{n} g_i$ and $f = \sum_{i=1}^{\infty} g_i$. Then $f_n \in \mathcal{I}^+$, $\int f_n \mu = \sum_{i=1}^{n} \int g_i \,\mathrm{d}\mu$, and $f_n \nearrow f$. Apply the Monotone Convergence Theorem.

Exercise 41:

a) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij} = 0. \qquad \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} A_{ij} = 3.$

b) Since $\mathcal{A} = \mathcal{P}(S)$, every subset of S is $\mathcal{A}$-measurable, in particular every subset of the form $\{(i, j) \in S \mid f(i, j) \leqslant a\}$. Hence, every function is $\mathcal{A}$-measurable.

c) + d) Since both $\int f^+ \,\mathrm{d}\tau$ and $\int f^- \,\mathrm{d}\tau$ are infinite, $\int f \,\mathrm{d}\tau$ is not defined and $f \notin \mathcal{I}$.

e) $\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty} B_{ij} = \frac{1}{2}$. $\qquad \sum\limits_{j=1}^{\infty}\sum\limits_{i=1}^{\infty} B_{ij} = \frac{1}{2}$.

f) Define $g_k^+ := \sum_{i=1}^{k} \frac{1}{2^i} \cdot \mathbb{1}_{\{(i,i)\}}$. Apply the MCT to find $\int g^+ \,\mathrm{d}\tau = 1$. Likewise, define $g_k^- := \sum_{i=1}^{k} \frac{1}{2^{i+1}} \cdot \mathbb{1}_{\{(i,i+1)\}}$ to find $\int g^- \,\mathrm{d}\tau = \frac{1}{2}$. Apparently $f \in \mathcal{I}$ and $\int g \,\mathrm{d}\tau = \frac{1}{2}$.

Exercise 42:

a) $\{\omega \in \Omega \mid X_k(\omega) \leqslant a\} = \{\omega \in \Omega \mid X(\omega) \leqslant a\} \cup \{\omega \in \Omega \mid X(\omega) > k\}$ if $a \geqslant 0$ and $\{\omega \in \Omega \mid X_k(\omega) \leqslant a\} = \emptyset$ if $a < 0$. Since X is a random variable, it is $\mathcal{A}$-measurable, and therefore these sets are $\mathcal{A}$-measurable.

b) The standard form of X_k is $X_k = \sum_{i=1}^{k} i \cdot \mathbb{1}_{\{w \in \Omega \mid X(\omega)=i\}}$. Verify that $X_1, X_2, X_3, \ldots$ converges monotonically to X. Then $\displaystyle\int X \,\mathrm{d}\mathbb{P} = \lim_{k\to\infty} \int X_k \,\mathrm{d}\mathbb{P} = \lim_{k\to\infty} \sum_{i=1}^{k} i \cdot \mathbb{P}\big(\{w \in \Omega \mid X(\omega) = i\}\big) = \mathbb{E}[X]$.

Exercise 43:

a) We have $\mathbb{P}[X_k = k] = \frac{1}{k}$ and $\mathbb{P}[X_k = 0] = \frac{k-1}{k}$. The sequence converges pointwise *but not monotonically* to the zero-function X and $\lim\limits_{k\to\infty} \mathbb{E}[X_k] = 1 \neq 0 = \mathbb{E}[X]$.

b) If $\mathbb{E}[Y]$ were finite, Theorem 3.23 would give that $\lim\limits_{k\to\infty} \mathbb{E}[X_k] = \mathbb{E}[X]$.

Index