1 Introduction

1a	**Integral buyer's guide: know what type is better**	**1**
1b	**Dimension 2 is a different story**	**3**
1c	**Gauge: a bridge from Riemann to Lebesgue**	**6**
1d	**Small but dense**	**8**
1e	**Sandwich (Riemann), zigzag sandwich (Lebesgue)**	**9**
1f	**Some achievements of Lebesgue theory**	**10**
1g	**Integral or measure? Functions or sets?**	**11**

This lightweight introduction is not a part of the theory of Lebesgue integration. It is rather a "comparative anatomy" of integration theories[1] *intended to explain what distinguishes Lebesgue's theory from others, and why it is the best but not all-mighty.*

1a Integral buyer's guide: know what type is better

Several types of integral are well-known:

∗ the Riemann integral;
∗ the improper Riemann integral (two types: 1-dim and n-dim);
∗ the Lebesgue integral;
∗ the gauge integral;
∗ the Newton integral (that is, antiderivative).

Probably, the most videly used in practice is the improper Riemann integral; the most well-known to a wide audience is the antiderivative; the most teached are the Riemann and Lebesgue integrals; and the gauge integral is controversial, seldom used, seldom teached, but contains a bright idea able to illuminate the Lebesgue integral.

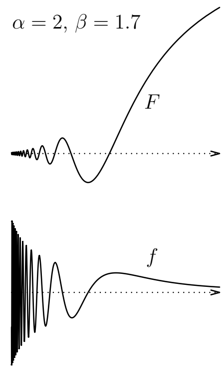
This course is devoted to the Lebesgue integral, but relations to other integrals are also of some interest.

For each type of integral we have

∗ the definition of integrable function;
∗ the definition of the integral of an integrable function;
∗ theorems on properties of the integral.

[1] This phrasing is borrowed from S. Berberian, see the quote on pp. 43-33 in "Theory of the integral" by B. Thomson.

1a1 Example.



Consider the function $F : (0, \infty) \to \mathbb{R}$,

$$F(x) = x^\alpha \sin x^{-\beta}$$

for given $\alpha, \beta \in (0, \infty)$, and its derivative

$$f(x) = F'(x) = -\beta x^{\alpha-\beta-1} \cos x^{-\beta} + \alpha x^{\alpha-1} \sin x^{-\beta} \,.$$

Note that

$$\begin{aligned}
&F(0+) = \lim_{x\to 0+} F(x) = 0 && \text{always}\,, \\
&f(0+) = \lim_{x\to 0+} f(x) = 0 && \text{if } \alpha > \beta + 1\,, \\
&\limsup_{x\to 0+} f(x) = +\infty && \text{if } \alpha < \beta + 1\,, \\
&\limsup_{x\to 0+} f(x) = \beta && \text{if } \alpha = \beta + 1\,, \\
&\liminf_{x\to 0+} f(x) = -\limsup_{x\to 0+} f(x) && \text{always}\,.
\end{aligned}$$

According to the Newton integral, f is integrable on $(0, 1)$ always, and

$$\int_0^1 f(x)\,\mathrm{d}x = F(1-) - F(0+) = \sin 1 \ (\approx 0.84)\,.$$

According to the Riemann integral, f is integrable on $(0, 1)$ if and only if $\alpha \ge \beta + 1$ (and then the integral is equal to $\sin 1$, of course).

The improper Riemann integral (1-dim) gives the same as the Newton integral (always). Also the gauge integral gives the same.

According to the Lebesgue integral, f is integrable on $(0, 1)$ if and only if $\alpha > \beta$ (and then the integral is $\sin 1$).[1] The same holds for the n-dim improper Riemann integral.[1] Ridiculously, for $n = 1$ it is still not the same as the 1-dim improper Riemann integral!

Integrals of all types apply (in particular) to compactly supported continuous functions, and conform on these functions.

Integrals of all types have three basic properties: linearity, positivity, and shift invariance.

Linearity: integrable functions (on a given set) are a vector space, and the integral is a linear functional on this space. That is, $\int af = a \int f$ and $\int (f + g) = \int f + \int g$ for all integrable f, g and all $a \in \mathbb{R}$.

[1]Since integrability of $|f|$ is necessary, see Sect. 1b.

Positivity: if $f(\cdot) \ge 0$ everywhere and f is integrable, then $\int f \ge 0$.

Monotonicity follows: if $f(\cdot) \le g(\cdot)$ everywhere and f, g are integrable, then $\int f \le \int g$.

Also, a basic convergence theorem follows: if continuous functions $f, f_1, f_2, \dots$ with a common compact support are such that $f_n \to f$ uniformly, then $\int f_n \to \int f$ (as $n \to \infty$).

Shift invariance: if $g(\cdot) = f(\cdot + x_0)$ for a given x_0 and f is integrable (on the whole space), then g is integrable and $\int g = \int f$ (integrals over the whole space).

1b Dimension 2 is a different story

In dimension 2, more advanced properies are desirable: iterated integral, and rotation invariance.

Iterated integral:

$$\int_{[a,b]\times[c,d]} f = \int_a^b \biggl(\int_c^d f(x,y)\,\mathrm{d}y \biggr) \mathrm{d}x$$

whenever f is integrable on $[a,b] \times [c,d]$. This is similar to the elementary equality

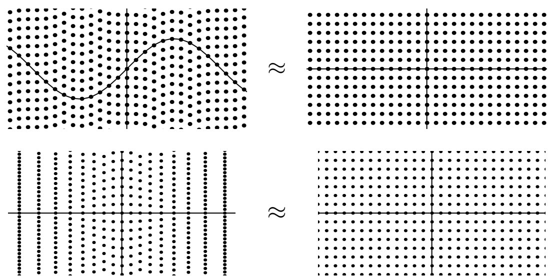
$$\sum_{(i,j)\in\{k,\dots,l\}\times\{m,\dots,n\}} a_{i,j} = \sum_{i=k}^{l} \sum_{j=m}^{n} a_{i,j}\,.$$

It holds for the Riemann integral with some technical reservations ($f(x,\cdot)$ may fail to be integrable on $[c,d]$ for some $x \in [a,b]$), and is quite problematic for the improper Riemann integral.

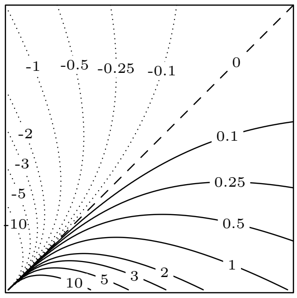
Rotation invariance: if $g(x,y) = f(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$ for a given θ and all $x, y \in \mathbb{R}$, and f is integrable (on $\mathbb{R}^2$), then g is integrable, and $\int_{\mathbb{R}^2} g = \int_{\mathbb{R}^2} f$.

In particular, when f is the indicator (in other words, characteristic function) of a disk, triangle etc., we get rotation invariance of the area.

There are other area preserving transformations; for instance, reflection $(x,y) \mapsto (y,x)$; also, $(x,y) \mapsto (cx, c^{-1}y)$ for a given $c \in (0,\infty)$. In fact, these generate all area preserving *linear* transformations of $\mathbb{R}^2$. But there are also nonlinear; for instance, $(x,y) \mapsto (x, y + \sin x)$, or $(x,y) \mapsto \bigl(x^3 + x, \frac{y}{3x^2+1}\bigr)$. They all should preserve integrals (which is a special case of the well-known change of variables).



1b1 Example.



Consider the function $f : (0,1) \times (0,1) \to \mathbb{R}$,

$$f(x,y) = \frac{x-y}{(x+y)^3}\,.$$

We have $f(x,y) = \frac{\mathrm{d}}{\mathrm{d}y}\frac{y}{(x+y)^2}$, thus $\int_0^1 f(x,y)\,\mathrm{d}y = \frac{1}{(x+1)^2}$ and $\int_0^1 \big(\int_0^1 f(x,y)\,\mathrm{d}y \big)\,\mathrm{d}x = \int_0^1 \frac{\mathrm{d}x}{(x+1)^2} = \frac12$. However, $f(y,x) = -f(x,y)$, thus, $\int_0^1 \big(\int_0^1 f(y,x)\,\mathrm{d}y \big)\,\mathrm{d}x = -\frac12$. Invariance under the reflection $(x,y) \mapsto (y,x)$ is broken. A paradox! How does it happen? Well, here is a simpler form of this paradox:

$$\tfrac12 = \tfrac12 + (-1+1) + (-1+1) + \dots = \tfrac12 - 1 + 1 - 1 + 1 - \dots = \\ = (\tfrac12 - 1) + (1-1) + (1-1) + \dots = -\tfrac12\,.$$

It is dangerous to cancel $+\infty$ and $-\infty$. Here is the positive part of our integral: $\int_0^x f(x,y)\,\mathrm{d}y = \frac{1}{4x}$, thus $\int_0^1 \big(\int_0^x f(x,y)\,\mathrm{d}y \big)\,\mathrm{d}x = \int_0^1 \frac{\mathrm{d}x}{4x} = +\infty$.

Do you feel that our integral must be 0 due to the (anti)symmetry? There is a way to force 0, the so-called Cauchy principal value. One excludes the ε-neighborhood of the singular point $(0,0)$ and takes the limit as $\varepsilon \to 0$. In dimension 1 this gives, for example,

$$\mathrm{v.\,p.} \int_{-1}^1 \frac{\mathrm{d}x}{x} = \lim_{\varepsilon\to0} \bigg(\int_{-1}^{-\varepsilon} + \int_\varepsilon^1 \bigg) = 0\,,$$

even though $\int_{-1}^1 \frac{\mathrm{d}x}{|x|} = +\infty$. The principal value is not the same as the improper integral: $\int_{-1}^1 \frac{\mathrm{d}x}{x} = \int_{-1}^0 \frac{\mathrm{d}x}{x} + \int_0^1 \frac{\mathrm{d}x}{x} = -\infty + \infty$, undefined. True, in Example 1a1, when $\alpha \le \beta$, the result ($\sin 1$) is given by the improper integral canceling $+\infty$ and $-\infty$, since $\int_0^1 |f(x)|\,\mathrm{d}x = +\infty$. Yes, but these $+\infty$ and $-\infty$ appear on the same side of the singular point. Such cancellation is rather harmless; it does not conflict with change of variables[1] (and iterated integral is irrelevant in dimension 1).

This is just a 1-dim good luck. Nothing like that is known in dim 2 (and higher).[2] Any attempt to cancel $+\infty$ and $-\infty$ conflicts with iterated integral

[1] In contrast, the 1-dim principal value conflicts with change of variables. For example, $\mathrm{v.\,p.}\int_{-1}^1 \frac{\mathrm{d}x}{x^3} = 0$, but $\mathrm{v.\,p.}\int_{\varphi^{-1}(-1)}^{\varphi^{-1}(1)} \frac{\varphi'(t)}{\varphi^3(t)}\,\mathrm{d}t = -\infty$ if $\varphi''(0) > 0$ and $+\infty$ if $\varphi''(0) < 0$.

[2] "We need to make this stronger definition of convergence in terms of $|f(x)|$ because cancellation in the integrals can occur in so many different ways in higher dimensions." Cooper, Jeffery (2005), "Working analysis", Gulf Professional (p. 538).
"However the process of defining improper integrals in dimension $n > 1$ is trickier than in dimension $n = 1$, (this is due to the great variety of ways in which a limit can be formed in $\mathbb{R}^n$)." M.A. Moskowitz, F. Paliogiannis, "Functions of several real variables" (p. 329).

(see above), and also with change of variables. Even the measure preserving transformation $(x,y) \mapsto (cx, c^{-1}y)$ is problematic:[1]

$$\text{v. p.} \int_{(0,c^{-1})\times(0,c)} \frac{cx - c^{-1}y}{(cx + c^{-1}y)^3} \, \mathrm{d}x\mathrm{d}y \ne 0 \,.$$

We conclude.

Conditionally convergent integrals are widely used in dimension 1 only. The mainstream of integration in higher dimensions stipulates absolute integrability.

Absolute integrability: if f is integrable then $|f|$ is integrable. (Here $|f|$ is the pointwise absolute value, $x \mapsto |f(x)|$.)

This is satisfied by the Riemann integral and the Lebesgue integral, but violated by the 1-dimensional improper Riemann integral, the gauge integral, the antiderivative, and the Cauchy principal value. Think twice before you apply iterated integral or change of variables to these "conditional" types of integral!

1b2 Example. One may hope that still, the calculation

$$\iint\limits_{(0,\infty)\times(0,2\pi)} \frac{\sin r}{r^2} \sin\varphi \, r \mathrm{d}r\mathrm{d}\varphi = \biggl(\int_0^\infty \frac{\sin r}{r} \, \mathrm{d}r \biggr) \biggl(\int_0^{2\pi} \sin\varphi \, \mathrm{d}\varphi \biggr) = \frac{\pi}{2} \cdot 0 = 0$$

is harmless in spite of nonabsolute convergence of $\int_0^\infty \frac{\sin r}{r} \, \mathrm{d}r$. That is,

$$\iint\limits_{\mathbb{R}^2} \frac{\sin\sqrt{x^2+y^2}}{x^2+y^2} \cdot \frac{y}{\sqrt{x^2+y^2}} \, \mathrm{d}x\mathrm{d}y = 0 \,.$$

But try the change of variables $(r,\varphi) \mapsto (r, \varphi + r)$;

$$\sin r \sin(\varphi + r) = \sin r (\sin\varphi \cos r + \cos\varphi \sin r) = \frac{1}{2} \sin\varphi \sin 2r + \cos\varphi \sin^2 2r \,;$$

integrate in r for a fixed φ:

$$\frac{1}{2} \sin\varphi \underbrace{\int_0^\infty \frac{\sin 2r}{r} \, \mathrm{d}r}_{\pi/2} + \cos\varphi \underbrace{\int_0^\infty \frac{\sin^2 r}{r} \, \mathrm{d}r}_{+\infty} \,.$$

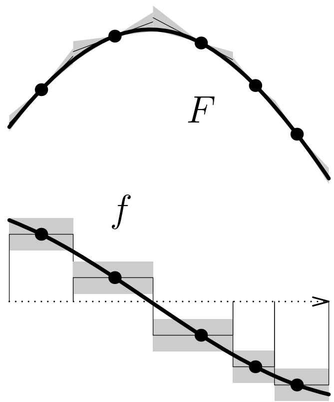
The radial integral is $+\infty$ for rays in the right half-plane, and $-\infty$ in the left half-plane. One more transformation $(r,\varphi) \mapsto \bigl(\frac{1}{r}, 4\varphi - \frac{\pi}{2}\bigr)$ with $0 < \varphi < \pi/2$ leads to a case as toxic as Example 1b1.

[1]In fact, $\iint_{x^2+y^2>\varepsilon^2} f(cx, c^{-1}y) \, \mathrm{d}x\mathrm{d}y = \frac{1}{2} \int_0^{\pi/2} \frac{\cos^2\theta - \sin^2\theta}{(c\cos\theta + c^{-1}\sin\theta)^2} = I(c)$ for all $\varepsilon < \min(c, c^{-1})$, and $I(0+) = 1/2$, $I(1) = 0$, $I(+\infty) = -1/2$.

1c Gauge: a bridge from Riemann to Lebesgue

Recall that the Riemann integral is the limit of Riemannian sums that correspond to tagged partitions, when the mesh tends to 0.

If a function $F : [0,1] \to \mathbb{R}$ is smooth enough, namely, $f = F'$ exists and is continuous, then $\int_0^1 f(x)\,\mathrm{d}x = F(1) - F(0)$, of course. Look at the picture: each term $f(t_i)(x_i - x_{i-1})$ of the Riemannian sum is the linear approximation to $F(x_i) - F(x_{i-1})$, its error being $\le \varepsilon(x_i - x_{i-1})$; thus, the sum is ε-close to $F(1) - F(0)$. Indeed,

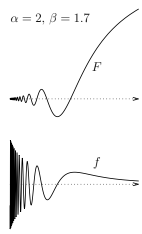


$$|F(x) - F(t) - f(t)(x - t)| \le \varepsilon |x - t|$$

whenever $|x - t| \le \delta$; it is crucial that for every ε there exists such δ that serves all t simultaneously.

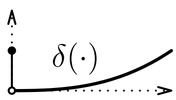
The condition $|x - t| \le \delta \implies |f'(x) - f'(t)| \le \varepsilon$ is sufficient, not necessary; note that it is sometimes violated on the given pictire.

Now we consider a harder case: the function(s) of Example 1a1 for $\max(\beta, 1) < \alpha < \beta + 1$. Clearly, the linear approximation to F near t works only for $|x^{-\beta} - t^{-\beta}| \ll 1$, that is, $|x - t| \ll t^{\beta+1}$. We need something like $\delta = \varepsilon t^{\beta+1}$ (up to a constant; never mind); the problem is that a single δ cannot serve all t near 0.

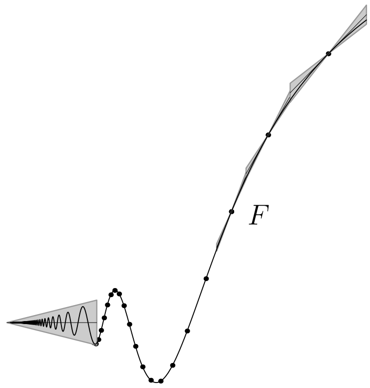


Could we take $\delta(t) = \varepsilon t^{\beta+1}$ and demand $x_i - x_{i-1} \le \delta(t_i)$? It may seem that this way leads to infinite partitions. But wait; F is differentiable at 0, and $F'(0) = 0$, since $|F(x)| \le x^\alpha \ll x$, even though f is unbounded near 0; we deal with the one-sided derivative at 0, assuming that $F(0) = 0$ (extended by continuity). The linear approximation to $F(x)$ near 0 is just 0, and $|F(x) - F(0)| \le \varepsilon |x - 0|$ for $x \le \varepsilon^{1/(\alpha-1)}$. That is,

$$\delta(t) = \begin{cases} \varepsilon^{1/(\alpha-1)} & \text{for } t = 0, \\ \varepsilon t^{\beta+1} & \text{for } t > 0. \end{cases}$$



Every point may represent its neighborhood,
but some points may represent larger neighborhoods than others.



In order to get a finite partition we must take $t_1 = 0$; then we may take $x_1 = \varepsilon^{1/(\alpha-1)}$, and divide the interval $[1, x_1^{-\beta}] = [1, \varepsilon^{-\beta/(\alpha-1)}]$ into intervals $[x_{i+1}^{-\beta}, x_i^{-\beta}]$ of length $\le \varepsilon$, getting a finite partition of (roughly) $\varepsilon^{-(\beta+1)/(\alpha-1)}$ intervals; much more than ε^{-1}, but still, finite. Thus, $\int_0^1 f(x)\,\mathrm{d}x = F(1) - F(0)$, where the integral is the **gauge integral.** Its definition looks very similar to the definition of the Riemann integral; still, "for every ε there exists δ such that for every tagged partition…"; but now δ is a function $[0,1] \to (0,\infty)$ rather than a number, and the tagged partition must be "finer than $\delta(\cdot)$", that is, safisfy $x_i - x_{i-1} \le \delta(t_i)$ (rather than $x_i - x_{i-1} \le \delta$). Surprizingly, such tagged partitions exist for *arbitrary* $\delta(\cdot)$, which is easy to prove.[1]

Similarly, $\int_0^1 F'(x)\,\mathrm{d}x = F(1) - F(0)$ for *all* differentiable $F : [0,1] \to \mathbb{R}$. Some of these F are monstrous! In particular, such F can be nowhere monotone. Did you know? Can you imagine it?[2]

Consider now the function(s) of Example 1a1 for $\alpha < 1$. In this case F has no one-sided derivative at 0, thus, $f(0)$ is undefined. So what? A single point should not contribute to integral, anyway. Let us define $f(0)$ arbitrarily.[3] We cannot say that the first term $f(0)(x_1 - 0)$ is $\varepsilon(x_1 - 0)$-close to $F(x_1) - F(0)$; so what? Both are small. A single term need not have a small *relative* error; a small *absolute* error is enough. We take $\delta(0)$ such that $|f(0)|\delta(0) + \delta^\alpha(0) \le \varepsilon$, then the first term contributes the error $|F(x_1) - F(0) - f(0)(x_1 - 0)| \le x_1^\alpha + |f(0)|x_1 \le \varepsilon$. The other terms contribute $\le \varepsilon$ as before, and so, the sum is 2ε-close to $F(1) - F(0)$. Well, having 2ε we can also get ε, of course.

We see that nondifferentiability of F at one point does not harm (as long as F is continuous). And no wonder: one point, as well as a finite set of points, does not contribute to the Riemann integral, and the more so, to the gauge integral. In contrast, the set of rational numbers contributes to the Riemann integral, but does not contribute to the gauge integral, as we'll see soon.

---

[1] Just consider the supremum u_0 of all $u \in [0,1]$ such that tagged partitions of $[0,u]$ finer than $\delta|_{[0,u]}$ exist; you know, $\delta(u_0) > 0$…

[2] See, for example, Sect. 9c of my advanced course "Measure and category".

[3] Feel free to take $f(0) = 0$; or $f(0) = -1234\pi^5$, if you prefer…

1d Small but dense

The indicator function $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$ of the set of all rational numbers on $[0,1]$ is a widely known example of a function that is not Riemann integrable. And no wonder: the δ-neighborhood of this set contains the whole $[0,1]$, for every $\delta > 0$. But now we use a gauge function $\delta(\cdot)$; what about the $\delta(\cdot)$-neighborhood?

Let us take $\delta\big(\frac{k}{n}\big) = \frac{6}{\pi^2}\frac{\varepsilon}{n^2(n+1)}$ for every irreducible fraction $\frac{k}{n}$, and $\delta(x) = 1$ for $x \in [0,1] \setminus \mathbb{Q}$. Then

$$\sum_{x\in\mathbb{Q}\cap[0,1]} \delta(x) \le \sum_{n=1}^{\infty}\sum_{k=0}^{n} \frac{6}{\pi^2}\frac{\varepsilon}{n^2(n+1)} = \varepsilon \cdot \frac{6}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2} = \varepsilon\,.$$

If a tagged partition of $[0,1]$ is finer than $\delta(\cdot)$, then the corresponding integral sum $\le \varepsilon$ (think, why). Thus, the gauge integral of $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$ is zero!

The Riemann integral does not feel the distinction between rationals and irrationals, but the gauge integral does.

By the way, existence of irrational numbers follows (think, why).[1]

The set $\cup_{x\in\mathbb{Q}\cap[0,1]}\big(x-\delta(x), x+\delta(x)\big)$ is open and contains $\mathbb{Q}\cap[0,1]$; it may be called an open neighborhood of $\mathbb{Q}\cap[0,1]$, but it is a *nonuniform* neighborhood.

More generally, let $G \subset (0,1)$ be an open set. What about the gauge integral of $\mathbb{1}_G$? If a gauge function δ on $[0,1]$ satisfies

$$\forall x \in G \;\; \big(x-\delta(x), x+\delta(x)\big) \subset G$$

(such $\delta(\cdot)$ surely exist), then for every tagged partition finer than $\delta(\cdot)$, the corresponding integral sum is the total length of finitely many nonoverlapping intervals inside G (think, why). Thus, $\int_0^1 \mathbb{1}_G(x)\,\mathrm{d}x$ does not exceed the inner Jordan measure of G. The converse inequality is trivial (think, why). This is instructive.

For an open set, its *inner* Jordan measure is relevant. For a compact set, its *outer* Jordan measure is relevant.

(For the second part, take the complement.)

The open neighborhood of $\mathbb{Q}\cap[0,1]$, constructed above, is ε-small but dense in $[0,1]$. Its closure is the whole $[0,1]$. Its boundary is large.

An open set can be much smaller than its closure (and its boundary).

[1] But did you note, where did we use the fact that $\mathbb{R}$ contains all real numbers, not only rationals?

Advice. Thinking about an arbitrary open set, do not imagine an open interval; it is too good. Rather, imagine a small but dense open set; this is nearly the worst case. Similarly, thinking about an arbitrary compact set, do not imagine a closed interval; rather, imagine the complement to a small but dense open set ("fat Cantor set"); it is large but nowhere dense.

The rationals are a countable set, but some null sets are nowhere countable. For example, all numbers of the form $\sum_{k=1}^{\infty} 3^{-k} c_k$ where $c_k \in \{0, 1, 2\}$ and $c_k = 1$ only finitely many times. Still, it can be approximated *from above* by a small open set. And a small open set can be approximated *from below* by a finite union of intervals. A zigzag approximation!

For functions, the situation is similar. The function $1\!\!1_{\mathbb{Q}}$ is not of the form $\lim_n f_n$ with continuous f_n, but is of the form $\lim_n \lim_m f_{m,n}$ with continuous $f_{m,n}$:

$$1\!\!1_{\mathbb{Q}}(x) = \lim_n \lim_m \cos^{2m}(\pi n! x)\,;$$

the convergence is monotone in both cases: decreasing in m but increasing in n. A zigzag, again.

1e Sandwich (Riemann), zigzag sandwich (Lebesgue)

It may seem that the Lebesgue integral is quite difficult to define. For example, the textbook by F. Jones defines first the Lebesgue measure, in 6 stages, and then the integral, in 2 stages. Nevertheless it is possible to define both integrals, Riemann and Lebesgue, on half a page, just now.

. . . .

For a compactly supported continuous function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} f = \lim_{m\to\infty} \frac{1}{m^d} \sum_{k\in\mathbb{Z}^d} f\Big(\frac{k}{m}\Big)\,.$$

. . . .

A function $f : \mathbb{R}^d \to \mathbb{R}$ is Riemann integrable, if there exist compactly supported continuous functions g_k, h_k such that $g_1 \le g_2 \le \dots \le f \le \dots \le h_2 \le h_1$ and $\lim_k \int g_k = \lim_k \int h_k$; in this case

$$\int f = \lim_k \int g_k = \lim_k \int h_k\,.$$

. . . .

A function $f : \mathbb{R}^d \to \mathbb{R}$ is Lebesgue integrable, if there exist compactly supported continuous functions $g_{k,l}, h_{k,l}$ such that $g_{k,1} \ge g_{k,2} \ge \dots$ and

$h_{k,1} \le h_{k,2} \le \dots$ for each k, and $g_1 \le g_2 \le \dots \le f \le \dots \le h_2 \le h_1$ where $g_k = \lim_l g_{k,l}$, $h_k = \lim_l h_{k,l}$ (pointwise; infinite values allowed), and $-\infty < \lim_k \lim_l \int g_{k,l} = \lim_k \lim_l \int h_{k,l} < +\infty$;[1] in this case

$$\int f = \lim_k \lim_l \int g_{k,l} = \lim_k \lim_l \int h_{k,l} \, .$$

. . . .

Done!

You may replace the compactly supported continuous functions with step functions. Or, if you prefer, with compactly supported infinitely differentiable functions. The final result is always the same.

Nice... but the shortest definition is not the best. It is not easy to understand and use when proving theorems. Thus, we'll follow the textbooks and define the integral differently, in several stages.

For now we note that the sandwich leads to the Riemann integral, while the zigzag sandwich leads to the Lebesgue integral. What about the next step, zigzag-zigzag sandwich and super-Lebesgue? A surprize:

One zigzag is enough.

More zigzags are futile: the result is the same Lebesgue integral, still. A good luck, isn't it?[2]

1f Some achievements of Lebesgue theory[3]

Trigonometric polynomials

$$f(x) = a_0 + \sqrt{2} \sum_{k=1}^{n} (a_k \cos 2\pi kx + b_k \sin 2\pi kx)$$

satisfy $\int_0^1 |f(x)|^2 \, \mathrm{d}x = a_0^2 + \sum_{k=1}^n (a_k^2 + b_k^2)$. They are a Euclidean space of dimension $2n+1$, w.r.t. the L_2-norm $\|f\| = \bigl(\int_0^1 |f(x)|^2 \, \mathrm{d}x \bigr)^{1/2}$. For $n \to \infty$ we hope for a bijective correspondence between sequences of coefficients

[1] Do you want to write $\int g_k$ instead of $\lim_l \int g_{k,l}$?

[2] This good luck is due to the sandwich. If you want to *represent* a given function f as $\lim_k \lim_l g_{k,l}$, you discover that sometimes you need more zigzags (and moreover, infinitely many zigzags do not suffice; countable ordinals are involved). It is hard to represent a function exactly, and much easier to squeeze it, that is, represent it almost everywhere.

[3] "It is to solve these problems, and not for love of complications..." Lebesque (quoted by Stein and Shakarchi on p. 49).

"In fact, as is so often the case in a new field of mathematics, many of the best consequences were given by the originator." (Jones, the preface).

satisfying $a_0^2+\sum_{k=1}^\infty(a_k^2+b_k^2) < \infty$ and functions satisfying $\int_0^1 |f(x)|^2 \, \mathrm{d}x < \infty$, or rather, their equivalence classes. This is indeed achieved by Lebesgue integration. In contrast, Riemann integration covers only a dense subset in the Hilbert space $L_2[0,1]$ of functions, and the corresponding dense subset in the Hilbert space l_2 of sequences. Lebesgue integral covers the whole space. The space l_2 is complete (it means, every Cauchy sequence converges), therefore $L_2[0,1]$ is also complete.

Lebesgue integral leads to complete spaces of functions, thus opening the door to functional analysis.

In this sense, Lebesgue integral is the ultimate truth.

. . . .

Given a continuous $f : [a,b] \to \mathbb{R}^2$, the length of the curve is defined as $L(a,b) = \sup \sum_{k=1}^n |f(t_k) - f(t_{k-1})|$ over all partitions $a = t_0 < t_1 < \dots < t_n = b$ (and all n). When f is continuously differentiable, $L(a,b) = \int_a^b |f'(t)| \, \mathrm{d}t$. Assume now that ($f$ need not be continuously differentiable, but) $L(a,b) < \infty$, and moreover, $L(c,d) \le M(d-c)$ for a given M and all $[c,d] \subset [a,b]$. Then f is differentiable almost everywhere, and $L(a,b) = \int_a^b |f'(t)| \, \mathrm{d}t$ (Lebesgue integral, of course).[1] However, without the assumption $L(c,d) \le M(d-c)$ it may happen that $L(a,b) > \int_a^b |f'(t)| \, \mathrm{d}t$.

. . . .

The last but not least. Lebesgue integration generalizes readily to measure spaces much more general than $\mathbb{R}^n$ with Lebesgue measure. In particular, probability measures are widely used in infinite-dimensional spaces of functions, of sets etc.;[2] but also on fractal subsets of $\mathbb{R}^d$.

1g Integral or measure? Functions or sets? [3]

It is possible to define the integral first, and then define the measure of a set A as $m(A) = \int \mathbb{1}_A$. But more often one defines the measure first, and then defines the integral.

It is possible to define $\int f$ as the measure under the graph of f (in $\mathbb{R}^{d+1}$). But more often one defines $\int f$ in terms of measures of sets of the form

[1] Riemann integrability may fail, no matter how we extend the integrand to the points of nondifferentiability.

[2] A more advanced example: the configuration space of the scaling limit of percolation.

[3] "One fundamental decision ... is whether to begin with measures or integrals, i.e. whether to start with sets or with functions. Functional analysts have tended to favour the latter approach, while the former is clearly necessary for the development of probability." Capinski, Kopp, p. ix.

$f^{-1}(a,b) = \{x : a < f(x) < b\}$; that is, in terms of the distribution of (the values of) the function f. One option is,

$$\int f = \int_0^\infty m\bigl(f^{-1}(y,\infty)\bigr)\,\mathrm{d}y - \int_0^\infty m\bigl(f^{-1}(-\infty,-y)\bigr)\,\mathrm{d}y\,;$$

these integrals of *monotone* functions are unproblematic. Another option:

$$\int f = \int y\,\mu_f(\mathrm{d}y)$$

where μ_f is the distribution of f, that is, the induced measure, $\mu_f(B) = m\bigl(f^{-1}(B)\bigr)$. The latter integral still needs a definition, but is simpler. More often, one approximates f by functions with discrete (finite or countable) set of values.

All these options are open for Lebesgue integral. For a non-absolute integral we still may write $m(A) = \int \mathbb{1}_A$ (when integrable); but other options fail badly when f is integrable while $|f|$ is not. The area under the graph becomes $\infty - \infty$; and $\mu_f = \mu_g$ does not imply $\int f = \int g$.

For the Riemann integral, the formula $m(A) = \int \mathbb{1}_A$ gives the Jordan measure. It may happen that f is integrable but $\mathbb{1}_{(y,\infty)}(f)$ is not, that is, $\{x : f(x) > y\}$ need not be Jordan measurable. But this is not really an obstacle. The decreasing function $y \mapsto \int \mathbb{1}_{(y,\infty)}(f)$ is still well-defined on $(0,\infty)$ except for its discontinuity points.