A Short Course on the Lebesgue Integral and Measure Theory

Steve Cheng

August 5, 2004

Contents

Preface

This article develops the basics of the Lebesgue integral and measure theory. In terms of content, it adds nothing new to any of the existing textbooks on the subject. But our approach here will be to avoid unduly abstractness and absolute generality, instead focusing on producing proofs of useful results as quickly as possible.

Much of the material here comes from lecture notes from a short real analysis course I had taken, and the rest are well-known results whose proofs I had worked out myself with hints from various sources. I typed this up mainly for my own benefit, but I hope it will be interesting for anyone curious about the Lebesgue integral (or higher mathematics in general).

I will be providing proofs of every theorem. If you are bored reading them, you are invited to do your own proofs. The bibliography outlines the background you need to understand this article.

Copyright matters

1 Motivation for the Lebesgue integral

If you have followed the rigorous definition of the Riemann integral in $\mathbb{R}$ or $\mathbb{R}^n$, you may be wondering why do we need to study yet another integral. After all, why should we even care to integrate nasty functions like:

$$D(x) = \begin{cases} 1\,, & x \in \mathbb{Q} \\ 0\,, & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Rephrased in another way, $D(x)$ is actually the indicator function[1] of the set

$$S = \{x \in \mathbb{Q}\} \subset \mathbb{R}\,,$$

and we want to find its "length". Continuing to rephrase this question, suppose we are taking many real-valued measurements x of a particular physical phenomenon. What is the probability, say, that x is rational? If we assume

[1] For any set S, this is the function χ_S defined by $\chi_S(x) = 1$ if $x \in S$ and $\chi_S(x) = 0$ if $x \notin S$. Math people call this the "characteristic function", while probability people call it the "indicator function" instead.

the measurements are distributed normally with a mean of μ and a standard deviation of σ, then this is given by:

$$\Pr[X \in \mathbb{Q}] = \int_S \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

So wild sets like S are theoretically worth considering, and it does not work to use Riemann integral to evaluate the above probability.

Another limitation to the Riemann integral is with limits. If a sequence of functions f_n is uniformly convergent (on a closed interval, or more generally a compact set $A \subseteq \mathbb{R}^n$), then we can interchange limits for the Riemann integral:

$$\lim_{n\to\infty} \int_A f_n(x)\, dx = \int_A \lim_{n\to\infty} f_n(x)\, dx\,,$$

but the criterion of uniform convergence is often too restrictive, e.g. when integrating Fourier series. On the other hand, it can be proven with the Lebesgue integral that the interchange is valid under weaker conditions (e.g. the functions f_n is bounded above somehow, and they converge *pointwise*).

As an added benefit, some sophisticated results concerning the Riemann integral, such as the Change of Variables Theorem in $\mathbb{R}^n$, are more easily proven using the Lebesgue integral, with its arsenal of limit theorems.

Finally, the Riemann integral does not deal with integration over "infinite bounds" very well. For example, the standard way to compute the probability integral

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx$$

goes like this:

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx\right)^2 &= \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\, dx \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}\, dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy \\
&= \int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy \\
&= \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr\, d\theta \qquad \text{(using polar coordinates)} \\
&= 2\pi \left[-e^{-\frac{1}{2}r^2}\right]_{r=0}^{r=\infty} \\
&= 2\pi\,.
\end{aligned}$$

So

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}\,.$$

The above computation seems easy, and although it can be justified using the Riemann integral alone, it is not entirely trivial, but it is with the Lebesgue integral. (For example, why should $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}$ be the same as $\int_{\mathbb{R}^2}$? Note that in the Riemann theory, the iterated integral and the area integral are proven to be equal only for *bounded* sets of integration.)

You will probably be able to find other sorts of limitations with the Riemann integral.

2 Basic measure theory

The setting of abstract integration is measure theory, which tells us what the areas or volumes of various sets are. Essentially we are *given* some function μ of sets which returns the area or volume — formally called the measure — of the given set. i.e. We assume at the beginning that such a function μ has already been defined for us. The abstract approach of the Lebesgue integral has the obvious advantage that the theory can be applied to many other measures besides volume in $\mathbb{R}^n$.

We begin with the axioms of measure theory.

Definition 2.1. Let X be any non-empty set. A *sigma algebra*[2] of subsets of X is a family $\mathcal{A}$ of subsets of X, with the properties:

1. $\mathcal{A}$ is non-empty.
2. If $E \in \mathcal{A}$, then $X \setminus E \in \mathcal{A}$.
3. If $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of sets in $\mathcal{A}$, then their union is in $\mathcal{A}$. That is, $\mathcal{A}$ is closed under countable unions.

The pair $(X, \mathcal{A})$ is called a *measurable space*, and the sets in $\mathcal{A}$ are called the *measurable sets*.

Notice that the axioms always imply that $X \in \mathcal{A}$. Also, by De Morgan's laws, $\mathcal{A}$ is closed under countable intersections as well as countable union.

Needless to say, we cannot insist that $\mathcal{A}$ is closed under arbitrary unions or intersections, as that would force $\mathcal{A} = 2^X$ if $\mathcal{A}$ contains all the singleton sets. That would be uninteresting. On the other hand, we want closure under countable set operations, rather than just finite ones, as we will want to take countable limits.

Example 2.1. Let X be any (non-empty) set. Then $\mathcal{A} = 2^X$ is a sigma algebra.

Example 2.2. Let X be any (non-empty) set. Then $\mathcal{A} = \{X, \emptyset\}$ is a sigma algebra.

To get non-trivial sigma algebras to work with we need the following, a very unconstructive(!) construction:

If we have a family of sigma algebras on X, then the intersection of all the sigma algebras from this family is also a sigma algebra on X. If all of the sigma

[2] I do not know why it has such a ridiculous name, other than the fact that it is often denoted by the Greek letter.

algebras from the family contains some fixed $\mathcal{G} \subseteq 2^X$, then the intersection of all the sigma algebras from the family, of course, is a sigma algebra containing $\mathcal{G}$.

Now if we are given $\mathcal{G}$, and we take *all* the sigma algebras on X that contain $\mathcal{G}$, and intersect all of them, we get the smallest sigma algebra that contains $\mathcal{G}$.

Definition 2.2. The smallest sigma algebra containing any given $\mathcal{G} \subseteq 2^X$, as constructed above, is denoted $\langle \mathcal{G} \rangle$, and is also called the sigma algebra *generated by* $\mathcal{G}$.

The following is an often-used sigma algebra.

Definition 2.3. If X is a topological space, we can construct the sigma algebra $\langle \mathcal{T} \rangle$, where $\mathcal{T}$ is the set of all open sets. This is called the *Borel sigma algebra* and is denoted $\mathcal{B}(X)$. When topological spaces are involved, we will always take the sigma algebra to be the Borel sigma algebra unless stated otherwise.

$\mathcal{B}(X)$, being generated by the open sets, then contains all open sets, all closed sets, and countable unions and intersections of open sets and closed sets. It seems unlikely, however, that every set in $\mathcal{B}(X)$ is expressible as a countable union and/or intersection of open sets and closed sets, although it is tempting to think that.

By the way, Theorem 9.3 shows the Borel sigma algebra is generally *not all* of 2^X.

Sigma algebras are the domain on which measures are defined.

Definition 2.4. Let $(X, \mathcal{A})$ be a measurable space. A *positive measure* on this space is a function $\mu \colon \mathcal{A} \to [0, \infty]$ such that

1. $\mu(\emptyset) = 0$
2. *Countable additivity*: For any sequence of *mutually disjoint* sets $E_n \in \mathcal{A}$,
$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)\,.$$

The set $(X, \mathcal{A}, \mu)$ will be called a *measure space*. Whenever convenient we will abbreviate this expression, as in "let X be a measure space", etc. Also, in this article, when we say "measure", we will be dealing with *positive* measures only. (There are also theories about signed measures and complex measures.)

Example 2.3. Let X be an arbitrary set, and $\mathcal{A}$ be a sigma algebra on X. Define $\mu \colon \mathcal{A} \to [0, \infty]$ as

$$\mu(A) = \begin{cases} |A|\,, & \text{if } A \text{ is a finite set} \\ \infty\,, & \text{if } A \text{ is an infinite set}\,. \end{cases}$$

This is called the *counting measure*.

We will be able to model the infinite series $\sum_{n=1}^{\infty} a_n$ in Lebesgue integration theory by using $X = \mathbb{N}$ and the counting measure, since integrals are essentially sums of the integrand values weighted by areas or measures.

Example 2.4. $X = \mathbb{R}^n$, and $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$. We can construct the Lebesgue measure λ which assigns to the rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$ its expected n-dimensional volume $(b_1 - a_1) \cdots (b_n - a_n)$. Of course this measure should also assign the correct volumes to the usual geometric figures, as well as for all the other sets in $\mathcal{A}$.

The *existence* of such a measure will be demonstrated later.

Intuitively, defining the volume of the rectangle only should suffice to uniquely also determine the volume of the other sets, since the volume of every set can be approximated by the volume of many small rectangles. Indeed, we will later show this intuition to be true. In fact, you will see that most theorems using Lebesgue measure really depend only on the definition of the volume of the rectangle.

Example 2.5. Any probability measure (as defined by the usual axioms of probability) is actually a measure in our sense. For example,

$$\Pr[Z \in B] = \mu(B) = \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, dt \,,$$ [3]

where the integration, of course, is with respect to the Lebesgue measure on the real line. Other examples include the uniform distribution, the Poisson distribution, and so forth.

Before we begin the prove more theorems, I must mention that we will be operating on the quantity ∞ as if it were a number, even though you may have been told this is "wrong" by some teachers. It is true, of course, that certain algebraic properties of $\mathbb{R}$ would fail to hold with ∞ included (i.e. $\mathbb{R} \cup \{\infty, -\infty\}$ is not a field), but the crucial point in real analysis is that ∞ obeys the usual ordering rules when used in inequalities. The rules we adopt are the following:

$$\begin{aligned} a &\leq \infty \\ \infty + \infty &= \infty \\ a \cdot \infty &= \infty \quad (a \neq 0) \\ 0 \cdot \infty &= 0 \end{aligned}$$

The first three rules are self-explanatory. The last rule may need explaining: when integrating functions, we often want to ignore "isolated" singularities, e.g. at zero for $\int_0^1 dx/\sqrt{x}$. The point 0 is supposed to have "measure zero", so even though the function is ∞ there, the area contribution at that point should still be $0 = 0 \cdot \infty$. Hence the rule. At this point a warning should be issued: the additive cancellation rule *will not work* with ∞. The danger should be sufficiently illustrated in the proofs of the following theorems.

Theorem 2.1. *The following are easy facts about measures:*

1. *It is finitely additive.*

[3] I guess this is my favorite integral. It's got all the important numbers in it — well, except for i.

2. *Monotonicity: If $E, F \in \mathcal{A}$, and $E \subseteq F$, then $\mu(E) \leq \mu(F)$.*

3. *If $E \subseteq F$ has finite measure ($\mu(E) < \infty$), then $\mu(F \setminus E) = \mu(F) - \mu(E)$.*

4. *If A or B has finite measure, then $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.*

Proof. The first fact is obvious. For the second fact, we have

$$\mu(F) = \mu((F \setminus E) \uplus E) = \mu(F \setminus E) + \mu(E)\,,$$

and $\mu(F \setminus E) \geq 0$. For the third fact, just subtract $\mu(E)$ from both sides. (The funny union symbol means that the union is disjoint.) Of course the fact that $\mathcal{A}$ is a sigma algebra is used throughout to know that the new sets also belong to $\mathcal{A}$.)

For the fourth fact, we decompose each of A, B, and $A \cup B$ into disjoint parts, to obtain the following:

$$\begin{aligned}
\mu(A) &= \mu(A \cap B^{\mathrm{c}}) + \mu(A \cap B)\,. \\
\mu(B) &= \mu(B \cap A^{\mathrm{c}}) + \mu(B \cap A)\,. \\
\mu(A \cup B) &= \mu(A \cap B^{\mathrm{c}}) + \mu(A^{\mathrm{c}} \cap B) + \mu(A \cap B)\,.
\end{aligned}$$

Adding the first two equations and then substituting in the third one,

$$\begin{aligned}
\mu(A) + \mu(B) &= \mu(A \cap B^{\mathrm{c}}) + \mu(A \cap B) + \mu(B \cap A^{\mathrm{c}}) + \mu(B \cap A) \\
&= \mu(A \cup B) + \mu(A \cap B)\,.
\end{aligned}$$

Since one of A or B has finite measure, so does $A \cap B \subseteq A, B$, by the second fact, so we may subtract $\mu(A \cap B)$ from both sides. Of course if one of A or B has infinite measure, the resulting equation says nothing interesting. □

The preceding theorem, as well as the next ones, are quite intuitive and you should have no trouble remembering them.

Theorem 2.2. *Let $(X, \mathcal{A}, \mu)$ be measure space, and let $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$ be subsets in $\mathcal{A}$ with union E. (The sets E_n are said to increase to E, and henceforth we will write $\{E_n\} \nearrow E$ for this.) Then*

$$\mu(E) = \mu\left(\bigcup_{n=0}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n)\,.$$

Proof. The sets E_k and E can be written as the disjoint unions

$$\begin{aligned}
E_k &= E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \cdots \cup (E_k \setminus E_{k-1}) \\
E &= E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \cdots\,,
\end{aligned}$$

(and set $E_0 = \emptyset$), so that

$$\mu(E) = \sum_{k=1}^{\infty} \mu(E_k \setminus E_{k-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(E_k \setminus E_{k-1}) = \lim_{n\to\infty} \mu(E_n)\,.$$ □

Theorem 2.3. *For any $E_n \in \mathcal{A}$,*

$$\mu\Big(\bigcup_{n=1}^{\infty} E_n\Big) \leq \sum_{n=1}^{\infty} \mu(E_n)\,.$$

Proof.

$$\begin{aligned}\mu\Big(\bigcup_{n=1}^{\infty} E_n\Big) &= \mu\Big(\bigcup_{n=1}^{\infty} E_n \setminus (E_1 \cup E_2 \cup \cdots \cup E_{n-1})\Big) \\ &= \sum_{n=1}^{\infty} \mu\big(E_n \setminus (E_1 \cup E_2 \cup \cdots \cup E_{n-1})\big) \leq \sum_{n=1}^{\infty} \mu(E_n)\,.\end{aligned}$$ □

Theorem 2.4. *Let $\{E_n\} \searrow E$ (that is, E_n are decreasing and their intersection is E), and $\mu(E_1) < \infty$. Then*

$$\lim_{n\to\infty} \mu(E_n) = \mu(E) = \mu\Big(\bigcap_{n=1}^{\infty} E_n\Big)\,.$$

Proof. We have $\{E_1 \setminus E_n\} \nearrow (E_1 \setminus E)$. So

$$\begin{aligned}\mu(E_1 \setminus E) &= \mu\Big(\bigcup_{n=1}^{\infty} (E_1 \setminus E_n)\Big) = \lim_{n\to\infty} \mu(E_1 \setminus E_n)\,, \\ \mu(E_1) - \mu(E) &= \lim_{n\to\infty} [\mu(E_1) - \mu(E_n)] = \mu(E_1) - \lim_{n\to\infty} \mu(E_n)\,,\end{aligned}$$

and cancel $\mu(E_1)$ on both sides. □

3 Measurable functions

To do integration theory, we of course need functions to integrate. You should not expect that arbitrary functions can be integrated, but only the "measurable" ones. The following definition is not difficult to motivate.

Definition 3.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A map $f\colon X \to Y$ is *measurable* if

$$\text{for all } B \in \mathcal{B},\ \text{the set } f^{-1}(B) = [f \in \mathcal{B}] \text{ is in } \mathcal{A}\,.$$

Example 3.1. A constant map is always measurable, for $f^{-1}(B)$ is either $\emptyset$ or X.

Theorem 3.1. *The composition of two measurable functions is measurable.*

Proof. Immediate from the definition. □

Theorem 3.2. *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces, and suppose $\mathcal{H}$ generates the sigma algebra $\mathcal{B}$: $\langle \mathcal{H} \rangle = \mathcal{B}$. A function $f \colon X \to Y$ is measurable if and only if for every $V \in \mathcal{H}$, $f^{-1}(V)$ is in $\mathcal{A}$.*

Proof. The "only if" part is just the definition of measurability. For the "if" direction, define $\mathcal{G} = \{f^{-1}(V) : V \in \mathcal{H}\}$, and also $\mathcal{C} = \{V \in \mathcal{B} : f^{-1}(V) \in \langle \mathcal{G} \rangle\}$. It is easily checked that $\mathcal{C}$ is a sigma algebra on Y, and it contains $\mathcal{H}$, and hence it is actually equal to $\mathcal{B}$. That is, for every $V \in \mathcal{B}$, $f^{-1}(V)$ is in $\langle \mathcal{G} \rangle \subseteq \langle \mathcal{A} \rangle = \mathcal{A}$. □

Corollary 3.3. *All continuous functions (between topological spaces) are measurable.*

A comment about infinities again. There is a natural topology on $[-\infty, +\infty]$ and $[0, \infty]$ that make them look like closed intervals. Some denote $[-\infty, +\infty]$ by $\overline{\mathbb{R}}$ ("the extended real numbers"). However, for convenience, I will just denote it as plain $\mathbb{R}$. So keep in mind that when we prove our theorems, we have to make sure that they work (or do not work) when infinite quantities are introduced.

Theorem 3.4. *Let $(X, \mathcal{A})$ be a measurable space. A map $f \colon X \to \mathbb{R}$ is measurable if and only if $[f > c] = f^{-1}((c, +\infty]) \in \mathcal{A}$ for all $c \in \mathbb{R}$.*

Proof. Let $\mathcal{B}$ be the set of all open intervals (a, b), along with $\{-\infty\}, \{+\infty\}$. Let $\mathcal{H}$ be the set of all intervals $(c, +\infty]$. (a, b, c are finite.) Evidently $\mathcal{H}$ generates $\mathcal{B}$:

$$
\begin{aligned}
(a, b) &= [-\infty, b) \cap (a, +\infty] \,, \\
[-\infty, b) &= \bigcup_{n=1}^{\infty} [-\infty, b - \frac{1}{n}] = \bigcup_{n=1}^{\infty} \mathbb{R} \setminus (b - \frac{1}{n}, +\infty] \,. \\
\{+\infty\} &= \bigcap_{n=1}^{\infty} (n, +\infty] \,. \\
\{-\infty\} &= \mathbb{R} \setminus \bigcup_{n=1}^{\infty} (-n, +\infty] \,.
\end{aligned}
$$

In turn, $\mathcal{B}$ generates the Borel sigma algebra on $\mathbb{R}$. Applying Theorem 3.2 to the generator $\mathcal{H}$ gives the result. (As $\mathcal{B} \subseteq \langle \mathcal{H} \rangle$ and $\mathcal{H} \subseteq \langle \mathcal{B} \rangle$ together mean $\langle \mathcal{H} \rangle = \langle \mathcal{B} \rangle$.) □

Remark 3.5. We can replace $(c, +\infty]$, in the statement of the theorem, by $[c, +\infty]$, $[-\infty, c]$, etc. and there is no essential difference.

The "countable union with $1/n$" trick used in the proof is widely applicable. It may be of interest to note that the Archimedean property of the real numbers is being used here — the same proof will not work with non-Archimedean ordered fields.

Theorem 3.6. *Let f_n be a sequence of measurable $\mathbb{R}$-valued functions. Then the functions*

$$\sup_n f_n,\ \inf_n f_n,\ \max_n f_n,\ \min_n f_n,\ \limsup_n f_n,\ \liminf_n f_n$$

(the limits are pointwise) are all measurable.

Proof. If $g(x) = \sup_n f_n(x)$, then $[g > c] = \bigcup_n [f_n > c]$, and we apply Theorem 3.4. Similarly, if $g(x) = \inf_n f_n(x)$, then $[g < c] = \bigcup_n [f_n < c]$. The rest can be expressed as in terms of supremums and infimums (over a countable set), so they are measurable also. $\square$

Not surprisingly, we will need to do arithmetic in integration theory, so we better know that

Theorem 3.7. *If $f, g \colon X \to \mathbb{R}$ are measurable, then so are $f+g$, fg, and f/g.*

Proof. Consider the countable union

$$[f + g < c] = \bigcup_{r \in \mathbb{Q}} [f < c - r] \cap [g < r]\,.$$

The set equality is justified as follows: Clearly $f(x) < c-r$ and $g(x) < r$ together imply $f(x) + g(x) < c$. Conversely, if we set $g(x) = t$, then $f(x) < c - t$, and we can increase t slightly to a rational number r such that $f(x) < c - r$, and $g(x) < t < r$. This shows that $f + g$ is measurable (by Theorem 3.4).

Since $[-f < c] = [f > -c]$, we see that $-f$ is measurable.

Therefore the functions

$$\begin{aligned} f^+(x) &= \max\{+f(x), 0\} \quad \text{(positive part of } f) \\ f^-(x) &= \max\{-f(x), 0\} \quad \text{(negative part of } f) \end{aligned}$$

are measurable (from Theorem 3.6 and Example 3.1). Since $f = f^+ - f^-$, f is measurable if f^+ and f^- are measurable separately also.

Since

$$fg = (f^+ - f^-)(g^+ - g^-) = f^+g^+ - f^+g^- - f^-g^+ + f^-g^-\,,$$

to prove that fg is measurable, it suffices to assume that f and g are both non-negative. Then just as with the sum,

$$[fg < c] = \bigcup_{r \in \mathbb{Q}} [f < c/r] \cap [g < r]\,.$$

Finally, for $1/g$,

$$[1/g < c] = \begin{cases} [1/c < g, cg > 0] \cup [1/c > g, cg > 0]\,, & c \neq 0 \\ [g < 0]\,, & c = 0\,. \end{cases} \qquad \square$$

Remark 3.8. You probably have already noticed there may be difficulty in defining what the arithmetic operations mean when the operands are infinite (or when dividing by zero). The usual way to deal with these problems is to simply redefine the functions whenever they are infinite to be some fixed value. In particular, if the function g is obtained by changing the original measurable function f on a *measurable* set A to be a constant c, we have:

$$[g \in B] = \big([g \in B] \cap A\big) \cup \big([g \in B] \cap A^{\mathrm{c}}\big)$$

$$[g \in B] \cap A = \begin{cases} A\,, & c \in B \\ \emptyset\,, & c \notin B \end{cases}$$
$$[g \in B] \cap A^{\mathrm{c}} = [f \in B] \cap A^{\mathrm{c}}\,,$$

so the resultant function g is also measurable. Very conveniently, any sets like $A = [f = +\infty], [f = 0]$ are automatically measurable. Thus the gaps in the previous proof with respect to infinite values can be repaired with this device.

As a final note, one intermediate result from the proof is quite useful and should be formally recognized:

Theorem 3.9. *An $\mathbb{R}$-valued function f is measurable if and only if f^+ and f^- are measurable. Moreover, if f is measurable, so is $|f| = f^+ + f^-$.*

Remark 3.10. Of course the converse to the second statement is not true. You may construct a counterexample to convince yourself of this fact.

4 Definition of the Lebesgue Integral

The idea behind Riemann integration is to try to measure the sums of area of the rectangles "below a graph" of a function and then take some sort of limit. The Lebesgue integral uses a similar approach: we perform integration on the "simple" functions first:

Definition 4.1. A function is *simple* if its range is a finite set.

An $\mathbb{R}$-valued simple function φ always has a representation

$$\varphi = \sum_{k=1}^{n} a_k\,\chi_{E_k}\,,$$

where a_k are the distinct values of φ, and $E_k = \varphi^{-1}(\{a_k\})$. Conversely, any expression of the above form, where a_k need not be distinct, and E_k is not necessarily $\varphi^{-1}(\{a_k\})$, also defines a simple function. For the purposes of integration, however, we will require that E_k be measurable, and that they partition X. It should be mentioned that χ_S is measurable if and only if S is.

Definition 4.2. Let (X, μ) be a measure space. The Lebesgue integral, over X, of a $\mathbb{R}^+$-valued measurable simple function φ is defined as

$$\int_X \varphi \, d\mu = \int_X \sum_{k=1}^n a_k \, \chi_{E_k} \, d\mu = \sum_{k=1}^n a_k \, \mu(E_k) \, .$$

(We restrict φ to being non-negative for now, to avoid mixed $+\infty$, $-\infty$ on the right-hand side.) Needless to say, the quantity on the right represents the sum of the areas below the graph of φ.

It had better be the case that the value of the integral does not depend on the representation of φ. If $\varphi = \sum_i a_i \chi_{A_i} = \sum_j b_j \chi_{B_j}$, where A_i and B_j partition X (so $A_i \cap B_j$ partition X), then

$$\sum_i a_i \, \mu(A_i) = \sum_j \sum_i a_i \, \mu(A_i \cap B_j) = \sum_j \sum_i b_j \, \mu(A_i \cap B_j) = \sum_j b_j \, \mu(B_j) \, .$$

The second equality follows because the value of φ is $a_i = b_j$ on $A_i \cap B_j$, so $a_i = b_j$ whenever $A_i \cap B_j \neq \emptyset$. So fortunately the integral is well-defined.

Using the same algebraic manipulations just now, you can prove that if we have two simple functions $\varphi \leq \psi$, then $\int_X \varphi \, d\mu \leq \int_X \psi \, d\mu$ (monotonicity of the integral).

Theorem 4.1. *The Lebesgue integral (for non-negative simple functions) is linear.*

Proof. Clearly $\int_X c\varphi \, d\mu = c \int_X \varphi \, d\mu$. And if $\varphi = \sum_i a_i \, \chi_{A_i}$, $\psi = \sum_j b_j \, \chi_{B_j}$, we have

$$\begin{aligned}
\int_X \varphi \, d\mu + \int_X \psi \, d\mu &= \sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) \\
&= \sum_i \sum_j a_i \mu(A_i \cap B_j) + \sum_j \sum_i b_j \mu(B_j \cap A_i) \\
&= \sum_i \sum_j (a_i + b_j) \, \mu(A_i \cap B_j) \\
&= \int_X (\varphi + \psi) \, d\mu \, .
\end{aligned}$$

□

Next, we integrate non-simple measurable functions like this:

Definition 4.3. Let $f \colon X \to [0, +\infty]$ be measurable. Consider te set S_f of all measurable simple functions $0 \leq \varphi \leq f$, and define the integral of f over X as

$$\int_X f \, d\mu = \sup_{\varphi \in S_f} \int_X \varphi \, d\mu \, .$$

Intuitively, the simple functions in S_f are supposed to approximate f as close as we like, and we find the integral of f by computing the integrals of these approximations. But logically we need to know that these approximations really do exist. This is the essence of the following theorem.

Theorem 4.2 (Approximation Theorem). *Let $f\colon X \to [0,\infty]$ be measurable. Then there exists a sequence of non-negative functions $\{\varphi_n\} \nearrow f$, meaning φ_n are increasing pointwise and converging pointwise to f. Moreover, if f is bounded, it becomes possible for the φ_n to converge to f uniformly.*

Proof. We prove the second statement first. Let N be any integer $> \sup f$, and set

$$\varphi_n = \sum_{k=1}^{N2^n} \frac{k-1}{2^n} \chi_{E_{n,k}}\,, \quad E_{n,k} = f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right)\,.$$

We have $0 \leq f - \varphi_n < 2^{-n}$ uniformly. The detailed verification is left to the reader.

The construction for the first statement is very similar. Set

$$\varphi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \chi_{E_{n,k}} + \chi_{F_n}\,, \quad F_n = [f \geq n]\,.$$

I leave it to you to check that $0 \leq f(x) - \varphi_n(x) < 2^{-n}$ whenever $n > f(x)$, and $\varphi_n(x) = n$ whenever $f(x) = \infty$. ☐

In case you were worrying about whether this new definition of the integral agrees with the old one in the case of the non-negative simple functions, well, it does. Use monotonicity to prove this.

Definition 4.4. If f is not necessarily non-negative, we define

$$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu\,,$$

provided that the two integrals on the right are not both ∞.

Of course we will want to integrate over subsets of X also. This can be accomplished in two ways. Let A be a measurable subset of X. Either we simply consider integrating over the measure space restricted to subsets of A, or we define

$$\int_A f\,d\mu = \int_X f\,\chi_A\,d\mu\,.$$

If φ is non-negative simple, a simple working out of the two definitions of the integral over A shows that they are equivalent. To prove this for the case of arbitrary measurable functions, we will need the tools of the next section.

Let us note some other basic properties of our integral:

Let f, g be non-negative. Since

$$S_{cf} = c \cdot S_f = \{c\varphi : \varphi \in S_f\}, \quad 0 \le c < \infty .$$

we have (we freely omit the "$d\mu$" and/or the integration limit "X" when they are implied by the context)

$$\int cf = c \int f .$$

This rule about constant multiplication also holds for f and c not necessarily non-negative, as you can easily check, but proving linearity requires the tools of the next section.

Moreover, if $0 \le f \le g$, then $S_f \subseteq S_g$, and therefore $\int f \le \int g$. In particular, if $A \subseteq B$, then $f\chi_A \le f\chi_B$, so

$$\int_A f \le \int_B f .$$

Unsurprisingly $\int f \le \int g$ also holds if f, g are not necessarily non-negative, and that is proven by considering the positive and negative parts of f, g separately. Then since $-|f| \le f \le |f|$, we also obtain

$$-\int |f| \le \int f \le \int |f|, \quad \text{i.e.} \ \left|\int f\right| \le \int |f| .$$

(This last inequality is sometimes called the "generalized triangle inequality", as integrals can be viewed as an advanced form of summing.)

Next, we make one more definition related to integrals.

Definition 4.5. A (μ-)measurable set is said to have (μ-)measurable zero if $\mu(E) = 0$.

Typical examples of a measure-zero set are the singleton points in $\mathbb{R}^n$, and lines and curves in $\mathbb{R}^n, n \ge 2$. By countable additivity, any countable set in $\mathbb{R}^n$ has measure zero also.

A particular property is said to hold almost everywhere if the set of points for which the property fails to hold is a set of measure zero. For example, "a function vanishes almost everywhere".

Clearly, if you integrate anything on a set of measure zero, you get zero.

Assuming that linearity of the integral has been proved, we can demonstrate the following intuitive result.

Theorem 4.3. *A measurable function $f \colon X \to [0, \infty]$ vanishes almost everywhere if and only if $\int_X f = 0$.*

Proof. Let $A = [f = 0]$, and $\mu(A^c) = 0$. Then

$$\int_X f = \int_X f \cdot (\chi_A + \chi_{A^c}) = \int_X f\chi_A + \int_X f\chi_{A^c} = \int_A f + \int_{A^c} f = 0 + 0 .$$

Conversely, if $\int_X f = 0$, consider $[f > 0] = \bigcup_n [f > \frac{1}{n}]$. We have

$$\mu[f > \frac{1}{n}] = \int_{[f>\frac{1}{n}]} 1 = n \int_{[f>\frac{1}{n}]} \frac{1}{n} \leq n \int_X f = 0\,.$$

Hence $\mu[f > 0] = 0$. $\square$

5 Convergence theorems

The following theorems are another feature of the Lebesgue integral that make it so much better than the Riemann definition.

Theorem 5.1 (Monotone Convergence Theorem). *Let (X, μ) be a measure space. Let f_n be non-negative measurable functions increasing pointwise to f. Then*

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

Proof. f is measurable because it is a limit of measurable functions. Since f_n is an increasing sequence of functions bounded by f, their integrals is an increasing sequence of numbers bounded by $\int_X f$; thus the following limit exists:

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

Next we show the inequality in the other direction.

Take any $0 < t < 1$. Given a fixed $\varphi \in S_f$, let $A_n = [f_n - t\varphi \geq 0]$. The A_n are obviously increasing.

If for a particular $x \in X$, we have $\varphi(x) = 0$, then $x \in A_n$, for all n. Otherwise, $\varphi(x) > 0$, so $f(x) \geq \varphi(x) > t\varphi(x)$, and there is going to be some n for which $f_n(x) \geq t\varphi(x)$, i.e. $x \in A_n$. Hence $X = \bigcup_n A_n$.

For all μ-measurable sets E, define a new measure on X by

$$\nu(E) = \int_E t\varphi\,d\mu\,.$$

Then

$$\begin{aligned}
\int_X t\varphi\,d\mu = \nu(X) = \nu\Big(\bigcup_n A_n\Big) &= \lim_{n\to\infty} \nu(A_n) = \lim_{n\to\infty} \int_{A_n} t\varphi\,d\mu \\
&\leq \lim_{n\to\infty} \int_{A_n} f_n\,d\mu\,, \quad \text{since on } A_n \text{ we have } t\varphi \leq f_n \\
&\leq \lim_{n\to\infty} \int_X f_n\,d\mu\,.
\end{aligned}$$

$$\begin{aligned}
t \int_X \varphi\,d\mu &\leq \lim_{n\to\infty} \int_X f_n\,d\mu\,, \quad \text{and take limit } t \to 1\,. \\
\int_X \varphi\,d\mu &\leq \lim_{n\to\infty} \int_X f_n\,d\mu\,, \quad \text{and take sup over } \varphi \in S_f\,.
\end{aligned}$$

$\square$

Using this theorem, we are now in the position to prove linearity of the Lebesgue integral for non-simple functions. Given any two non-negative measurable functions f, g, by the approximation theorem (Theorem 4.2), we know that are non-negative simple functions $\{\varphi_n\} \nearrow f$, and $\{\psi_n\} \nearrow g$. Then $\{\varphi_n + \psi_n\} \nearrow f + g$, and so

$$\int f + g = \lim_{n\to\infty} \int \varphi_n + \psi_n = \lim_{n\to\infty} \int \varphi_n + \int \psi_n = \int f + \int g\,.$$

(The second equality follows because we already know the integral is linear for simple functions. For the first and third equality we apply the Monotone Convergence Theorem.)

And if f, g not necessarily non-negative, then

$$\begin{aligned}
\int f + g &= \int (f^+ - f^-) + (g^+ - g^-) \\
&= \int f^+ + g^+ - (f^- + g^-) \\
&= \int f^+ + g^+ - \int f^- + g^- \\
&= \int f^+ + \int g^+ - (\int f^- + \int g^-) = \int f + \int g\,.
\end{aligned}$$

(Only at the fourth equality we apply what we had just proved for non-negative functions. The third and last equality are just by the definition of the integral.)

Here is another application of the Monotone Convergence Theorem.

Theorem 5.2 (Beppo Levi). *Let $f_n\colon X \to [0,\infty]$ be measurable. Then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n\,.$$

Proof. Let $g_N = \sum_{n=1}^{N} f_n$, and $g = \sum_{n=1}^{\infty} f_n$. The Monotone Convergence Theorem applies to g_N, and:

$$\int g = \int \lim_{N\to\infty} g_N = \lim_{N\to\infty} \int g_N = \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \sum_{n=1}^{\infty} \int f_n\,. \qquad \square$$

There are many more applications like this. We postpone those for now, since you will probably be even more amazed by the next convergence theorem. We first need a lemma.

Lemma 5.3 (Fatou's Lemma). *Let $f_n\colon X \to [0,\infty]$ be measurable. Then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n\,.$$

Proof. Set $g_n = \inf_{k \geq n} f_k$, so that $g_n \leq f_n$, and $\{g_n\} \nearrow \liminf_n f_n$. Then

$$\int \liminf_n f_n = \int \lim_n g_n = \lim_n \int g_n = \liminf_n \int g_n \leq \liminf_n \int f_n \,. \qquad \square$$

Remark 5.4. By adding and subtracting a constant, the hypotheses may be weakened to allow functions that are bounded below by any fixed number, not just non-negative functions. (This lower bound condition cannot be dropped.) The same considerations apply to the Monotone Convergence Theorem.

The following definition is used to formulate a crucial hypothesis of the theorem that is about to follow.

Definition 5.1. A function $f\colon X \to \mathbb{R}$ is called *integrable* if it is measurable and $\int_X |f| < \infty$.

It is immediate that f is integrable if and only if f^+ and f^- are both integrable. It is also helpful to know, that $\int |f| < \infty$ must imply $|f| < \infty$ almost everywhere.

Theorem 5.5 (Dominated Convergence Theorem). *Let (X, μ) be a measure space. Let $f_n\colon X \to \mathbb{R}$ be a sequence of measurable functions converging pointwise to f. Moreover, suppose that there is an* integrable *function g such that $|f_n| \leq g$, for all n. Then f_n and f are also integrable, and*

$$\lim_{n\to\infty} \int_X |f_n - f| \, d\mu = 0 \,.$$

Proof. Obviously f_n and f are integrable. Also, $2g - |f_n - f|$ is measurable and non-negative. By Fatou's lemma,

$$\int \liminf_n (2g - |f_n - f|) \leq \liminf_n \int (2g - |f_n - f|) \,.$$

Since f_n converges to f, the left-hand quantity is just $\int 2g$. The right-hand quantity is:

$$\begin{aligned}
\liminf_n \left(\int 2g - \int |f_n - f| \right) &= \int 2g + \liminf_n \left(- \int |f_n - f| \right) \\
&= \int 2g - \limsup_n \int |f_n - f| \,.
\end{aligned}$$

Since $\int 2g$ is finite, it may be cancelled from both sides. Then we obtain

$$\limsup_n \int |f_n - f| \leq 0 \,, \quad \text{i.e.} \ \lim_{n\to\infty} \int |f_n - f| = 0 \,. \qquad \square$$

Remark 5.6. It obviously suffices to only require that f_n converge to f pointwise almost everywhere, or that $|f_n|$ is bounded above by g almost everywhere. (Of course, if f_n only converges to f almost everywhere, then the theorem would not automatically say that f is measurable.)

Remark 5.7. By the generalized triangle inequality, we conclude from the hypotheses that also

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

which is usually how this theorem is applied.

Remark 5.8. The theorem also holds for continuous limits of functions, not just countable limits. That is, if we have a continuous sequence of functions, say f_t, $0 \leq t < 1$, we can also say

$$\lim_{t\to 1} \int_X |f_t - f| \, d\mu = 0 \,;$$

for given any sequence $\{a_n\}$ convergent to 1, we can apply the theorem to f_{a_n}. Since this can be done for *any* sequence convergent to 1, the above limit is established.

6 Some Results of Integration Theory

This section contains some nice applications proven using the convergence theorems from the last section.

Theorem 6.1 (Generalization of Beppo Levi). *Let X be a measure space, and $f_n \colon X \to \mathbb{R}$ be measurable functions, with $\int \sum |f_n| = \sum \int |f_n| < \infty$. Then*

$$\sum_{n=1}^{\infty} \int f_n = \int \sum_{n=1}^{\infty} f_n \,.$$

Proof. Let $g_N = \sum_{n=1}^{N} f_n$, $g = \limsup_{N\to\infty} g_N$, and $h = \sum_{n=1}^{\infty} |f_n|$. Then $|g_N| \leq h = |h|$. Since $\int |h| < \infty$ by hypothesis, we have $|h| < \infty$ almost everywhere, so $\sum_{n=1}^{\infty} f_n$ is absolutely convergent almost everywhere. That is, g_N converges pointwise to g almost everywhere.

By the Dominated Convergence Theorem, $\lim\limits_{N\to\infty} \int g_N = \int g$, whence

$$\sum_{n=1}^{\infty} \int f_n = \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n = \int \sum_{n=1}^{\infty} f_n \,. \qquad \square$$

Example 6.1. Here's a perhaps unexpected application. Suppose we have a countable set of real numbers $a_{n,m}$, $n, m \in \mathbb{N}$. Let μ be the counting measure on $\mathbb{N}$. Then $\int_{m\in\mathbb{N}} a_{n,m} \, d\mu = \sum_{m=1}^{\infty} a_{n,m}$. Moreover, Theorem 6.1 says that we can sum either along n first or m first and get the same results ($\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{n,m} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{n,m}$) if the double sum is absolutely convergent. Of course, this fact can also be proven in an entirely elementary way.

Theorem 6.2. *Let $g \colon X \to [0, \infty]$ be measurable in the measure space $(X, \mathcal{A}, \mu)$. Let*

$$\nu(E) = \int_E g \, d\mu \,, \quad E \in \mathcal{A} \,.$$

Then ν is a measure on $(X, \mathcal{A})$, and for any measurable function f on X,

$$\int_X f \, d\nu = \int_X fg \, d\mu \,,$$

often written as $d\nu = g \, d\mu$.

Proof. We prove ν is a measure. $\nu(\emptyset) = 0$ is trivial. For countable additivity, let $\{E_n\}$ be measurable with union E, so that $\chi_E = \sum_{n=1}^{\infty} \chi_{E_n}$, and

$$\begin{aligned} \nu(E) &= \int_E g \, d\mu = \int_X g\chi_E \, d\mu \\ &= \int_X \sum_{n=1}^{\infty} g\chi_{E_n} \, d\mu = \sum_{n=1}^{\infty} \int_X g\chi_{E_n} \, d\mu = \sum_{n=1}^{\infty} \nu(E_n) \,. \end{aligned}$$

Next, if $f = \chi_E$ for some $E \in \mathcal{A}$, then

$$\int_X f \, d\nu = \int_X \chi_E \, d\nu = \nu(E) = \int_X \chi_E g \, d\mu = \int_X fg \, d\mu \,.$$

By linearity, we see that $\int f \, d\nu = \int fg \, d\mu$ whenever f is non-negative simple. For general non-negative f, we use a sequence of simple approximations $\{\varphi_n\} \nearrow f$, so $\{\varphi_n g\} \nearrow fg$. Then by the monotone convergence,

$$\int_X f \, d\nu = \lim_{n\to\infty} \int_X \varphi_n \, d\nu = \lim_{n\to\infty} \int_X \varphi_n g \, d\mu = \int_X \lim_{n\to\infty} \varphi_n g \, d\mu = \int_X fg \, d\mu \,.$$

Finally, for f not necessarily non-negative, we apply the above to its positive and negative parts, and use linearity. $\square$

The procedure of proving some fact about integrals by first reducing to the case of simple functions and non-negative functions is used quite often. (It will get quite monotonous if we had to detail the procedure every time we use it, so we won't anymore if the circumstances permit.)

Also, we should note that if f is only measurable but not integrable, then the integrals of f^+ or f^- might be infinite. If both are infinite, the integral of f is not defined, although the equation of the theorem might still be interpreted as saying that the left-hand and right-hand sides are undefined at the same time. For this reason, and for the sake of the clarity of our exposition, we will not bother to modify the hypotheses of the theorem to state that f must be integrable.

Problem cases like this also occur for some of the other theorems we present, and there I will also not make too much of a fuss about these problems, trusting that you understand what happens when certain integrals are undefined.

Theorem 6.3 (Change of variables). *Let X, Y be measure spaces, and $g \colon X \to Y$, $f \colon Y \to \mathbb{R}$ be measurable. Then*

$$\int_X (f \circ g) \, d\mu = \int_Y f \, d\nu \,,$$

where $\nu(B) = \mu(g^{-1}(B))$ is a measure defined for all measurable $B \subseteq Y$.

Proof. First suppose $f = \chi_B$. Let $A = g^{-1}(B) \subseteq X$. Then $f \circ g = \chi_A$, and we have

$$\int_Y f \, d\nu = \int_Y \chi_B \, d\nu = \nu(B) = \mu(g^{-1}(B)) = \mu(A) = \int_X (f \circ g) \, d\mu \, .$$

Since both sides of the equation are linear in f, the equation holds whenever f is simple. Applying the "standard procedure" mentioned above, the equation is then proved for all measurable f. □

Remark 6.4. The change of variables theorem can also be applied "in reverse". Suppose we want to compute $\int_Y f \, d\nu$, where ν is already given to us. Further assume that g is bijective and its inverse is measurable. Then we can define $\mu(A) = \nu(g(A))$, and it follows that $\int_Y f \, d\nu = \int_X (f \circ g) \, d\mu$.

Our theorem (especially when stated in the reverse form) is clearly related to the usual "change of variables" theorem in calculus. If $g \colon X \to Y$ is a bijection between open subsets of $\mathbb{R}^n$, and both it and its inverse are continuously differentiable (i.e. g is a *diffeomorphism*), and $\nu = \lambda$ is the Lebesgue measure in $\mathbb{R}^n$, then (as we shall prove rigorously in Lemma 12.1),

$$\mu(A) = \lambda(g(A)) = \int_A |\det \mathrm{D}g| \, d\lambda \, .$$

Appealing to Theorems 6.2 and 6.3, we obtain:

Theorem 6.5 (Differential change of variables in $\mathbb{R}^n$). *Let $g \colon X \to Y$ be a diffeomorphism of open sets in $\mathbb{R}^n$. If $A \subseteq X$ is measurable, and $f \colon Y \to \mathbb{R}$ is measurable, then*

$$\int_{g(A)} f \, d\lambda = \int_A (f \circ g) \, d\mu = \int_A (f \circ g) \cdot |\det \mathrm{D}g| \, d\lambda \, .$$

The next two theorems are the Lebesgue versions of well-known results about the Riemann integral.

Theorem 6.6 (First Fundamental Theorem of Calculus). *Let $I \subseteq \mathbb{R}$ be an interval, and $f \colon I \to R$ be integrable (with Lebesgue measure in $\mathbb{R}$). Then the function*

$$F(x) = \int_a^x f(t) \, dt$$

is continuous. Furthermore, if f is continuous at x, then $F'(x) = f(x)$.

Proof. To prove continuity, we compute:

$$F(x+h) - F(x) = \int_x^{x+h} f(t) \, dt = \int_I f(t) \cdot \chi_{[x,x+h]}(t) \, dt \, .$$

(Naturally, when $h < 0$, $\chi_{[x,x+h]}$ should be interpreted as $-\chi_{[x+h,x]}$ here.) Since $|f \cdot \chi_{[x,x+h]}| \leq |f|$, by the Dominated Convergence Theorem (along with remark 5.8),

$$\begin{aligned}
\lim_{h\to 0} F(x+h) - F(x) &= \lim_{h\to 0} \int_I f(t) \cdot \chi_{[x,x+h]}(t)\, dt \\
&= \int_I \lim_{h\to 0} f(t) \cdot \chi_{[x,x+h]}(t)\, dt \\
&= \int_I f(t) \cdot \chi_{\{x\}}(t)\, dt = 0\,.
\end{aligned}$$

The proof of differentiability is the same as for the Riemann integral:

$$\begin{aligned}
\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \left| \frac{\int_x^{x+h} (f(t) - f(x))\, dt}{h} \right| \\
&\leq \frac{\int_{[x,x+h]} |f(t) - f(x)|\, dt}{|h|} \\
&\leq \frac{\sup_{t\in[x,x+h]} |f(t) - f(x)| \cdot |h|}{|h|}\,,
\end{aligned}$$

which goes to zero as h does. $\square$

The First Fundamental Theorem was easy, but the Second Fundamental Theorem (which states that $\int_a^b f' = f(b) - f(a)$) is not entirely trivial. The difficulty is that we should not assume as hypotheses that f' is continuous, or even that it is Lebesgue-integrable. It turns out that a theorem without such strong hypotheses is possible; we will not reproduce its proof here, but just settle for a weaker version:

Theorem 6.7 (Second Fundamental Theorem of Calculus). *Suppose $f\colon [a,b] \to \mathbb{R}$ is measurable and* bounded *above and below. If $f = g'$ for some g, then*

$$\int_a^b f(x)\, dx = g(b) - g(a)\,.$$

Proof. We first note that $\left|\frac{g(x+h)-g(x)}{h}\right|$ can be bounded by a constant using the Mean Value Theorem, and a constant is obviously integrable on a finite interval.

Then

$$\begin{aligned}
\int_a^b f(x)\,dx &= \int_a^b \lim_{h\to 0} \frac{g(x+h)-g(x)}{h}\,dx \\
&= \lim_{h\to 0} \int_a^b \frac{g(x+h)-g(x)}{h}\,dx \\
&= \lim_{h\to 0} \frac{1}{h}\Big[\int_{a+h}^{b+h} g(x)\,dx - \int_a^b g(x)\,dx\Big] \\
&= \lim_{h\to 0} \frac{1}{h}\Big[\int_{b}^{b+h} g(x)\,dx - \int_a^{a+h} g(x)\,dx\Big] \\
&= g(b) - g(a)\,.
\end{aligned}$$

(The last equality follows from the First Fundamental Theorem and that g must be continuous at a and b if it is differentiable there.) □

Remark 6.8. If $g'(x)$ exists, then it can also be computed as the countable limit $\lim_{n\to\infty} n(g(x+1/n)-g(x))$, thus showing that g' is measurable. Thus we can drop the hypothesis that f is measurable in Theorem 6.7.

Remark 6.9. You might have noticed that I cheated a bit in the proof, in assuming the integral of a function is invariant under horizontal translations. But of course, this can be proven readily using the fact that Lebesgue measure is translation-invariant.

The following theorems are often not found in calculus texts even though they are quite important for applications.

Theorem 6.10 (Continous dependence on integral parameter). *Let (X,μ) be a measure space, T be any metric space (e.g. $\mathbb{R}^n$), and $f\colon X\times T\to\mathbb{R}$, with $f(\cdot,t)$ being measurable for each $t\in T$. Consider the function*

$$F(t) = \int_{x\in X} f(x,t)\,.$$

Then we have F continuous at $t_0\in T$ if the following conditions are met:

1. *For each $x\in X$, $f(x,\cdot)$ is continuous at $t_0\in I$.*
2. *There is an integrable function g such that $|f(x,t)|\le g(x)$ for all $t\in T$.*

Proof.

$$\lim_{t\to t_0}\int_{x\in X} f(x,t) = \int_{x\in X}\lim_{t\to t_0} f(x,t) = \int_{x\in X} f(x,t_0)\,.$$ □

Theorem 6.11 (Differentiation under the integral sign). *Using the same notation as Theorem 6.10, with T being an open real interval, we have*

$$F'(t) = \frac{d}{dt}\int_{x\in X} f(x,t) = \int_{x\in X}\frac{\partial}{\partial t} f(x,t)$$

if the following conditions are satisfied:

1. *For each $x \in X$, $\frac{\partial}{\partial t} f(x,t)$ exists.*

2. *There is an integrable function g such that $\left|\frac{\partial}{\partial t} f(x,t)\right| \leq g(x)$ for all $t \in T$.*

Proof. This theorem is often proven by using iterated integrals and switching the order of integration, but that method is theoretically troublesome because it requires more stringent hypotheses. It is easier, and better, to prove it directly from the definition of the derivative.

The straightforward computation yields:

$$\begin{aligned}
\lim_{h \to 0} \frac{F(t+h) - F(t)}{h} &= \lim_{h \to 0} \int_{x \in X} \frac{f(x,t+h) - f(x,t)}{h} \\
&= \int_{x \in X} \lim_{h \to 0} \frac{f(x,t+h) - f(x,t)}{h} = \int_{x \in X} \frac{\partial}{\partial t} f(x,t)
\end{aligned}$$

(noting that $\left|\frac{f(x,t+h)-f(x,t)}{h}\right|$ is bounded by $g(x)$). □

Remark 6.12. It is easy to see that we may generalize Theorem 6.11 to T being any open set in $\mathbb{R}^n$, taking partial derivatives. I won't write it out in full because the notation is somewhat complicated.

Example 6.2. Check that the function $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt, x > 0$ is continuous, and differentiable with the obvious formula for the derivative.

7 $\mathbf{L}^p$ spaces

The contents in this section do not have applications in this article, but they are so well known that it would not do justice to omit them.

Definition 7.1. Let X be a measure space, and let $p \in [1, \infty)$. The space $\mathbf{L}^p$ consists of all measurable functions $f \colon X \to \mathbb{R}$ such that

$$\int |f|^p < \infty \,.$$

Definition 7.2. For each $f \in \mathbf{L}^p$, define

$$\|f\|_p = \left(\int |f|^p\right)^{1/p} .$$

(If $f \notin \mathbf{L}^p$, this quantity is of course defined as ∞.)

As suggested by the notation, $\|\cdot\|_p$ is a real norm on the vector space $\mathbf{L}^p$, provided that we declare two functions to be equivalent if they differ only on a set of measure zero (so that $\|f\|_p = 0$ if and only if $f = 0$ as equivalence classes). Only the verification of the triangle inequality presents any difficulties — this will be solved by the theorems below.

Definition 7.3. Two numbers $p, q \in (1, \infty)$ are called *conjugate exponents* when $\frac{1}{p} + \frac{1}{q} = 1$.

Theorem 7.1 (Hölder's inequality). *For $\mathbb{R}$-valued measurable functions f and g,*

$$\left|\int fg\right| \leq \int |f||g| \leq \|f\|_p \|g\|_q \,.$$

Proof. The first inequality is trivial. For the second inequality, since it only involves absolute values, for the rest of the proof we may assume that f, g are non-negative.

If $\|f\|_p = 0$, then $|f|^p = 0$ almost everywhere, and so $f = 0$ and $fg = 0$ almost everywhere too. Thus the inequality is valid in this case. (Similarly when $\|g\|_q = 0$.)

If $\|f\|_p$ or $\|g\|_q$ is infinite, the inequality is trivial.

So we now assume these two quantities are both finite and non-zero. Define $F = f/\|f\|_p$, $G = g/\|g\|_q$, so that $\|F\|_p = \|G\|_q = 1$. We must then show that $\int FG \leq 1$.

To do this, we employ the fact that log is concave:

$$\frac{1}{p}\log s + \frac{1}{q}\log t \leq \log\left(\frac{s}{p} + \frac{t}{q}\right), \quad 0 \leq s, t \leq \infty$$

or,

$$s^{\frac{1}{p}}\, t^{\frac{1}{q}} \leq \frac{s}{p} + \frac{t}{q}\,.$$

Substitute $s = F^p$, $t = G^q$, and integrate both sides:

$$\int FG \leq \frac{1}{p}\int F^p + \frac{1}{q}\int G^q = \frac{1}{p}\|F\|_p^p + \frac{1}{q}\|G\|_q^q = \frac{1}{p} + \frac{1}{q} = 1. \qquad \square$$

You may have seen a special case of this theorem, for $p = q = 2$, as the Cauchy-Schwarz inequality.

Theorem 7.2 (Minkowski's inequality). *For $\mathbb{R}$-valued measurable functions f and g,*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p\,.$$

Proof. The inequality is trivial when $p = 1$ or when $\|f + g\|_p = 0$. Also, since $\|f + g\|_p \leq \||f| + |g|\|_p$, it again suffices to consider only the case when f, g are non-negative.

Next, we employ the convexity of $t \mapsto t^p$, $p > 1$,

$$\left(\frac{s+t}{2}\right)^p \leq \frac{s^p + t^p}{2}, \quad 0 \leq s, t \leq \infty\,.$$

When we substitute $s = f$, $t = g$, we get $(f + g)^p \leq 2^{p-1}(f^p + g^p)$. This inequality shows that if $\|f + g\|_p$ is infinite, then one of $\|f\|_p$ or $\|g\|_p$ is also infinite, so Minkowski's inequality holds true in that case.

We may now assume $\|f + g\|_p$ is finite. We write:

$$\int (f + g)^p = \int f(f + g)^{p-1} + \int g(f + g)^{p-1}\,.$$

By Hölder's inequality, and noting that $(p-1)q = p$ for conjugate exponents,

$$\int f(f+g)^{p-1} \le \|f\|_p \left\|(f+g)^{p-1}\right\|_q = \|f\|_p \Big(\int |f+g|^p\Big)^{1/q} = \|f\|_p \|f+g\|_p^{p/q}\,.$$

A similar inequality holds for $\int g(f+g)^{p-1}$. Putting these together:

$$\|f+g\|_p^p \le (\|f\|_p + \|g\|_p)\|f+g\|_p^{p/q}\,.$$

Dividing by $\|f+g\|_p^{p/q}$ yields the desired result. $\square$

Definition 7.4. A measure space (X,μ) has *finite measure* if $\mu(X)$ is finite.

Theorem 7.3. *Let (X,μ) have finite measure. Then whenever $1 \le r < p < \infty$, $\mathbf{L}^p \subseteq \mathbf{L}^r$. Moreover, the inclusion map from $\mathbf{L}^p$ to $\mathbf{L}^r$ is continuous.*

Proof. If $f \in \mathbf{L}^p$, apply the Hölder inequality with conjugate exponents $\frac{p}{r}$ and $s = \frac{p}{p-r}$:

$$\|f\|_r^r = \int |f|^r \le \Big(\int |f|^{r\cdot\frac{p}{r}}\Big)^{r/p} \Big(\int 1^s\Big)^{1/s} = \|f\|_p^r\, \mu(X)^{1/s}\,,$$

and so

$$\|f\|_r \le \|f\|_p\, \mu(X)^{1/rs} = \|f\|_p\, \mu(X)^{\frac{1}{r}-\frac{1}{p}} < \infty\,.$$

To show continuity of the inclusion map, replace f with $f-g$ above where $\|f-g\|_p < \varepsilon$. $\square$

Example 7.1. $\int_0^1 x^{-\frac{1}{2}}\,dx = 2 < \infty$, so automatically $\int_0^1 x^{-\frac{1}{4}}\,dx < \infty$. On the other hand, the condition that $\mu(X) < \infty$ is indeed necessary: $\int_1^\infty x^{-2}\,dx < \infty$, but $\int_1^\infty x^{-1}\,dx = \infty$.

Theorem 7.4. *Let $f_n\colon X \to \mathbb{R}$ be measurable functions converging (almost everywhere) pointwise to f, and $|f_n| \le g$ for some $g \in \mathbf{L}^p$. Then $f, f_n \in \mathbf{L}^p$, and f_n converges to f in the $\mathbf{L}^p$ norm, meaning:*

$$\lim_{n\to\infty} \Big(\int |f_n - f|^p\Big)^{1/p} = \lim_{n\to\infty} \|f_n - f\|_p = 0.$$

Proof. $|f_n - f|^p$ converges to 0 and $|f_n - f|^p \le (2g)^p \in \mathbf{L}^1$. Apply the Dominated Convergence Theorem on these functions. $\square$

One wonders whether the converse is true: if f_n converges to f in the $\mathbf{L}^p$ norm, do the functions f_n themselves converge pointwise to f? The answer is no (the counterexamples are not difficult), but we do have the following.

Theorem 7.5. *Let $\{f_n\}$ be a Cauchy sequence in $\mathbf{L}^p$. Then it has a subsequence converging pointwise almost everywhere.*

Corollary 7.6. *If* $\lim_{n\to\infty} \|f_n - f\|_p = 0$ *then there is a subsequence* $\{f_{n(k)}\}$ *converging to* f *pointwise almost everywhere.*

Corollary 7.6 is used in the following result.

Theorem 7.7. $\mathbf{L}^p$ *is a complete metric space. (This means every Cauchy sequence in* $\mathbf{L}^p$ *converges.)*

The proofs of these theorems are collected in the next section.

It is also possible to define "$\mathbf{L}^\infty$":

Definition 7.5. Let X be a measure space, and $f\colon X \to \mathbb{R}$ be measurable. A number $M \in [0, \infty]$ is an almost-everywhere upper bound for $|f|$ if $|f| \leq M$ almost everywhere. The infimum of all almost-everywhere upper bounds for $|f|$ is denoted by $\|f\|_\infty$.

Definition 7.6. $\mathbf{L}^\infty$ is the set of all measurable functions f with $\|f\|_\infty < \infty$. Its norm is given by $\|f\|_\infty$.

The use of the subscript "∞" is explained by the following theorem, whose proof is left to the reader:

Theorem 7.8. *If* $\mu(X) < \infty$*, then* $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$.

Remark 7.9. Observe that there is a similar thing for vectors $\vec{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$:

$$\lim_{p\to\infty} \|\vec{a}\|_p = \lim_{p\to\infty} \left(|a_1|^p + \cdots + |a_n|^p\right)^{1/p} = \max(|a_1|, \ldots, |a_n|) = \|\vec{a}\|_\infty \,.$$

This remark may serve as a hint.

If we (naturally) define the conjugate exponent of $p = 1$ to be $q = \infty$, the Hölder inequality remains valid, since $|fg| \leq |f|\|g\|_\infty$ almost everywhere. Integrating both sides gives $\|fg\|_1 \leq \|f\|_1\|g\|_\infty$.

Before closing, we mention two more results. From Theorem 7.4, it is clear that

Theorem 7.10. *Let* $f\colon X \to \mathbb{R} \in \mathbf{L}^p(X)$*,* $1 \leq p < \infty$*. Then for any* $\varepsilon > 0$*, there exist simple functions* $\varphi : X \to \mathbb{R}$ *such that*

$$\|\varphi - f\|_p = \left(\int_{\mathbb{R}^n} |\varphi - f|^p \, d\lambda\right)^{1/p} < \varepsilon \,.$$

(This is also true for $p = \infty$.) This may be summarized by saying that the set of all simple functions is dense in $\mathbf{L}^p$ (in the topological sense).

When $X = \mathbb{R}^n$ with Lebesgue measure λ, we have another result. The set of infinitely differentiable functions on $\mathbb{R}^n$ with compact support[4], denoted $\mathbf{C}_0^\infty$,

[4] The support of a function $\psi\colon X \to \mathbb{R}$ is the closure of the set $\{x \in X : \psi(x) \neq 0\}$. "$\psi$ has compact support" means that the support of ψ is compact (when $X = \mathbb{R}^n$, same as closed and bounded).

is also dense in $\mathbf{L}^p(\mathbb{R}^n, \lambda)$, $1 \le p < \infty$. (This fact will be fully proven in Section 14.)

These two facts are typically used as tools to prove other theorems about functions $f \in \mathbf{L}^p$. One first proves that a certain theorem holds for all simple φ (or $\varphi \in \mathbf{C}_0^\infty$), and then prove that the same result holds for arbitrary $f \in \mathbf{L}^p$ by approximating such f by φ. An example of this procedure follows.

Theorem 7.11 (Riemann-Lebesgue Lemma). *For all* $f \in \mathbf{L}^1(\mathbb{R})$,

$$\lim_{|\omega|\to\infty} \int_{\mathbb{R}} f(x)\sin(\omega x)dx = 0\,.$$

Proof. For convenience, assume $\omega > 0$.

Suppose first that $f = \psi \in \mathbf{C}_0^\infty$. Since ψ has compact support, integrating $\psi(x)\sin(\omega x)$ over $\mathbb{R}$ is the same as integrating over some compact interval $[a, b]$ containing the support of ψ. And since every function involved is infinitely differentiable, we may use integration by parts ($u = \psi(x)$, $dv = \sin(\omega x)\,dx$):

$$\int_a^b \psi(x)\sin(\omega x)\,dx = -\psi(x)\frac{\cos(\omega x)}{\omega}\bigg|_a^b + \int_a^b \frac{\cos(\omega x)}{\omega}\psi'(x)\,dx\,.$$

As $|\psi|$ and $|\psi'|$ are continuous on $[a, b]$, they are bounded by constants M and M' respectively. We have,

$$\begin{aligned}\left|\int_a^b \psi(x)\sin(\omega x)\,dx\right| &\le \frac{1}{\omega}|\psi(b)\cos(\omega b) - \psi(a)\cos(\omega a)| + \frac{1}{\omega}\int_a^b |\cos(\omega x)\psi'(x)|\,dx \\ &\le \frac{2M}{\omega} + \frac{(b-a)M'}{\omega} \to 0\,, \quad \omega \to \infty\,.\end{aligned}$$

Thus we have proven the result when $f = \psi \in \mathbf{C}_0^\infty$. Now suppose f is arbitrary. By the denseness of $\mathbf{C}_0^\infty$, for every $\varepsilon > 0$ we can find $\psi \in \mathbf{C}_0^\infty$ such that $\|f - \psi\|_1 < \varepsilon$. Then

$$\begin{aligned}\left|\int_{\mathbb{R}} f(x)\sin(\omega x)\,dx - \int_{\mathbb{R}} \psi(x)\sin(\omega x)\,dx\right| &\le \int_{\mathbb{R}} \Big|\big(f(x) - \psi(x)\big)\sin(\omega x)\Big|\,dx \\ &\le \int_{\mathbb{R}} |f(x) - \psi(x)|\,dx \\ &< \varepsilon\,,\end{aligned}$$

or,

$$\left|\int_{\mathbb{R}} f(x)\sin(\omega x)\,dx\right| < \varepsilon + \left|\int_{\mathbb{R}} \psi(x)\sin(\omega x)\,dx\right|\,.$$

We take $\limsup_{\omega\to\infty}$ of both sides:

$$\limsup_{\omega\to\infty}\left|\int_{\mathbb{R}} f(x)\sin(\omega x)\,dx\right| \le \varepsilon + 0\,.$$

But $\varepsilon > 0$ is arbitrary, so we must have:

$$\lim_{\omega\to\infty}\left|\int_{\mathbb{R}} f(x)\sin(\omega x)\,dx\right| = \limsup_{\omega\to\infty}\left|\int_{\mathbb{R}} f(x)\sin(\omega x)\,dx\right| = 0\,. \qquad \square$$

8 Construction of Lebesgue Measure

We now come to actually construct Lebesgue measure, as promised. The idea is to extend an existing measure μ which has been only partially defined, to an "outer measure" μ^*. The extension is remarkably simple and intuitive:

$$\mu^*(E) = \inf_{\substack{A_1, A_2, \ldots \in \mathcal{A} \\ E \subseteq \bigcup_n A_n}} \sum_n \mu(A_n)\,, \quad \text{for all } E \subseteq X\,.$$

(One interesting point to note: the definition of μ^* "works" with pretty much any non-negative function μ, provided that μ is defined on some set $\mathcal{A} \subseteq 2^X$ with $\emptyset, X \in \mathcal{A}$, and $\mu(\emptyset) = 0$. In fact, for the first few proofs, these are the only formal properties of μ that we need. Of course later we will need stronger conditions on μ, such as additivity.)

Lemma 8.1. *μ^* has the following properties:*

1. *$\mu^*(\emptyset) = 0$.*
2. *It is monotone: $\mu^*(E) \leq \mu^*(F)$ when $E \subseteq F \subseteq X$.*
3. *$\mu^*(A) \leq \mu(A)$ for all $A \in \mathcal{A}$.*
4. *μ^* is countably subadditive: if $E_1, E_2, \ldots \subseteq X$, then $\mu^*(\bigcup_n E_n) \leq \sum_n \mu^*(E_n)$.*

Proof. The first three properties are obvious. For the fourth, first observe that if any of the $\mu^*(E_n)$ is infinite, there is nothing to prove. Otherwise, let $\varepsilon > 0$. For each E_n, by the definition of μ^*, there are sets $\{A_{n,m}\}_m \in \mathcal{A}$ covering E_n, with $\sum_m \mu(A_{n,m}) \leq \mu^*(E_n) + \varepsilon/2^n$. All of the $A_{n,m}$ together cover $\bigcup_n E_n$, and $\sum_{n,m} \mu^*(A_{n,m}) \leq \sum_n \mu^*(E_n) + \varepsilon$. Since ε was arbitrary, we have $\sum_{n,m} \mu^*(A_{n,m}) \leq \sum_n \mu^*(E_n)$. $\square$

There is of course no guarantee that μ^* satisfies all the properties of a proper measure on 2^X. (It often does not.) Instead we will claim that μ^* is a proper measure on the following subcollection of 2^X:

$$\begin{aligned} \mathcal{M} &= \{B \in 2^X \mid \mu^*(B \cap E) + \mu^*(B^{\mathrm{c}} \cap E) = \mu^*(E) \quad \text{for all } E \subseteq X\} \\ &= \{B \in 2^X \mid \mu^*(B \cap E) + \mu^*(B^{\mathrm{c}} \cap E) \leq \mu^*(E) \quad \text{for all } E \subseteq X\}\,. \end{aligned}$$

(The two subcollections are the same since by subadditivity we always have $\mu^*(B \cap E) + \mu^*(B^{\mathrm{c}} \cap E) \geq \mu^*(E)$.)

Lemma 8.2. *$\mathcal{M}$ is a sigma algebra.*

Proof. It is immediate from the definition that $\mathcal{M}$ is closed under taking complements, and that $\emptyset \in \mathcal{M}$. We first show $\mathcal{M}$ is closed under finite intersection (and hence under finite union). Let $B, C \in \mathcal{M}$.

$$
\begin{aligned}
&\mu^*\big((B\cap C)\cap E\big) + \mu^*\big((B\cap C)^{\mathrm{c}}\cap E\big)\\
=&\mu^*(B\cap C\cap E) + \mu^*\big((B^{\mathrm{c}}\cap C\cap E)\cup(B\cap C^{\mathrm{c}}\cap E)\cup(B^{\mathrm{c}}\cap C^{\mathrm{c}}\cap E)\big)\\
\le&\mu^*(B\cap C\cap E) + \mu^*(B^{\mathrm{c}}\cap C\cap E) + \mu^*(B\cap C^{\mathrm{c}}\cap E) + \mu^*(B^{\mathrm{c}}\cap C^{\mathrm{c}}\cap E)\\
=&\mu^*(C\cap E) + \mu^*(C^{\mathrm{c}}\cap E)\,, \quad \text{by definition of } B\in\mathcal{M}\\
=&\mu^*(E)\,, \quad \text{by definition of } C\in\mathcal{M}\,.
\end{aligned}
$$

Thus $B \cap C \in \mathcal{M}$.

We now have to show that if $B_1, B_2, \ldots \in \mathcal{M}$, $\bigcup_n B_n \in \mathcal{M}$. We may assume that B_n are disjoint, for otherwise we just consider $B_n' = B_n \setminus (B_1 \cup \cdots \cup B_{n-1})$, which are in $\mathcal{M}$ by the previous paragraph.

We shall need to know that, for all $N \geq 1$, and all $E \subseteq X$,

$$
\mu^*\left(\bigcup_{n=1}^N B_n \cap E\right) = \sum_{n=1}^N \mu^*(B_n\cap E)\,.
$$

The proof will be by induction on N. The statement is trivial for $N = 1$. For the induction step, let $D_N = \bigcup_{n=1}^N B_n$ which are increasing and are all in $\mathcal{M}$. Then

$$
\begin{aligned}
\mu^*(D_{N+1}\cap E) &= \mu^*\big(D_N\cap(D_{N+1}\cap E)\big) + \mu^*\big(D_N^{\mathrm{c}}\cap(D_{N+1}\cap E)\big)\\
&= \mu^*(D_N\cap E) + \mu^*(B_{N+1}\cap E)\\
&= \sum_{n=1}^{N+1}\mu^*(B_n\cap E) \quad \text{(induction hypothesis).}
\end{aligned}
$$

Using the fact just proven, we now have:

$$
\begin{aligned}
\mu^*(E) &= \mu^*(D_N\cap E) + \mu^*(D_N^{\mathrm{c}}\cap E)\\
&= \sum_{n=1}^N \mu^*(B_n\cap E) + \mu^*(D_N^{\mathrm{c}}\cap E)\\
&\geq \sum_{n=1}^N \mu^*(B_n\cap E) + \mu^*\left(\left(\bigcup_{n=1}^\infty B_n\right)^{\mathrm{c}}\cap E\right) \quad \text{(monotonicity).}
\end{aligned}
$$

Taking $N \to \infty$, we obtain:

$$
\begin{aligned}
\mu^*(E) &\geq \sum_{n=1}^\infty \mu^*(B_n\cap E) + \mu^*\left(\left(\bigcup_{n=1}^\infty B_n\right)^{\mathrm{c}}\cap E\right)\\
&\geq \mu^*\left(\left(\bigcup_{n=1}^\infty B_n\right)\cap E\right) + \mu^*\left(\left(\bigcup_{n=1}^\infty B_n\right)^{\mathrm{c}}\cap E\right) \quad \text{(subadditivity).}
\end{aligned}
$$

But this shows $\bigcup_{n=1}^\infty B_n \in \mathcal{M}$. $\square$

Lemma 8.3. *If $B_1, B_2, \ldots \in \mathcal{M}$ are disjoint, then $\mu^*(\bigcup_n B_n) = \sum_n \mu^*(B_n)$.*

Proof. The finite case $\mu^*(\bigcup_{n=1}^N B_n) = \sum_{n=1}^N \mu^*(B_n)$ was proven in the previous lemma (set $E = X$). By monotonicity, $\sum_{n=1}^N \mu^*(B_n) \le \mu^*(\bigcup_{n=1}^\infty B_n)$. Taking $N \to \infty$ gives $\sum_{n=1}^\infty \mu^*(B_n) \le \mu^*(\bigcup_{n=1}^\infty B_n)$. Inequality in the other direction is implied by subadditivity of μ^*. □

We now know μ^* satisfies all the properties of a measure on $\mathcal{M}$. In order for μ^* to be a sane extension of μ to $\mathcal{M}$, we need to impose some conditions on μ. The ones that work from experience are:

1. $\mathcal{A}$ should be an *algebra*, meaning that it is non-empty, and closed under intersection and *finite* union (and intersection).

2. If $A_1, \ldots, A_n \in \mathcal{A}$ are disjoint, then $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$. It follows that μ is monotone and finitely subadditive.

3. Also, if $A_1, A_2, \ldots \in \mathcal{A}$ are disjoint, and $\bigcup_i A_i$ happens to be in $\mathcal{A}$, the previous equation must also hold. (This is equivalent to requiring that $\mu(\bigcup_i A_i) \le \sum_i \mu(A_i)$.)

Then we have the following important result. (By the way, the name of "Carathéodory Extension Process" is often used to refer to the constructions in this section.)

Theorem 8.4. *$\mathcal{A} \subseteq \mathcal{M}$, and $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$.*

Thus μ^* is a measure extending of μ onto the sigma algebra $\mathcal{M}$ containing the algebra $\mathcal{A}$. ($\mathcal{M}$ must also then contain the sigma algebra $\mathcal{B}$ generated by $\mathcal{A}$, although $\mathcal{M}$ may be larger than $\mathcal{B}$.)

Proof. Fix $A \in \mathcal{A}$. For any $E \subseteq X$ and $\varepsilon > 0$, by definition we can find $A_1, A_2, \ldots \in \mathcal{A}$ with $E \subseteq \bigcup_n A_n$ and $\sum_n \mu(A_n) \le \mu^*(E) + \varepsilon$. Then we have

$$\begin{aligned}
\mu^*(A \cap E) + \mu^*(A^c \cap E) &\le \mu^*\Big(A \cap \bigcup_n A_n\Big) + \mu^*\Big(A^c \cap \bigcup_n A_n\Big) \\
&\le \sum_n \mu^*(A \cap A_n) + \sum_n \mu^*(A^c \cap A_n) \\
&\le \sum_n \mu(A \cap A_n) + \sum_n \mu(A^c \cap A_n) \\
&= \sum_n \mu(A_n) \quad \text{(finite additivity of } \mu) \\
&\le \mu^*(E) + \varepsilon\,.
\end{aligned}$$

ε being arbitrary, $\mu^*(A \cap E) + \mu^*(A^c \cap E) \le \mu^*(E)$, showing that $A \in \mathcal{M}$.

Now we show $\mu^*(A) = \mu(A)$. Note $\mu(A) \geq \mu^*(A)$ is always true. Consider any $A_1, A_2, \ldots \in \mathcal{A}$ with $A \subseteq \bigcup_n A_n$. By countable subadditivity and monotonicity of μ,

$$\mu(A) = \mu\Bigl(\bigcup_n A \cap A_n\Bigr) \leq \sum_n \mu(A \cap A_n) \leq \sum_n \mu(A_n)\,.$$

This implies $\mu(A) \leq \mu^*(A)$, directly from the definition of μ^*. □

Our final results for this section concern the uniqueness of this extension.

Theorem 8.5. *Assume $\mu(X) < \infty$. Let $\mathcal{B}$ be the sigma algebra generated from the algebra $\mathcal{A}$. If ν is another measure on $\mathcal{B}$, which agrees with μ^* on $\mathcal{A}$, then μ^* and ν agree on $\mathcal{B}$ as well.*

Proof. Let $B \in \mathcal{B}$.

$$\mu^*(B) = \inf_{\substack{A_1, A_2, \ldots \in \mathcal{A} \\ B \subseteq \bigcup_n A_n}} \sum_n \mu(A_n) = \inf \sum_n \nu(A_n) \geq \inf \nu\Bigl(\bigcup_n A_n\Bigr) \geq \inf \nu(B)\,.$$

So $\mu^*(B) \geq \nu(B)$. Similarly, we have $\mu^*(B^{\mathrm{c}}) \geq \nu(B^{\mathrm{c}})$, so that $\mu^*(X) - \mu^*(B) \geq \nu(X) - \nu(B) = \mu^*(X) - \nu(B)$, or $\mu^*(B) \leq \nu(B)$. (In fact, this actually proves μ^* and ν agree on $\mathcal{M}$ also, if ν is defined on $\mathcal{M}$.) □

But the hypothesis that $\mu(X) < \infty$ is clearly too restrictive. The fix is easy:

Definition 8.1. A measure space $(X, \mathcal{B}, \mu)$ is *sigma-finite*, if there are measurable sets $X_1, X_2, \ldots \subseteq X$, such that $\bigcup_n X_n = X$ and $\mu(X_n) < \infty$ for all n.

(Clearly, we may as well assume that the X_n are increasing in this definition.)

Theorem 8.6. *Theorem 8.5 holds also in the case that $(X, \mathcal{B}, \mu)$ is sigma-finite.*

Proof. Let $\{X_n\} \nearrow X$, $\mu(X_n) < \infty$ as in the definition of sigma-finiteness. (Of course we also need to assume that the X_n can be chosen from $\mathcal{A}$.) For each $B \in \mathcal{B}$, Theorem 8.5 says that $\mu^*(B \cap X_n) = \nu(B \cap X_n)$. Taking limits as $n \to \infty$ gives $\mu^*(B) = \nu(B)$. □

The restriction that X be sigma-finite is not too severe, since the usual spaces such as $\mathbb{R}^n$ *are* sigma-finite. Sigma-finiteness also comes back in the theorems of Section 11.

Finally, the corollary below is just Theorem 8.6 restated without reference to the outer measure:

Corollary 8.7 (Uniqueness of measures). *Let $\mathcal{B}$ be the sigma algebra generated by the algebra $\mathcal{A}$. Then if two measures μ and ν agree on $\mathcal{A}$, and X is sigma-finite (under either μ or ν), then $\mu = \nu$ on $\mathcal{B}$.*

9 Lebesgue Measure in $\mathbb{R}^n$

In this section we rigorously construct the n-dimensional volume measure on $\mathbb{R}^n$. As hinted before, the idea is to define the measure for rectangles and then use the extension process in Section 8. Unfortunately, there is some grunt work to do in order to verify that the hypotheses of those theorems are indeed satisfied.

Our setting will be the collection $\mathcal{R}$ of rectangles $I_1 \times \cdots \times I_n$ in $\mathbb{R}^n$, where I_k is any open, half-open or closed, bounded or unbounded, interval in $\mathbb{R}$. The following definition is merely an abstracted version of the formal facts we need about these rectangles.

Definition 9.1. Let X be any set. A *semi-algebra* is any $\mathcal{R} \subseteq 2^X$ with the following properties:

1. The empty set is in $\mathcal{R}$.
2. The intersection of any two sets in $\mathcal{R}$ is also in $\mathcal{R}$.
3. For any set in $\mathcal{R}$, its complement is expressible as a finite disjoint union of other elements of $\mathcal{R}$.

It is easy to see, although tiresome to write down formally, that the collection of all rectangles is indeed a semi-algebra. But immediately from this, we can automatically construct an algebra $\mathcal{A}$:

Theorem 9.1. *The set $\mathcal{A}$ of all finite disjoint unions of elements of a semi-algebra $\mathcal{R}$ is an algebra on X.*

Proof. We check the properties for an algebra:

1. The empty set is trivially in $\mathcal{A}$.
2. If $A = \biguplus_i R_i$, and $B = \biguplus_j S_j$, where R_i and S_i denote a finite number of sets chosen from $\mathcal{R}$, then $A \cap B = \biguplus_i R_i \ \cap \ \biguplus_j S_j = \biguplus_{i,j} R_i \cap S_j \in \mathcal{A}$.
3. If $A = \biguplus_i R_i$, then $A^{\mathrm{c}} = \bigcap_i R_i^{\mathrm{c}} = \bigcap_i \biguplus_j S_{i,j}$ for some $S_{i,j} \in \mathcal{R}$. The finite intersection belongs to $\mathcal{A}$ by the previous step. So $A^{\mathrm{c}} \in \mathcal{A}$.
4. Finally, given $A_i = \biguplus_j R_{i,j}$, for a finite number of i, let $D_0 = \emptyset$, and $D_i = D_{i-1} \uplus (A_i \setminus D_{i-1}) \in \mathcal{A}$. Then $\bigcup_i A_i = \bigcup_i D_i = \biguplus_i (D_i \setminus D_{i-1}) \in \mathcal{A}$. □

By the way, the sigma algebra generated by $\mathcal{A}$ will contain the Borel sigma algebra: every open set $U \in \mathbb{R}^n$ obviously can be written as a union of open rectangles, and in fact we can use a *countable* union of open rectangles. For, given an arbitrary collection of rectangles covering $U \subseteq \mathbb{R}^n$, there always exists a *countable* subcover[5].

The volume of $A = I_1 \times \cdots \times I_n$ is naturally defined as $\lambda(A) = \lambda(I_1) \cdot \cdots \cdot \lambda(I_n)$, $\lambda(I_k)$ being the length of the interval I_k, with the usual rules about multiplying zeroes and infinities together in force.

[5] If you are not aware of this theorem, you are invited to prove it yourself.

Now let's suppose that the rectangle A has been partitioned into a disjoint smaller rectangles. Then the sum of the volumes of the smaller rectangles, as we have defined it, should equal the volume of A. This is true, of course, although it is again tedious to write down formally. Essentially, one draws a rectangular grid on A using the boundaries of the smaller rectangles, and show that the sum of the volumes of each cell in the grid equals the volume of A, by applying the distributive property of multiplication over addition.

In the even more general case, suppose $A \in \mathcal{A}$ is a disjoint union of rectangles, but A is not necessarily in the semi-algebra of rectangles. The volume of A is defined as the sum of the volumes of the component rectangles. This is obvious — we mention it only to note that, although A may certainly have different decompositions ($A = \bigcup_i R_i = \bigcup_j S_j$), the volume sum is always the same. To see this, simply take the common refinement $R_i \cap S_j$. Then $\sum_i \lambda(R_i) = \sum_i \sum_j \lambda(R_i \cap S_j)$ by applying the result of the previous paragraph on each rectangle R_i. But $\sum_j \lambda(R_j)$ equals this double sum also.

It follows easily then, that λ is finitely additive. Thus there is only one final thing left to show: if the disjoint union of $A_1, A_2, \ldots \in \mathcal{A}$ is $C \in \mathcal{A}$, then $\lambda(C) = \sum_{i=1}^{\infty} \lambda(A_i)$. From monotonicity and taking limits we always have $\lambda(C) \geq \sum_{i=1}^{\infty} \lambda(A_i)$.

We show $\lambda(C) \leq \sum_{i=1}^{\infty} \lambda(A_i)$. Observe that this also ought to be true if C is contained in, but not necessarily equal to, the union of the A_i, and A_i need not be disjoint at all. Henceforth these are our new hypotheses.

Suppose first that C happens to be compact, and A_i are all open. In other words, $\{A_i\}$ form an open cover of the compact set C. So there is a finite subcover $A_1, \ldots, A_n$. By finite subadditivity, we have $\lambda(C) \leq \sum_{i=1}^{n} \lambda(A_i) \leq \sum_{i=1}^{\infty} \lambda(A_i)$.

Now continue to assume that C is compact, but A_i are not open. But it is easy to make the A_i open and still cover C, by slightly expanding each A_i. In particular, stipulate that the volume of each new A_i grows by at most $\varepsilon/2^i$. (We can assume the A_i are plain rectangles, rather than disjoint unions of them, and that they are bounded, since C is bounded.) Then we have $\lambda(C) \leq \sum_{i=1}^{\infty} \lambda(A_i) + \varepsilon$, and $\varepsilon > 0$ is arbitrary.

All that remains is the case that C is not compact. If C is bounded, so is its closure, and hence by the Heine-Borel theorem, C is compact. Similarly take the closure of the A_i. But taking closures of elements of $\mathcal{A}$ does not change their volumes.

Finally consider C unbounded. But $C \cap [-N, N]^n$ is bounded for each N, so from the previous case we have $\lambda(C \cap [-N, N]^n) \leq \sum_{i=1}^{\infty} \lambda(A_i)$. It is easily checked that the limit as $N \to \infty$ of the left side is exactly $\lambda(C)$.

Thus, using the theorems of Section 8, we can conclude:

Theorem 9.2. *Lebesgue measure in $\mathbb{R}^n$ exists, and it is uniquely determined, given our hypotheses.*

It is obvious from our constructions that Lebesgue measure is invariant under translations of sets. It is also invariant under other rigid motions (rotations, reflections); this will be a consequence of Lemma 12.1 and some linear algebra.

An interesting question to ask is whether there are any sets that are not Borel, or that cannot be assigned any volume. The following theorem gives a classic example (and should also serve to convince you why our strenuous efforts are necessary).

Theorem 9.3 (Vitali). *There exists a non-measurable set in* $[0,1]$ *using Lebesgue measure. In other words, Lebesgue measure cannot be defined consistently for* all *subsets of* $[0,1]$.

Proof. The key fact in this proof is translation-invariance. In particular, given any measurable $H \subseteq [0,1]$, define its "shift with wrap-around":

$$H \oplus x = \{h + x : h \in H,\, h + x \leq 1\} \cup \{h + x - 1 : h \in H,\, h + x > 1\}\,.$$

Then $\lambda(H \oplus x) = \lambda(H)$.

Define two real numbers to be equivalent if their difference is rational. The interval $[0,1]$ is partitioned by this equivalence relation. Compose a set $H \subset [0,1]$ consisting of exactly one element from each equivalence class, and also say $0 \notin H$. Then $(0,1]$ equals the disjoint union of all $H \oplus r$, for $r \in [0,1) \cap \mathbb{Q}$. Consequently, by countable additivity,

$$1 = \lambda((0,1]) = \sum_{r \in [0,1) \cap \mathbb{Q}} \lambda(H \oplus r) = \sum_{r \in [0,1) \cap \mathbb{Q}} \lambda(H)\,,$$

a contradiction, because the sum on the right can only be 0 or ∞. Hence H cannot be measurable. $\square$

A fact related to these matters is that Lebesgue measure is *complete*, meaning if $\mu(A) = 0$, then every $B \subseteq A$ is Lebesgue-measurable and $\mu(B) = 0$. (This follows directly from the construction of the outer measure in the previous section.) On the other hand, one can show that the Lebesgue measure restricted to the Borel sets in $\mathbb{R}^n$ is *not* complete. This means a slight complication in the theorems we prove about Lebesgue measure, but fortunately the extension process allows us to complete any (sigma-finite) measure if necessary.

10 Riemann integrability implies Lebesgue integrability

You have probably already suspected that any function that any Riemann-integrable function is also Lebesgue-integrable, and certainly with the same values for the two integrals. We shall prove this fact here. Let us first review the definition of the Riemann integral.

Let $A \subset \mathbb{R}^m$ be a (bounded) rectangle. Usually a bounded function $f\colon A \to \mathbb{R}$ is said to be (proper-) Riemann-integrable if the supremum of its lower sums and the infimum of its upper sums are equal:

$$\sup_{\mathcal{P}} \Bigl\{ \sum_{R \in \mathcal{P}} \mu(R) \cdot \inf_{x \in R} f(x) \Bigr\} = \inf_{\mathcal{P}} \Bigl\{ \sum_{R \in \mathcal{P}} \mu(R) \cdot \sup_{x \in R} f(x) \Bigr\}$$

(where $\mathcal{P}$ denotes a rectangular partition of A).

We can rephrase the definition by considering not just lower sums and upper sums for f, but the integral of *any simple function* $s \le f$ or $s \ge f$ (simple with respect to a rectangular partition). Such simple functions are obviously both Riemann- and Lebesgue- integrable with the same values for the integral. It is also easily seen that for every such $s \le f$ there exists some lower sum for f (in the usual sense) such that the integral of s is less than or equal to that lower sum. Similarly for the upper simple functions and the upper sums. Therefore

$$\sup_{\text{all simple } s \le f} \left\{ \int_A s \right\} = \sup_{\mathcal{P}} \left\{ \sum_{R \in \mathcal{P}} \mu(R) \cdot \inf_{x \in R} f(x) \right\},$$

$$\inf_{\text{all simple } s \ge f} \left\{ \int_A s \right\} = \inf_{\mathcal{P}} \left\{ \sum_{R \in \mathcal{P}} \mu(R) \cdot \sup_{x \in R} f(x) \right\},$$

and we may equivalently define f to be Riemann-integrable if the supremum of the integrals of the lower simple functions is equal to the infimum of the integrals of the upper simple functions.

It follows from the usual arguments, that if s_1 and s_2 are simple with $s_1 \le f \le s_2$, then $\int_A s_1 \le \int_A s_2$, and that f is Riemann-integrable if and only if there exists a sequence of lower simple functions $l_n \le f$, and upper simple functions $u_n \ge f$ such that

$$\lim_{n\to\infty} \int_A l_n = \int_A f = \lim_{n\to\infty} \int_A u_n \,.$$

This more relaxed definition of Riemann integrability is easier to work with in the proof of the following theorem.

Theorem 10.1. *Let $A \subset \mathbb{R}^m$ be a rectangle. If $f \colon A \to \mathbb{R}$ is proper Riemann-integrable, then it is also Lebesgue-integrable (with respect to Lebesgue measure) with the same value for the integral.*

Proof. f is Riemann-integrable, so choose a sequence of simple functions $l_n \le f \le u_n$ with $\lim_n \int_A l_n = \int_A f = \lim_n \int_A u_n$. Let $g_n(x) = \max_{k \le n} l_k(x)$, so that $g_n(x)$ increase to $g(x) = \sup_n g_n(x)$. By our construction, we have

$$l_n \le g_n \le g \le f \le u_n \,.$$

Since the integrals of l_n and u_n converge onto each other, we know that g is Riemann-integrable. Riemann-integrating and applying limits to the above inequality,

$$\lim_{n\to\infty} \int_A l_n \le \lim_{n\to\infty} \int_A g_n \le \int_A g \le \int_A f \le \lim_{n\to\infty} \int_A u_n \,.$$

Thus the non-negative function $f - g$ has a Riemann integral of zero, and so $f - g = 0$ almost everywhere with respect to Lebesgue measure[6]. In turn, $f - g$ must be measurable. (This follows from Lebesgue measure being complete.)

[6] This follows from a very famous theorem about Riemann integrability, whose proof you can find in [Spivak2] or [Munkres].

On the other hand, g is measurable, because g_n and l_n are, so f is measurable. Since $|f|$ is bounded and A has finite measure, f must also be Lebesgue-integrable. Then the same inequality above with the Riemann integrals changed to Lebesgue integrals shows that the Lebesgue and Riemann integrals of f are the same. $\square$

The following theorem concerns the absolutely convergent improper Riemann integral as defined in [Munkres].

Theorem 10.2. *Let $A \subseteq \mathbb{R}^m$ be open, and $f\colon A \to \mathbb{R}$ be locally bounded on A and continuous almost everywhere on A. If f is improper-Riemann-integrable (i.e. $\int_A |f| < \infty$), then its Lebesgue integral exists with the same value for the integral. Also $\int_A |f|$ diverges simultaneously for the improper Riemann integral and the Lebesgue integral.*

Proof. Suppose first that $f \geq 0$. Let C_n be a sequence of compact Jordan-measurable subsets of A whose union is A and $C_n \subset \operatorname{interior} C_{n+1}$. Then

$$\int_A f = \lim_{n\to\infty} \int_{C_n} f\,.$$

Note that $\int_{C_n} f$ is valid as both a Riemann and Lebesgue integral, by Theorem 10.1, and it can also be written as the Lebesgue integral $\int_A f \cdot \chi_{C_n}$ which converges monotonically, as $n \to \infty$, to the Lebesgue integral $\int_A f$. This must of course be equal to the left side of the equation above, which is the improper Riemann integral.

For general f, repeating the same reasoning for the non-negative functions $|f|$, f^+, f^-, in turn proves the theorem. $\square$

11 Product measures and Fubini's Theorem

Fubini's theorem concerns integrals in "multiple dimensions" and their evaluation using iterated integrals. The concept should be familiar from multi-dimensional calculus, so I won't launch myself into an extended discussion here.

But before we start writing down integral signs, we need to discuss the measurability of the sets involved in multiple integration.

Definition 11.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces. A *measurable rectangle* in $X \times Y$ is a set of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

The sigma algebra generated by all the measurable rectangles is denoted by $\mathcal{A} \otimes \mathcal{B}$, and this will be the sigma algebra we use for $X \times Y$.

Theorem 11.1. *Let E be a measurable set from $(X \times Y, \mathcal{A} \otimes \mathcal{B})$. Let*

$$E_x = \{y \in Y : (x,y) \in E\}\,, \quad x \in X\,.$$
$$E_y = \{x \in X : (x,y) \in E\}\,, \quad y \in Y\,.$$

Then $E_y \in \mathcal{A}$, $E_x \in \mathcal{B}$.

Proof. We prove the theorem for E_y; the proof for E_x is the same.

We consider the collection $\mathcal{D} = \{E \in \mathcal{A} \otimes \mathcal{B} : E_y \in \mathcal{A}\}$.

Suppose $E = A \times B$ is a measurable rectangle. Then $E_y = A$ when $y \in \mathcal{B}$, otherwise $E_y = \emptyset$. In both cases $E_y \in \mathcal{A}$, so $E \in \mathcal{D}$.

$\mathcal{D}$ is a sigma algebra, because:

1. $\emptyset_y = \{x \in X : (x, y) \in \emptyset\} = \emptyset \in \mathcal{A}$, so $\emptyset \in \mathcal{D}$.
2. If $E^n \in \mathcal{D}$, then $(\bigcup E^n)_y = \bigcup E^n_y \in \mathcal{A}$. So $\bigcup E^n \in \mathcal{D}$.
3. If $E \in \mathcal{D}$, and $F = E^{\mathrm{c}}$, then $F_y = \{x \in X : (x, y) \in F\} = \{x \in X : (x, y) \notin E\} = X \setminus E_y \in \mathcal{A}$. So $E^{\mathrm{c}} \in \mathcal{D}$.

Thus $\mathcal{D}$ is a sigma algebra containing the measurable rectangles, i.e. $\mathcal{D}$ is all of $\mathcal{A} \otimes \mathcal{B}$. □

Theorem 11.2. *Let $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ and $(Z, \mathcal{C})$ be measurable spaces. If $f \colon X \times Y \to Z$ is measurable, then the functions $f_y \colon X \to Z$, $f_x \colon Y \to Z$ obtained by holding one variable fixed are also measurable.*

Proof. Again we consider only f_y. Let $I_y \colon X \to X \times Y$ be defined by $I_y(x) = (x, y)$. Then given $E \in \mathcal{A} \otimes \mathcal{B}$, by Theorem 11.1. $I_y^{-1}(E) = E_y \in \mathcal{A}$, so I_y is a measurable function. But $f_y = f \circ I_y$. □

The next theorem on multiple integration requires the following technical tool, which comes equipped with a definition.

Definition 11.2. A family $\mathcal{A}$ of subsets of X is a *monotone class* if it is closed under increasing unions and decreasing intersections.

The intersection of any set of monotone classes is a monotone class. The smallest monotone class containing a given set $\mathcal{G}$ is the intersection of all monotone classes containing $\mathcal{G}$. This construction is analogous to the one for sigma algebras, and the result is also said to be the *monotone class generated by $\mathcal{G}$*.

Theorem 11.3 (Monotone Class Theorem). *If $\mathcal{A}$ be an algebra on X, then the monotone class generated by $\mathcal{A}$ is the same as the sigma algebra generated by $\mathcal{A}$.*

Proof. Since a sigma algebra is a monotone class, the generated sigma algebra contains the generated monotone class $\mathcal{M}$. So we only need to show $\mathcal{M}$ is a sigma algebra.

We first claim that $\mathcal{M}$ is actually closed under complementation. Let $\mathcal{M}' = \{S \in \mathcal{M} : X \setminus S \in \mathcal{M}\} \subseteq \mathcal{M}$. This is a monotone class, and it contains the algebra $\mathcal{A}$. So $\mathcal{M} = \mathcal{M}'$ as desired.

To prove that $\mathcal{M}$ is closed under countable unions, we only need to prove that it is closed under finite unions, for it is already closed under countable increasing unions.

First let $A \in \mathcal{A}$, and $\mathcal{N}(A) = \{B \in \mathcal{M} : A \cup B \in \mathcal{M}\} \subseteq \mathcal{M}$. Again this is a monotone class containing the algebra $\mathcal{A}$; thus $\mathcal{N}(A) = \mathcal{M}$.

Finally, let $S \in \mathcal{M}$, with the same definition of $\mathcal{N}(S)$. The last paragraph, rephrased, says that $\mathcal{A} \subseteq \mathcal{N}(S)$. And $\mathcal{N}(S)$ is a monotone class containing $\mathcal{A}$ by the same arguments as the last paragraph. Thus $\mathcal{N}(S) = \mathcal{M}$ as desired. □

Theorem 11.4. *Let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ be sigma-finite measure spaces. If $E \in \mathcal{A} \otimes \mathcal{B}$, then*

1. *$\nu(E_x)$ is a measurable function of $x \in X$.*
2. *$\mu(E_y)$ is a measurable function of $y \in Y$.*

Proof. We concentrate on $\mu(E_y)$. Let $\{X_m\} \nearrow X$, with $\mu(X_m) < \infty$. Fix m for now and let $\mathcal{D} = \{E \in \mathcal{A} \otimes \mathcal{B} : \mu(E_y \cap X_m) \text{ is a measurable function of } y\}$. $\mathcal{D}$ is equal to $\mathcal{A} \otimes \mathcal{B}$, because:

1. If $E = A \times B$ is a measurable rectangle, then $\mu(E_y \cap X_m) = \mu(A \cap X_m) \cdot \chi_B(y)$ which is a measurable function of y.

   If E is a finite disjoint union of measurable rectangles E^n, then $\mu(E_y \cap X_m) = \sum_n \mu(E^n_y \cap X_m)$ which is also measurable.

   The measurable rectangles form a semi-algebra (just like the rectangles in $\mathbb{R}^n$). Therefore, applying Theorem 9.1, $\mathcal{D}$ contains the algebra of finite disjoint unions of measurable rectangles.

2. If E^n are increasing sets in $\mathcal{D}$ (not necessarily measurable rectangles), then $\mu((\bigcup E^n)_y \cap X_m) = \mu(\bigcup E^n_y \cap X_m) = \lim_{n\to\infty} \mu(E^n_y \cap X_m)$ is measurable, so $\bigcup E^n \in \mathcal{D}$.

   Similarly, if E^n are decreasing sets in $\mathcal{D}$, then using limits we see that $\bigcap E^n \in \mathcal{D}$. (Here it is crucial that $E^n_y \cap X_m$ have finite measure, for the limiting process to be valid.)

3. These arguments show that $\mathcal{D}$ is a monotone class, and it contains the monotone class generated by the algebra of finite unions of measurable rectangles. By the Monotone Class Theorem, $\mathcal{D}$ must therefore be the same as the sigma algebra $\mathcal{A} \otimes \mathcal{B}$.

We now know that for each $E \in \mathcal{A} \otimes \mathcal{B}$, $\mu(E_y \cap X_m)$ is measurable, for every m. Taking limits as $m \to \infty$, we conclude that $\mu(E_y)$ is also measurable. □

One thing has been deliberately left out of our discussion so far: the construction of the product measure $\mu \otimes \nu$, which, as in the case of $\mathbb{R}^n$, should assign a measure $\mu(A)\nu(B)$ to the measurable rectangle $A \times B$. The problem is that we cannot prove countable additivity of $\mu \otimes \nu$ defined this way. We take an indirect route instead, defining it by iterated integrals:

Theorem 11.5. *Let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ be sigma-finite measure spaces. There exists a unique product measure $\mu \otimes \nu \colon \mathcal{A} \otimes \mathcal{B} \to [0, \infty]$, with*

$$(\mu \otimes \nu)(E) = \int_{x \in X} \underbrace{\int_{y \in Y} \chi_E(x, y)\, d\nu}_{\nu(E_x)}\, d\mu = \int_{y \in Y} \underbrace{\int_{x \in X} \chi_E(x, y)\, d\mu}_{\mu(E_y)}\, d\nu.$$

Proof. Let $\lambda_1(E)$ denote the double integral on the left, and $\lambda_2(E)$ denote the one on the right. (These integrals exist by Theorem 11.4.) It is obvious that both λ_1 and λ_2 are countably additive, so they are both measures on $\mathcal{A} \otimes \mathcal{B}$. Moreover, if $E = A \times B$, then just expanding the two integrals shows $\lambda_1(E) = \mu(A)\nu(B) = \lambda_2(E)$. Now $X \times Y$ is sigma-finite if X and Y are, so by uniqueness of measures[7] (Corollary 8.7), $\lambda_1 = \lambda_2$ on all of $\mathcal{A} \otimes \mathcal{B}$. $\square$

There's not much work left for our final theorems:

Theorem 11.6 (Fubini). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be sigma-finite measure spaces. If $f \colon X \times Y \to \mathbb{R}$ is $\mu \otimes \nu$-integrable, then*

$$\int_{X \times Y} f \, d(\mu \otimes \nu) = \int_{x \in X} \Bigl[\int_{y \in Y} f(x, y) \, d\nu \Bigr] \, d\mu = \int_{y \in Y} \Bigl[\int_{x \in X} f(x, y) \, d\mu \Bigr] \, d\nu \, .$$

This equation also holds when $f \geq 0$ (if it is merely measurable, not integrable).

Proof. The case $f = \chi_E$ is just Theorem 11.5. Since all three integrals are additive, they are equal for non-negative simple f, and hence also for all other non-negative f, by approximation and monotone convergence. From linearity on $f = f^+ - f^-$, we see that they are equal for any integrable f. (We will allow $\infty - \infty$ to occur on a set of measure zero.) $\square$

Theorem 11.7 (Tonelli). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be sigma-finite measure spaces, and $f \colon X \times Y \to \mathbb{R}$ be $\mu \otimes \nu$-measurable. Then f is $\mu \otimes \nu$-integrable if and only if*

$$\int_{x \in X} \Bigl[\int_{y \in Y} |f(x, y)| \, d\nu \Bigr] \, d\mu < \infty$$

(or with X and Y reversed).

Consequently, if any one of these conditions hold, then it is valid to switch the order of integration when integrating f.

Proof. Immediate from Fubini's theorem applied to the function $|f|$. $\square$

Remark 11.8. Note that it is possible that $\int_{y \in Y} |f(x, y)| \, d\nu = \infty$ on a set of measure zero in X, and still have integrability, and conversely.

12 Change of variables in $\mathbb{R}^n$

This section will be devoted to completing the proof of the differential change of variables formula, Theorem 6.5. As we noted in the remarks preceding that theorem, it suffices to prove the following.

[7] If you prefer, you can also prove this theorem using the Monotone Class Theorem instead.

Lemma 12.1. *Let $g\colon X \to Y$ be a diffeomorphism between open sets in $\mathbb{R}^n$. Then for all measurable sets $A \subseteq X$,*

$$\lambda(g(A)) = \int_{g(A)} 1 = \int_A |\det \mathrm{D}g| \,.$$

Proof. We first begin with two simple reductions.

(I) It suffices to prove the lemma locally.

That is, suppose there exists an open cover of X, $\{U_\alpha\}$, so that the equation of the lemma holds for measurable A contained inside one of the U_α. Then the equation actually holds for all measurable $A \subseteq X$.

Proof. By taking a countable subcover, we may assume there are only countably many U_i. Define the disjoint measurable sets $E_i = U_i \setminus (U_1 \cup \dots \cup U_{i-1})$, which cover X. Also define the two measures:

$$\mu(A) = \lambda(g(A))\,, \quad \nu(A) = \int_A |\det \mathrm{D}g| \,.$$

Now let $A \subseteq X$ be any measurable set. We have $A \cap E_i \subseteq U_i$, so $\mu(A \cap E_i) = \nu(A \cap E_i)$ by hypothesis. Therefore,

$$\mu(A) = \mu\Bigl(\bigcup_i A \cap E_i\Bigr) = \sum_i \mu(A \cap E_i) = \sum_i \nu(A \cap E_i) = \nu(A)\,.$$

(II) Suppose the lemma holds for two diffeomorphisms g and h, and all measurable sets. Then it holds for the composition diffeomorphism $g \circ h$ (and all measurable sets).

Proof. For any measurable A,

$$\int_{g(h(A))} 1 = \int_{h(A)} |\det \mathrm{D}g| = \int_A |(\det \mathrm{D}g) \circ h| \cdot |\det \mathrm{D}h| = \int_A |\det \mathrm{D}(g \circ h)| \,.$$

The second equality follows from Theorem 6.5 applied to the diffeomorphism h, which is valid once we know $\lambda(h(B)) = \int_B |\det \mathrm{D}h|$ for all measurable B.

We proceed to prove the lemma by induction, on the dimension n.

Base case $n = 1$. Cover X by a countable set of bounded intervals I_k in $\mathbb{R}$. By Reduction I, it suffices to prove the lemma for measurable sets contained in each of the I_k individually. By the uniqueness of measures (Corollary 8.7), it also suffices to show $\mu = \nu$ only for the intervals $[a, b]$, (a, b), etc. But this is just the Fundamental Theorem of Calculus:

$$\int_{g([a,b])} 1 = |g(b) - g(a)| = \Bigl|\int_a^b g'\Bigr| = \int_a^b |g'| \,.$$

(For the last equality, remember that the g' must be either positive on all of $[a, b]$ or negative on all of $[a, b]$. If the interval is open or half-open, we may not be able to apply the Fundamental Theorem, but the preceding equation can still be obtained via a limiting procedure.)

Induction step. Locally (i.e. on a sufficiently small open set around each point $x \in X$), g can always be factored[8] as $g = h_k \circ \cdots \circ h_2 \circ h_1$, where each h_i is a diffeomorphism and fixes one coordinate of $\mathbb{R}^n$. By Reduction I, it suffices to consider this local case only. By Reduction II, it suffices to prove the lemma for each of the diffeomorphisms h_i.

So suppose g fixes one coordinate. For convenience in notation, assume g fixes the last coordinate: $g(u, v) = (h_v(u), v)$, for $u \in \mathbb{R}^{n-1}, v \in \mathbb{R}$, and h_v are functions on (open subsets of) $\mathbb{R}^{n-1}$, in fact diffeomorphisms. Clearly h_v are one-to-one, and most importantly, $\det \mathrm{D}h_v(u) = \det \mathrm{D}g(u, v) \neq 0$.

Next, let a measurable set A be given, and consider its projection $V = \{v \in \mathbb{R} : (u, v) \in A\}$, and its cross-section $U_v = \{u \in \mathbb{R}^{n-1} : (u, v) \in A\}$.

We now apply Fubini's theorem and the induction hypothesis on h_v:

$$\begin{aligned}
\int_{g(A)} 1 &= \int_{v \in V} \int_{h_v(U_v)} 1 \\
&= \int_{v \in V} \int_{u \in U_v} |\det \mathrm{D}h_v(u)| \\
&= \int_{v \in V} \int_{u \in U_v} |\det \mathrm{D}g(u, v)| = \int_A |\det \mathrm{D}g| \,. \qquad \square
\end{aligned}$$

13 Vector-valued integrals

This section, in short, is a remark that everything we have done so far generalizes to vector-valued functions, the vectors being from real finite-dimensional spaces (or $\mathbb{C}$). These often occur in applications.

Given $f \colon X \to \mathbb{R}^n$ measurable, let $\{e_k\}$ denote the standard basis vectors in $\mathbb{R}^n$, and $\{f^k\}$ the components of f with respect to this basis. Of course we define

$$\int f = \sum_{k=1}^{n} \Bigl(\int f^k \Bigr) e_k \,,$$

provided the integrals on the right exist. Generally, to say that $\int_X f$ exists, we do not allow any one of the components to be infinite, for this is usually not useful when $n \geq 2$.

It is not hard to see that f is measurable if and only if each f^k is.

It follows immediately that this integral is linear, and hence, the definition is independent of the basis, as it ought to be. That is, if $\{e_k\}$ is *any* basis of $\mathbb{R}^n$, the sum in the definition does not change.

[8] The proof of this fact can be found in [Munkres], but it is not difficult to prove it yourself. Hint: Inverse Function Theorem.

Since by our convention that the components of the integral are not allowed to be infinite, it seems we do not need to define separately what it means for f to be integrable. But we define it, because we want to take note of some facts (admittedly they are not very interesting): f being integrable means $\int \|f\| < \infty$.

The norm can be arbitrary, for in $\mathbb{R}^n$, every norm is equivalent: if $\|\cdot\|_1$ and $\|\cdot\|_2$ are any two given norms, then there always exist constants $\alpha, \beta > 0$ such that $\alpha\|\cdot\|_2 \leq \|\cdot\|_1 \leq \beta\|\cdot\|_2$. Thus being integrable in one norm implies integrability in another norm. In particular, by using the norm $\|x\|_\Sigma = \sum_{k=1}^n |x^k|$, we see that f is integrable if and only if its components f^k are integrable.

We want to show that $\|\int f\| \leq \int \|f\|$; this basic inequality will enable us to make estimates without having to separate components. As a start, this is clearly true if f is a simple function, i.e. $f = \sum_{j=1}^m a_j \chi_{E_j}$ for $a_j \in \mathbb{R}^n$. It is also trivial if $\int \|f\| = \infty$. To prove the inequality for the other f, we use the following easy lemma.

Lemma 13.1. *Let $f \colon X \to \mathbb{R}^n$ be integrable. There exists a sequence of simple functions $\varphi_j \colon X \to \mathbb{R}^n$ converging pointwise to f, with*

$$\lim_{j\to\infty} \int \|\varphi_j - f\| = 0 \,.$$

Proof. By equivalence of norms, it suffices to prove this only for the norm $\|\cdot\|_\Sigma$ as defined above.

For each f^{k+}, by the approximation theorem in $\mathbb{R}$, there exists measurable simple φ_j^{k+} increasing to f^{k+}. Similarly for f^{k-}. Let $\varphi_j^k = \varphi_j^{k+} - \varphi_j^{k-}$. Using the Dominated Convergence Theorem applied to each component,

$$\lim_{j\to\infty} \int \|\varphi_j - f\|_\Sigma = \lim_{j\to\infty} \int \sum_{k=1}^n |\varphi_j^k - f^k| = 0 \,.$$

We also note that by the Monotone Convergence Theorem, the φ_j satisfy $\lim\limits_{j\to\infty} \int \varphi_j = \int f$. ☐

We finish our demonstration of the generalized triangle inequality. Let φ_j as in the lemma. Then

$$\left\| \int \varphi_j \right\| \leq \int \|\varphi_j\| \leq \int \|f\| + \int \|\varphi_j - f\| \,,$$

so that

$$\lim_{j\to\infty} \left\| \int \varphi_j \right\| = \left\| \lim_{j\to\infty} \int \varphi_j \right\| = \left\| \int f \right\| \leq \int \|f\| + \lim_{j\to\infty} \int \|\varphi_j - f\| = \int \|f\|.$$

14 $\mathbf{C}_0^\infty$ functions are dense in $\mathbf{L}^p(\mathbb{R}^n)$

This section is devoted to the result that the space of $\mathbf{C}_0^\infty$ functions is dense in $\mathbf{L}^p(\mathbb{R}^n)$, which was discussed at the end of Section 7.

Theorem 14.1. *Let $f\colon \mathbb{R}^n \to \mathbb{R} \in \mathbf{L}^p(\mathbb{R}^n)$, $1 \le p < \infty$. Then for any $\varepsilon > 0$, there exists $\psi \in \mathbf{C}_0^\infty$ such that*

$$\|\psi - f\|_p = \left(\int_{\mathbb{R}^n} |\psi - f|^p \, d\lambda \right)^{1/p} < \varepsilon \, .$$

Our strategy for proving this theorem is straightforward. Since we already know that the simple functions $\varphi = \sum_i a_i \chi_{E_i}$ are dense in $\mathbf{L}^p$, we should try approximating χ_{E_i} by $\mathbf{C}_0^\infty$ functions. Since $\mathbf{C}_0^\infty$ functions are non-zero on compact sets, it stands to reason that we should approximate the sets E_i by compact sets K_i. If this can be done, then it suffices to construct the $\mathbf{C}_0^\infty$ functions on the sets K_i.

Our constructions start with this last step. You might even have seen some of these constructions before.

Lemma 14.2. *Let A be a compact rectangle in $\mathbb{R}^n$. Then there exists $\phi \in \mathbf{C}_0^\infty$ which is positive on the interior of A and zero elsewhere.*

Proof. Consider the infinitely differentiable function

$$f(x) = \begin{cases} e^{-1/x^2} \, , & x > 0 \\ 0 \, , & x \le 0 \, . \end{cases}$$

If $A = [0, 1]$, then $\phi_(x) = f(x) \cdot f(1 - x)$ is the desired function of $\mathbf{C}_0^\infty$. (Draw pictures!)

If $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$, then we let

$$\phi_A(x) = \phi\left(\frac{x_1 - a_1}{b_1 - a_1}\right) \cdots \phi\left(\frac{x_n - a_n}{b_n - a_n}\right) . \qquad \square$$

Lemma 14.3. *For any $\delta > 0$, there exists an infinitely differentiable function $h\colon \mathbb{R} \to [0, 1]$ such that $h(x) = 0$ for $x \le 0$ and $h(x) = 1$ for $x \ge \delta$.*

Proof. Take the function ϕ from Lemma 14.2 for the rectangle $[0, \delta]$, and let

$$h(x) = \frac{\int_{-\infty}^x \phi(t) \, dt}{\int_{-\infty}^\infty \phi(t) \, dt} \, . \qquad \square$$

Theorem 14.4. *Let U be open, and $K \subset U$ compact. Then there exists $\psi \in \mathbf{C}_0^\infty$ which is positive on K and vanishes outside some other compact set L, $K \subset L \subset U$.*

Proof. For each $x \in U$, let $A_x \subset U$ be a bounded open rectangle containing x, whose closure $\overline{A_x}$ lies in U. The $\{A_x\}$ together form an open cover of K. Take a finite subcover $\{A_{x_i}\}$. Then the compact rectangles $\{\overline{A_{x_i}}\}$ also cover K.

From Lemma 14.2, obtain functions $\psi_i \in \mathbf{C}_0^\infty$ that are positive on $\overline{A_{x_i}}$ and vanish outside $\overline{A_{x_i}}$. Let $\psi = \sum_i \psi_i \in \mathbf{C}_0^\infty$. ψ vanishes outside $L = \bigcup_i \overline{A_{x_i}}$, which is compact. $\square$

Corollary 14.5. *In Theorem 14.4, it is even possible to require in addition that $0 \le \psi(x) \le 1$ for all $x \in \mathbb{R}^n$ and $\psi(x) = 1$ for $x \in K$.*

Proof. Let ψ be from Theorem 14.4. Since ψ is positive on the compact set K, it has a positive minimum δ there. Take the function h of Lemma 14.3 for this δ. The new candidate function is $h \circ \psi$. □

As we have said, we must now approximate arbitrary Borel sets $B \in \mathcal{B}(\mathbb{R}^n)$ by compact sets. (We will also need approximation by open sets.) It turns out that this part of the proof is purely topological, and generalizes to other metric spaces X besides $\mathbb{R}^n$. Henceforth we consider the more general case.

Let d denote the metric for the metric space X.

Theorem 14.6. *Let $(X, \mathcal{B}(X), \mu)$ be a finite measure space, and let $B \in \mathcal{B}(X)$. For every $\varepsilon > 0$, there exists a closed set V and an open set U such that $V \subseteq B \subseteq U$ and $\mu(U \setminus V) < \varepsilon$.*

Proof. Let $\mathcal{M}$ be the set of all $B \in \mathcal{B}(X)$ for which the statement is true. We show that $\mathcal{M}$ is a sigma algebra containing all the open sets in X.

1. Let $B \in \mathcal{M}$ with V and U as above. Then V^{c} open $\supseteq B^{\mathrm{c}} \supseteq U^{\mathrm{c}}$ closed, and $\mu(V^{\mathrm{c}}) - \mu(U^{\mathrm{c}}) < \varepsilon$. This shows $B^{\mathrm{c}} \in \mathcal{M}$.

2. Let $B_n \in \mathcal{M}$. Choose V_n and U_n for each B_n such that $\mu(U_n \setminus V_n) < \varepsilon/2^n$. Let $U = \bigcup_n U_n$ which is open, and $V = \bigcup_n V_n$, so that $V \subseteq \bigcup_n B_n \subseteq U$.

   Of course V is not necessarily closed, but $W_N = \bigcup_{n=1}^N V_n$ are, and these W_N increase to V. Hence $\mu(V \setminus W_N) \to 0$ as $N \to \infty$, meaning that for large enough N, $\mu(V \setminus W_N) < \varepsilon$.

   Next, we have

   $$\begin{aligned} U \setminus W_N &= (U \setminus V) \uplus (V \setminus W_N) \\ &\subseteq \bigcup\nolimits_n (U_n \setminus V_n) \cup (V \setminus W_N)\,, \\ \mu(U \setminus W_N) &= \mu(U \setminus V) + \mu(V \setminus W_N) \\ &\le \sum\nolimits_n (U_n \setminus V_n) + \mu(V \setminus W_N) < \varepsilon + \varepsilon\,. \end{aligned}$$

   This shows that $\bigcup_n B_n \in \mathcal{M}$.

3. Let B be open, and $A = B^{\mathrm{c}}$. Also let $d(x, A) = \inf_{y \in A} d(x, y)$ be the distance from $x \in X$ to A. Set $D_n = \{x \in X : d(x, A) \ge 1/n\}$. D_n is closed, because $d(\cdot, A)$ is a continuous function, and $[1/n, \infty]$ is closed.

   Clearly $d(x, A) \ge 1/n > 0$ implies $x \in A^{\mathrm{c}} = B$, but since A is closed, the converse is also true: for every $x \in A^{\mathrm{c}} = B$, $d(x, A) > 0$. Obviously the D_n are increasing, so we have just shown that they in fact increase to B. Hence $\mu(B \setminus D_n) < \varepsilon$ for large enough n. Thus $B \in \mathcal{M}$. □

The case that μ is not a finite measure is taken care of, as you would expect, by taking limits like we did for sigma-finite measures in Section 8. But since compact and open sets are involved, we need stronger hypotheses:

1. There exists $\{K_n\} \nearrow X$, with K_n compact and $\mu(K_n) < \infty$.
2. There exists $\{X_n\} \nearrow X$, with X_n open and $\mu(X_n) < \infty$.

It is easily seen that these properties are satisfied by $X = \mathbb{R}^n$ and the Lebesgue measure λ, as well as many other "reasonable" measures μ on $\mathcal{B}(\mathbb{R}^n)$. We will discuss this more later.

We assume henceforth that X and μ have the properties just listed.

Theorem 14.7. *Let $B \in \mathcal{B}(X)$ with $\mu(B) < \infty$. For every $\varepsilon > 0$, there exists a compact set V and an open set U such that $K \subseteq B \subseteq U$ and $\mu(U \setminus K) < \varepsilon$.*

Proof. It suffices to show that $\mu(U \setminus B) < \varepsilon$ and $\mu(B \setminus K) < \varepsilon$ separately.

Existence of K. Since $\{B \cap K_n\} \nearrow B$, there exists some n such that $\mu(B) - \mu(B \cap K_n) < \varepsilon/2$.

For this n, define the finite measure $\mu_{K_n}(E) = \mu(E \cap K_n)$, for $E \in \mathcal{B}(X)$. By Theorem 14.6, there are sets $V \subseteq B \subseteq U$, V closed, and $\mu_{K_n}(B \setminus V) \leq \mu_{K_n}(U \setminus V) < \varepsilon/2$. Since X_n is compact, it is closed. Then $K = V \cap K_n$ is also closed, and hence compact, because it is contained in the compact set K_n. We have,

$$\begin{aligned}\mu(B \setminus K) &= \mu(B) - \mu(B \cap K_n) + \mu(B \cap K_n) - \mu(K) \\ &= \mu(B) - \mu(B \cap K_n) + \mu_{K_n}(B) - \mu_{K_n}(V) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\,.\end{aligned}$$

Existence of U. For every n, define the finite measure $\mu_{X_n}(E) = \mu(E \cap X_n)$, for $E \in \mathcal{B}(X)$. By Theorem 14.6, there are sets $V_n \subseteq B \subseteq U_n$, U_n open, and $\mu_{X_n}(U_n \setminus B) \leq \mu_{X_n}(U_n \setminus V_n) < \varepsilon/2^n$.

Let $U = \bigcup_n U_n \cap X_n \supseteq B$. We have,

$$\mu(U \setminus B) \leq \sum\nolimits_n \mu(U_n \cap X_n \setminus B) = \sum\nolimits_n \mu_{X_n}(U_n \setminus B) < \varepsilon\,. \qquad \square$$

We return to the case of $X = \mathbb{R}^n$.

Theorem 14.8. *Let $B \in \mathcal{B}(\mathbb{R}^n)$ with $\mu(B) < \infty$. For every $\varepsilon > 0$, there exists $\psi \in \mathbf{C}_0^\infty$ such that*

$$\|\psi - \chi_B\|_p = \Big(\int_{\mathbb{R}^n} |\psi - \chi_B|^p \, d\mu\Big)^{1/p} < \varepsilon\,.$$

Proof. By Theorem 14.7, there is compact K and open U, $K \subseteq B \subseteq U$, $\mu(U \setminus K) < \varepsilon$. From Corollary 14.5, there is $\psi \in \mathbf{C}_0^\infty$ such that $\psi = 1$ on K, $\psi = 0$ outside U, and $0 \le \psi \le 1$. Then

$$\begin{aligned}\int_{\mathbb{R}^n} |\psi - \chi_B|^p &= \int_{\mathbb{R}^n \setminus U} 0 + \int_{U \setminus B} \psi^p + \int_{B \setminus K} (1-\psi)^p + \int_K 0 \\ &\le \mu(U \setminus B) + \mu(B \setminus K) \\ &= \mu(U) - \mu(B) + \mu(B) - \mu(K) \\ &< \varepsilon\,. \end{aligned}$$

□

Proof of Theorem 14.1. Let $\varphi = \sum_i a_i \chi_{E_i}$, $a_i \neq 0$ be a simple function such that $\|\varphi - f\|_p < \varepsilon/2$. Let $\psi_i \in \mathbf{C}_0^\infty$ such that $\|\psi_i - \chi_{E_i}\|_p < \varepsilon/2|a_i|$. (Note that E_i must have finite measure; otherwise φ would not be integrable.) Let $\psi = \sum_i a_i \psi_i$. Then (Minkowski's inequality),

$$\begin{aligned}\|f - \psi\|_p &\le \|f - \varphi\|_p + \|\varphi - \psi\|_p \\ &\le \|f - \varphi\|_p + \sum_i |a_i| \cdot \|\chi_{E_i} - \psi_i\|_p < \varepsilon\,. \end{aligned}$$

□

Actually, even the last part of theorem can be generalized to spaces other than $\mathbb{R}^n$: instead of infinitely differentiable functions with compact support, we consider continuous functions, defined on the metric space X, with compact support. In this case, a topological argument must be found to replace Lemma 14.2. This is easy:

Lemma 14.9. *Let A be any compact set in X. Then there exists a continuous function $\phi\colon X \to \mathbb{R}$ which is positive on the interior of A and zero elsewhere.*

Proof. Let $C = X \setminus \operatorname{interior} A$, so C is closed. Then $\phi(x) = d(x, C)$ works. ($d(x, C)$ was defined in the proof of Theorem 14.6.) □

The proof of Theorem 14.4 goes through verbatim for metric spaces X, provided that X is *locally compact*. This means: given any $x \in X$ and an open neighborhood U of x, there exists another open neighborhood V of x, such that $\overline{V}$ is compact and $\overline{V} \subseteq U$.

Finally, we need to consider when properties (1) and (2) (in the remarks preceding Theorem 14.7) are satisfied. These properties are somewhat awkward to state, so we will introduce some new conditions instead.

Definition 14.1. A measure μ on a topological space X is *locally finite* if for each $x \in X$, there is an open neighborhood U of x such that $\mu(U) < \infty$.

It is easily seen that when μ is locally finite, then $\mu(K) < \infty$ for *every* compact set K.

Definition 14.2. A topological space X is *strongly sigma-compact* if there exists a sequence of open sets X_n with compact closure, and $\{X_n\} \nearrow X$.

If X is strongly sigma-compact, and μ is locally finite, then properties (1) and (2) are automatically satisfied. It is even true that strong sigma-compactness implies local compactness in a metric space. (The proof requires some topology and is left as an exercise.) Then we have the following theorem:

Theorem 14.10. *Let X be a strongly sigma-compact metric space, and μ be any locally finite measure on $\mathcal{B}(X)$. Then the space of continuous functions with compact support is dense in $\mathbf{L}^p(X, \mathcal{B}(X), \mu)$, $1 \leq p < \infty$.*

15 Other examples of measures

Since so far we have chiefly worked only in $\mathbb{R}^n$ with Lebesgue measure, it should be of interest to give a few more useful examples of measures.

k-dimensional volume of a k-dimensional manifold

A manifold is a generalization of curves and surfaces to higher dimensions, and sometimes even to spaces other than $\mathbb{R}^n$. But here we shall concentrate on differentiable manifolds inside $\mathbb{R}^n$; the theory is elucidated in [Spivak2] or [Munkres]. Here we give a definition of the k-dimensional volume for k-dimensional manifolds which does not require those dreaded "partitions of unity".

Suppose a k-dimensional manifold $M \subseteq \mathbb{R}^n$ is covered by a single coordinate chart $\alpha \colon U \to M$, $U \subseteq \mathbb{R}^k$ open. Let $\mathrm{D}\alpha$ denote the n-by-k matrix

$$D\alpha = \begin{bmatrix} \dfrac{d\alpha}{dt_1} & \dfrac{d\alpha}{dt_2} & \dots & \dfrac{d\alpha}{dt_k} \end{bmatrix}.$$

(More precisely, each vector $\frac{d\alpha}{dt_i}$ is represented by a column vector in the standard basis of $\mathbb{R}^n$. Actually our definition works using any orthonormal basis also.)

Define, for any vectors $v_1, \dots, v_k \in \mathbb{R}^n$ (again represented in an orthnormal basis):

$$\begin{aligned} \mathrm{V}(v_1, \dots, v_k) &= \sqrt{\det \begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix}^{\mathrm{tr}} \begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix}} \\ &= \sqrt{\det \left[v_i \cdot v_j \right]_{i,j=1,\dots,k}}\,, \end{aligned}$$

This is the k-dimensional volume of a k-dimensional parallelopiped spanned by the vectors $v_1, \dots, v_k$ in $\mathbb{R}^n$. One easily shows that this volume is invariant under orthogonal transformations, and that it agrees with the usual k-dimensional volume (as defined by the Lebesgue measure) when the parallelopiped lies in the subspace $\mathbb{R}^k \times 0 \subseteq \mathbb{R}^n$.

The k-dimensional volume of any $E \in \mathcal{B}(M)$ is defined as:

$$\nu(E) = \int_{\alpha^{-1}(E)} \mathrm{V}(\mathrm{D}\alpha)\, d\lambda\,.$$

(Since α is continuous, $\alpha^{-1}(E) \in \mathcal{B}(\mathbb{R}^k)$.)

The integrand, of course, is supposed to represent "infinitesimal" elements of surface area (k-dimensional volume), or approximations of the surface area of E by polygons that are "close" to E. As indicated by the quotation marks, these assertions about "surface area" are completely non-rigorous, and we won't belabour to prove them, since the equation above *is* our definition of k-dimensional volume. But it should be pointed out that there are better theories of k-dimensional volume available, which are intrinsic to the sets being measured, instead of our computational theory. (I don't know these other theories well enough though.)

Back to our definitions. If M is not covered by a single coordinate chart, but more than one, say $\alpha_i \colon U_i \to M$, $i = 1, 2, \ldots$, then partition M with $V_1 = \alpha_1(U_1)$, $V_i = \alpha_i(U_i) \setminus V_{i-1}$, and define

$$\nu(E) = \sum_i \int_{\alpha_i^{-1}(E \cap V_i)} \mathrm{V}(D\alpha)\, d\lambda\,.$$

It is left as an exercise to show that $\nu(E)$ is well-defined: it is independent of the coordinate charts α_i used for M.

Finally, the scalar integral of $f \colon M \to \mathbb{R}$ over M is simply

$$\int_M f \, d\nu\,.$$

And the integral of a differential form ω on an oriented manifold M is

$$\int_{p \in M} \omega\bigl(p\,; T(p)\bigr)\, d\nu\,,$$

where $T(p)$ is an orthonormal frame of the tangent space of M at p, oriented according to the given orientation of M. (If you don't know what I'm talking about, just ignore this definition — essentially it generalizes the line and surface integrals in calculus.)

Again it is not hard to show that the formulae I have given are exactly equivalent to the classical ones for evaluating scalar integrals and integrals of differential forms, which are of course needed for actual computations. But there are several advantages to our new definitions. First is that they are elegant: they are mostly coordinate-free, and all the different integrals studied in calculus have been unified to the Lebesgue integral by employing different measures. In turn, this means that the nice properties and convergence theorems we have proven all carry over to integrals on manifolds.

For example, everybody "knows" that on a sphere, any circular arc C has "measure zero", and so may be ignored when integrating over the sphere. To prove this rigorously using our definitions, we only have to remark that $\nu(C) = 0$, since $\lambda(\alpha^{-1}(C)) = 0$ for a coordinate chart α for the sphere.

Stieltjes measure

The definition of the Stieltjes measure is best motivated by probability theory. Suppose we have a random variable Z with distribution μ, and the cumulative distribution function $F\colon \mathbb{R} \to [0,1]$ — by definition, they satisfy $F(z) = \Pr[Z \leq z] = \mu\bigl([-\infty, z]\bigr)$. It follows that F is (non-strict) increasing, and $\mu((a,b]) = F(b) - F(a)$.

The idea here is to try to reverse this procedure: given any increasing function F, can we construct a measure μ on $\mathcal{B}(\mathbb{R})$ that assigns, to any interval $(a,b]$, a "length" of $F(b) - F(a)$?

Actually we will need to impose some conditions on F first. Since F is increasing, it always has only a countable number of discontinuities, and these discontinuities must all be jump discontinuities. At these jumps, we will insist that F is right-continuous, i.e. $\lim_{x \searrow a} F(x) = F(a)$. Otherwise, taking countable limits may fail: for example, if at the point a, F is left-continuous instead of right-continuous, then

$$\begin{aligned} \mu\bigl((a,b]\bigr) &= F(b) - F(a) \\ &\neq \lim_{x \searrow a} (F(b) - F(x)) \\ &= \lim_{x \searrow a} \mu\bigl((a,x]\bigr) \,. \end{aligned}$$

(Of course, the preference of "right" over "left" comes from our convention that we used intervals $(a,b]$ that are open on the left and closed on the right.)

We will also insist that $F(x) < \infty$ for $x \neq -\infty, +\infty$, so that the subtraction $F(b) - F(a)$ makes sense. However, it should be allowed that, say, $F(-\infty) = -\infty$ (and $F(x)$ does not have to be in $[0,1]$ either). This allows sigma-finite measures to be constructed.

With the necessary conditions now stated, we can begin the construction of μ, which is not much different from the construction of the Lebesgue measure on $\mathbb{R}$ — not surprising, since $F(x) = x$ is exactly the Lebesgue measure on $\mathbb{R}$.

First, it is easily checked, by drawing pictures of the intervals $(a,b]$, that they actually form a semi-algebra on $\mathbb{R}$. (Just ignore the point $-\infty$ for now.) The measure μ on the generated *algebra* $\mathcal{A}$ is defined in the obvious way, and it follows from the same arguments as in Section 9 that μ is finitely additive.

Countable additivity requires the typical approximation arguments. Suppose that we have $J_n \in \mathcal{A}$ with infinite disjoint union $I \in \mathcal{A}$. By monotonicity we automatically have $\sum_{n=1}^{\infty} \mu(J_n) \leq \mu(I)$, so we only have to prove the other inequality. We can assume that J_n are simple intervals $(a_n, b_n]$, instead of finite disjoint unions of intervals. Also assume, for now, that I is the finite interval $(a,b]$.

Since F is right-continuous, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$0 \leq F(b) - F(a) < F(b) - F(a+\delta) + \varepsilon \,.$$

Also there exists $\delta_n > 0$ such that

$$0 \leq F(b_n + \delta_n) - F(a_n) < F(b_n) - F(a_n) + \frac{\varepsilon}{2^n} \,.$$

There exists a finite set $n_1, \ldots, n_k$ such that $(a+\delta, b] \subseteq \bigcup_{i=1}^{k} (a_{n_i}, b_{n_i} + \delta_{n_i}]$, since the open sets $(a_n, b_n + \delta)$ cover the compact set $[a + \delta, b]$. Therefore,

$$\begin{aligned} F(b) - F(a+\delta) &\leq \sum_{i=1}^{k} F(b_{n_i} + \delta_{n_i}) - F(a_{n_i}) \\ &\leq \sum_{n=1}^{\infty} F(b_n + \delta_n) - F(a_n)\,, \\ F(b) - F(a) &\leq 2\varepsilon + \sum_{n=1}^{\infty} F(b_n) - F(a_n)\,. \end{aligned}$$

and we take $\varepsilon \to 0$.

Finally, the proof for general $I = I_1 \uplus \cdots \uplus I_k \in \mathcal{A}$ just follows from finite additivity and that the finite sum of limits equals the limit of the finite sum. Infinite intervals are handled in the same way as in Section 9.

Thus using the theorems of Section 8, μ can thereby be extended to a measure on $\mathcal{B}(\mathbb{R})$. This is called the *Stieltjes measure* on $\mathbb{R}$, and the integral

$$\int_{\mathbb{R}} g\, d\mu = \int_{\mathbb{R}} g\, dF$$

is the Stieltjes integral, and it generalizes the Riemann-Stieltjes integral that is sometimes studied in real analysis courses. (The Riemann-Stieltjes integral is defined by taking limits of Riemann-like sums $\sum_i g(\xi_i) \cdot (F(x_i) - F(x_{i-1}))$. Showing this limit exists requires some effort, however.)

Of course, if F is differentiable, the Stieltjes integral just reduces to

$$\int_{\mathbb{R}} g\, dF = \int_{\mathbb{R}} g \cdot F'\, d\lambda\,.$$

Lastly, we should mention that if we admit *signed measures*, which are differences of two (positive) measures, then the condition that F be increasing can even be relaxed. We will not pursue that theory here though.

16 Egorov's Theorem

The following theorem does not really belong in a first course, but it is quite a surprising and interesting result, and I want to record its proof.

Theorem 16.1 (Egorov). *Let (X, μ) be a measure space of finite measure, and $f_n \colon X \to \mathbb{R}$ be a sequence of measurable functions convergent almost everywhere to f. Then given any $\varepsilon > 0$, there exists a measurable subset $A \subseteq X$ such that $\mu(X \setminus A) < \varepsilon$ and the sequence f_n converges uniformly to f on A.*

Proof. First define

$$B_{n,m} = \bigcap_{k=n}^{\infty} \Big[\, |f - f_k| < \tfrac{1}{m} \,\Big]\,.$$

Fix m. For most $x \in X$, $f_n(x)$ converges to $f(x)$, so there exists n such that $|f_k(x) - f(x)| < 1/m$ for all $k \geq n$, so $x \in B_{n,m}$. Thus we see $\{B_{n,m}\}_n \nearrow X \setminus C$ (C is some set of measure zero).

We construct the set A inductively as follows. Set $A_0 = X \setminus C$. For each $m > 0$, since $\{A_{m-1} \cap B_{n,m}\}_n \nearrow A_{m-1}$, we have $\mu(A_{m-1} \setminus B_{n,m}) \to 0$, so we can choose $n(m)$ such that

$$\mu(A_{m-1} \setminus B_{n(m),m}) < \frac{\varepsilon}{2^m} .$$

Furthermore set

$$A_m = A_{m-1} \cap B_{n(m),m} .$$

Since $A_m \uplus (A_{m-1} \setminus B_{n(m),m}) = A_{m-1}$, we have

$$\begin{aligned} \mu(A_m) &> \mu(A_{m-1}) - \frac{\varepsilon}{2^m} \\ &> \mu(X) - \frac{\varepsilon}{2} - \frac{\varepsilon}{4} - \cdots - \frac{\varepsilon}{2^m} \geq \mu(X) - \varepsilon . \end{aligned}$$

The sets A_m are decreasing, so letting

$$A = \bigcap_{m=1}^{\infty} A_m = \bigcap_{m=1}^{\infty} B_{n(m),m} ,$$

we have $\mu(A) \geq \mu(X) - \varepsilon$, or $\mu(X \setminus A) \leq \varepsilon$. Finally, for $x \in A$, $x \in B_{n(m),m}$ for all m, showing that $|f(x) - f_k(x)| < 1/m$ whenever $k \geq n(m)$. This condition is uniform for all $x \in A$. $\square$

17 Exercises

I have been suggested to provide some more exercises to this text. Here they are.

1. In $\mathbb{R}^n$ with Lebesgue measure, find an uncountable set of measure zero.
2. Show that if $f\colon \mathbb{R}^n \to \mathbb{R}$ is continuous and equal to zero almost everywhere, then f is in fact equal to zero everywhere.
3. Find a sequence of integrable functions f_n such that $f_n(x) \to 0$ for every x but $\int f_n \to \infty$.
4. Let $f \in \mathbf{L}^1(\mathbb{R}^n)$, $1 \leq p < \infty$. Compute the limits

$$\lim_{h \to 0} \int |f(x+h) - f(x)|^p \, dx , \qquad \lim_{\|h\| \to \infty} \int |f(x+h) - f(x)|^p \, dx .$$

5. Let $f \in \mathbf{L}^1(\mathbb{R}^n)$. Show that

$$\lim_{\|y\| \to \infty} \int_{\mathbb{R}^n} f(x) e^{i\langle x,y \rangle} dx = 0 .$$

6. Let $f_n \in \mathbf{L}^p, 1 < p < \infty$ be a sequence of functions converging to $f \in \mathbf{L}^p$ almost everywhere, and suppose there is a constant M such that $\|f_n\|_p \leq M$ for all n. Then for each $g \in \mathbf{L}^q$,

$$\int fg = \lim_{n\to\infty} \int f_n g \,.$$

Hint: Use Egovov's Theorem and a density argument. Is this also true for $p = 1$?

7. Let $f_n \in \mathbf{L}^p, 1 \leq p < \infty$ be a sequence of functions converging $f \in Le^p$ almost everywhere. Prove that f_n converges to f in the Le^p norm if and only if $\|f_n\|_p \to \|f\|_p$.

8. Let $f \in \mathbf{L}^p(\mathbb{R})$, $g \in \mathbf{L}^q(\mathbb{R})$, with $1 \leq p, q \leq \infty$. Show that the function $F(x) = \int_0^x f(t)\,dt$ is defined and continuous for all $x \in \mathbb{R}$, and that the function $h(x) = (|x|+1)^{-a} F(x) g(x)$ is in $\mathbf{L}^1(\mathbb{R})$, the constant a being larger than $2 - \frac{1}{p} - \frac{1}{q}$.

9. Let $f \in \mathbf{L}^1(\mathbb{R})$. Show that the series

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} f(x\sqrt{n})$$

is convergent for almost all $x \in \mathbb{R}$.

10. Let f be in $\mathbf{L}^1(\mathbb{R})$, and g be a continuous periodic function with period 1. Show that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x) g(nx)\,dx = \int_{-\infty}^{\infty} f(x)\,dx \int_0^1 g(y)\,dy \,.$$

18 Bibliography

The following outlines the prerequesites for this article. (Although I'm not suggesting that you must first know everything here before you read this article; you could be learning these as you go along.)

First, you need a respectable first-year calculus course, dealing with limits rigorously. The course I took used [Spivak1], possibly the best math book ever.

You probably should be at least somewhat familiar with multi-dimensional calculus, if only to have a motivation for the theorems we prove (e.g. Fubini's Theorem, Change of Variables). I learned multi-dimensional calculus from [Spivak2] and [Munkres]. As you'd expect, these are theoretical books, and not very practical, but we will need a few elementary results that these books prove.

Point-set topology is also introduced in the study of multi-dimensional calculus. We will not need a deep understanding of that subject here, but just the basic definitions and facts about open sets, closed sets, compact sets, continous maps between topological spaces, and metric spaces. I don't have particular references for these, as it has become popular to learn topology with Moore's method (as I have done), where you are given lists of theorems that you are supposed to prove alone.

The last book, [Rosenthal], (not a prerequesite) is what I mostly referred to while writing up Section 8. It contains applications to probability of the abstract measure stuff we do here, and it is not overly abstract. I recommend it, and it's cheap too.

I don't mention any of the standard real analysis or measure theory books here, since I don't have them handy, and this text is supposed to supplant a fair portion of these books anyway. But surely you can find references elsewhere.

References

[Spivak1] Michael Spivak, *Calculus* (3rd ed.). Publish or Perish, 1994; ISBN 0-914098-89-6.

[Spivak2] Michael Spivak, *Calculus on Manifolds*. Perseus, 1965; ISBN 0-8053-9021-9.

[Munkres] James R. Munkres, *Analysis on Manifolds*. Westview Press, 1991; ISBN 0-201-51035-9.

[Rosenthal] Jeffrey S. Rosenthal, *A First Look at Rigorous Probability Theory*. World Scientific, 2000; ISBN 981-02-4303-0.