Liam O'Carroll Summer 2021

Notes on Measure Theory and Lebesgue Integration

Contents

1 About these notes

These notes follow this series of lectures on measure theory by the YouTube channel "The Bright Side of Mathematics." (I highly recommend this channel for its concise yet formal lectures on a variety of mathematical subjects.) Each section corresponds to a different lecture in the series. While I plan to include all definitions, theorems, and important exposition/intuition, I likely will not cover proofs in formal detail. Rather, I will seek to convey the intuition behind proofs while referring the reader back to the corresponding part of the lecture for formal details.

If you end up using these notes, I would highly recommend also reading Chapter 11 of [Rud76]. It covers many of the same results but with less intuition provided, although it also includes full proofs unlike these notes.

2 Sigma algebras

Key takeaways: *When you see a σ-algebra, think "subsets I can measure." The definition of a σ-algebra was constructed with this in mind!*

Measure theory is the study of the generalization and formalization of geometric measures (distance/length, area, volume). If we just have an interval on the real number line $[a, b]$, then we know of course that its length is $b - a$. But we may want to measure more complex sets (including higher-dimensional sets) or deal with different notions of length (depending on the problem, maybe we want to add weights at different parts)—measure theory allows us to formalize all of this.

Thus, let X be an arbitrary set which we want to measure subsets of, and consider the power set $\mathcal{P}(X)$. It typically won't be possible to measure every element of $\mathcal{P}(X)$, but we want to develop a theory which can measure a lot of $\mathcal{P}(X)$ (or "good" elements of $\mathcal{P}(X)$). This leads to the following definition:

Definition 2.1 (Sigma algebra) *Let X be some set. $\mathscr{A} \subseteq \mathcal{P}(X)$ is called a σ-algebra ("sigma-algebra") if it satisfies the following rules:*

1. $\emptyset, X \in \mathscr{A}$,

2. $A \in \mathscr{A} \Rightarrow A^c := X \setminus A \in \mathscr{A}$ *(closed under complementation),*

3. $A_i \in \mathscr{A}$ *for* $i \in \mathbb{N}$ *implies* $\bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$ *(closed under countable unions). (Note that it can be that* $A_i = A_j$ *for* $i \neq j$*, meaning* $\mathscr{A}$ *is also closed under finite unions.)*

Using these three properties, one can show that a σ*-algebra is also closed under countable intersections. A* σ*-algebra is also closed under set difference since* $A \setminus B = A \cap B^c$*. Also, note that due to Property 2, Property 1 can be reduced to either* $\emptyset \in \mathscr{A}$ *or* $X \in \mathscr{A}$*.*

We call $A \in \mathscr{A}$ *a measurable set. (In other words, it is measurable with respect to the* σ*-algebra* $\mathscr{A}$*. We may say that* A *is* $\mathscr{A}$*-measurable if we want to make it clear which sigma algebra it is measurable with respect to.)*

Note that Definition 2.1 is very natural and corresponds to our intuition of "what should be measurable." (For the third requirement, the point being that intuitively we should be able to measure a union of measurable sets by adding the measures of the individual sets in the union.)

$\mathscr{A} = \{\emptyset, X\}$ is the smallest σ-algebra and $\mathscr{A} = \mathcal{P}(X)$ is the largest. Typically it won't be possible to measure every set in $\mathcal{P}(X)$, so $\mathscr{A}$ will lie somewhere between these extremes. We would like to be closer to $\mathcal{P}(X)$ though so that we can measure many sets.

3 The Borel σ-algebra

Key takeaways: *We can construct new* σ*-algebras by taking the intersection of other* σ*-algebras. The Borel* σ*-algebra on* $\mathbb{R}^n$ *is what we typically think about when we think of measurable subsets of* $\mathbb{R}^n$*.*

Once again, let X be some set which we want to measure subsets of. (In the future, we will not keep redefining X.) The following is easy to show:

Theorem 3.1 (σ-algebras are closed under intersection) *Suppose* $\mathscr{A}_i$ *is a sigma algebra on* X *for all* $i \in I$*, where* I *is a (potentially uncountable) index set. Then* $\bigcap_{i \in I} \mathscr{A}_i$ *is also a* σ*-algebra on* X*.*

Theorem 3.1 is helpful if there are many properties which you want your measurable sets to have. You can just put the properties into different $\mathscr{A}_i$'s and then take the intersection to ensure that every element of your final σ-algebra satisfies all of the properties.

Definition 3.2 (Smallest σ-algebra) *Let* $M \subseteq \mathcal{P}(X)$*. Then there exists a smallest* σ*-algebra which contains* M*. This follows due to Theorem 3.1; we can take all of the* σ*-algebras which contain* M *and take their intersection to get the smallest* σ*-algebra which contains* M*, which we know is a* σ*-algebra due to Theorem 3.1. We denote this smallest* σ*-algebra as* $\sigma(M)$*:*

$$\sigma(M) := \bigcap_{\substack{\mathscr{A} \supseteq M, \\ \mathscr{A} \text{ is a } \sigma\text{-algebra}}} \mathscr{A}.$$

For example, if $X = \{a, b, c, d\}$ and $M = \{\{a\}, \{b\}\}$, then one can check that

$$\sigma(M) = \{\emptyset, X, \{a\}, \{b\}, \{a, b\}, \{b, c, d\}, \{a, c, d\}, \{c, d\}\}.$$

Definition 3.3 (Borel σ-algebra) *Let (X, τ) be a topological space.*[1] *The Borel σ-algebra on X, denoted $\mathcal{B}(X)$, is the σ-algebra generated by the open sets of X. In other words, $\mathcal{B}(X) := \sigma(\tau)$.*

Note that we just write $\mathcal{B}(X)$ instead of $\mathcal{B}(X, \tau)$ since typically the topology is clear.

Also, note that clearly the Borel σ-algebra contains all of the open subsets of X. It also contains all of the closed subsets of X by Condition 2 in Definition 2.1 (since by definition, a closed set is a set whose complement is an open set). See this for a more detailed characterization of the Borel σ-algebra, which is also sometimes just referred to as the Borel algebra.

Though, as noted in Definition 3.3, we only write $\mathcal{B}(X)$ instead of $\mathcal{B}(X, \tau)$, keep in mind that $\mathcal{B}(X)$ is intimately tied to the topological structure of X. For example, if $X = \mathbb{R}^n$ and we are working with the *standard topology* (the topology on $\mathbb{R}^n$ which induces the usual open/closed sets), then $\mathcal{B}(X) = \mathcal{B}(\mathbb{R}^n)$ is very large, which is good because it means we can measure lots of sets. But as it turns out, $\mathcal{B}(\mathbb{R}^m) \neq \mathcal{P}(\mathbb{R}^n)$. As we will see in the next lecture though, this is very purposeful. In the case of $\mathbb{R}^n$, we need $\mathcal{B}(\mathbb{R}^n)$ and not $\mathcal{P}(\mathbb{R}^n)$ as our σ-algebra because our measure (defined soon) must satisfy some meaningful rules. These rules can be fulfilled on the whole power set, but luckily $\mathcal{B}(\mathbb{R}^n)$ contains everything we really care to measure.

4 Measures

Key takeaways: *A measure formalizes what it means to take the volume of a set. However, there are nonstandard measures even when our set is $\mathbb{R}^n$.*

To define what a measure is, we first need to make the following definition:

Definition 4.1 (Measurable space) *A pair $(X, \mathscr{A})$ is a measurable space, where X is a set and $\mathscr{A}$ is a σ-algebra on X.*

Then:

Definition 4.2 (Measure) *Given a measure space $(X, \mathscr{A})$, a map $\mu : \mathscr{A} \to [0, \infty]$*[2] *is called a measure if it satisfies:*

1. $\mu(\emptyset) = 0$,

[1] Here, τ is the set of open subsets of X. If you don't know what a topological space is, this definition works just as well if (X, d) is a metric space or if X is just $\mathbb{R}^n$. All we really need is "open sets" to be defined. (Note that every metric space is a topological space since the metric induces a topology.)

[2] We have that $[0, \infty] := [0, \infty) \cup \{\infty\}$, where the symbol ∞ (or $+\infty$) comes from the extended real number line and satisfies certain properties. The point being that a measure can assign ∞ to an element of $\mathscr{A}$. One thing to note though (this is also mentioned on the linked Wikipedia page): Unlike in other areas of math where $0 \cdot \pm\infty$ is typically undefined, in measure theory and probability we usually use the convention $0 \cdot \pm\infty := 0$.

2. *$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ for $A_j \in \mathscr{A}$ if $A_i \cap A_j = \emptyset$ for all $i \neq j$. (Additivity under a countable union as long as the subsets are disjoint. The property "additivity under a countable union" is also known as "σ-additive.")*

Again, Definition 4.2 is very intuitive. The empty set should have zero "volume." Adding disjoint sets should add their volumes. Adding a countably infinite number of sets can be thought of as approximating a larger set, in which case we still want the volumes to add to result in the volume of the larger set. Also, of course the volume of any set should be nonnegative. (I'm using "volume" loosely here.)

Definition 4.3 (Measure space) *A measure space is a triple $(X, \mathscr{A}, \mu)$ where X is some set, $\mathscr{A}$ is a σ-algebra on X, and μ is measure on $\mathscr{A}$.*

As an example, let X be some set and let $\mathscr{A} = \mathcal{P}(X)$. The *counting measure* is defined as

$$\mu(A) := \begin{cases} \text{number of elements in } A & \text{if } A \text{ has finitely many elements,} \\ \infty & \text{otherwise,} \end{cases}$$

for $A \in \mathscr{A}$.

Another example is the *Dirac measure* which is parameterized by some $p \in X$ and denoted $\delta_p(A)$ for $A \in \mathscr{A}$:

$$\delta_p(A) := \begin{cases} 1 & \text{if } p \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

You can think of p as being a point charge: If a set contains p it has charge 1; otherwise it has charge 0.

Finally, we of course want to define a measure when $X = \mathbb{R}^n$ which corresponds to our usual notion of "volume." In particular, it should satisfy

1. $\mu([0,1]^n) = 1$ (volume of a unit cube is 1), and
2. $\mu(x + A) = \mu(A)$ for all $x \in \mathbb{R}^n, A \in \mathscr{A}$ (invariant under translation).

As we will see, we will be able to come up with such a σ and it will be called the Lebesgue measure. However, the underlying σ-algebra $\mathscr{A}$ will *not* be $\mathcal{P}(\mathbb{R}^n)$ but the Borel σ-algebra on $\mathbb{R}^n$.

5 Not everything is Lebesgue measurable

Key takeaways: *If we want to come up with a measure when our underlying set X is $\mathbb{R}$ which corresponds to our intuition of length, our σ-algebra must be more restrictive than $\mathcal{P}(\mathbb{R})$ due to the presence of "nasty sets."*

In this lecture, we solve the "measure problem": We search for a measure μ on some σ-algebra $\mathscr{A}$ of $\mathbb{R}$ which corresponds to our usual notion of length. As stated at the end of the previous lecture, μ should satisfy

1. $\mu([a, b]) = b - a$ for $b > a$ (usual notion of the length of an interval), and
2. $\mu(x + A) = \mu(A)$ for $x \in \mathbb{R}, A \in \mathscr{A}$ (invariant under translation).

We will refer to these as Condition 1 and Condition 2.

In the case where $\mathscr{A} = \mathcal{P}(\mathbb{R})$, it has been known for over a hundred years that such a measure μ *does not exist*. In other words,

Theorem 5.1 (Can't have $\mathscr{A} = \mathcal{P}(\mathbb{R})$) *There is no measure μ on $\mathcal{P}(\mathbb{R})$ which satisfies Conditions 1 and 2.*

In this lecture, the following claim is proven instead which can easily be shown to imply Theorem 5.1:

Lemma 5.2 *Let μ be a measure on $\mathcal{P}(\mathbb{R})$ with $\mu((0, 1]) < \infty$ which also satisfies Condition 2. Then it must be the trivial measure ($\mu = 0$).*

Proof: At a very high level, the idea is to construct a special set which can't be given a reasonable "length" or measure. In other words, if we weren't dealing with the trivial measure, the existence of such a set leads to contradictions. Ultimately, the only way to deal with such strange sets is to remove them from our σ-algebra (meaning we won't try to assign a measure to them), hence why we can't let $\mathscr{A} = \mathcal{P}(\mathbb{R})$. ■

6 Measurable maps

Key takeaways: *Measurable functions are exactly those functions which we plan to be able to integrate.*

Definition 6.1 (Measurable function) *Let $(\Omega_1, \mathscr{A}_1), (\Omega_2, \mathscr{A}_2)$ denote two measurable spaces (Definition 4.1). A function $f : \Omega_1 \to \Omega_2$ is measurable if $f^{-1}(A_2) \in \mathscr{A}_1$ for all $A_2 \in \mathscr{A}_2$. ($f^{-1}(A_2)$ denotes the preimage of A_2.)*

We can express this more cleanly as follows: Recall from Definition 2.1 that elements of a σ-algebra are referred to as measurable sets. Then f is measurable if the preimage of any measurable set is itself measurable.

If we are working with multiple σ-algebras on Ω_1, Ω_2, then we may specifically state that f is measurable with respect to $\mathscr{A}_1, \mathscr{A}_2$.

He provides some really nice intuition behind why we care about measurable functions early on in the video. Basically, you should think of the function f in Definition 6.1 as being a function which you want to integrate. Suppose for simplicity that $\Omega_1 = \Omega_2 = \mathbb{R}$, and consider the following:

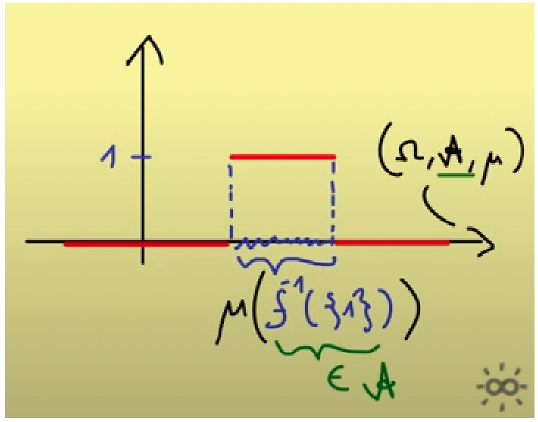

Figure 1: Intuition behind a measurable function

Here f is denoted by the read lines; in other words, f takes the value 1 on some unspecified interval (which could potentially actually be disconnected/look a lot more complicated) and is 0 everywhere else. Since f is 1 everywhere on this interval, the integral of f from $-\infty$ to ∞ should just be

$$1 \cdot \text{length of the interval where } f \text{ is } 1 = 1 \cdot \mu_x(f^{-1}(\{1\})),$$

where μ_x is some measure defined on the input set $\mathbb{R}$ ($\Omega = \mathbb{R}$ in the picture). Assuming $\{1\}$ is a measurable set with respect to whatever measure we are associating with the output (μ_y if you want), our hope then is that the preimage of this measurable set is also measurable so that $\mu_x(f^{-1}(\{1\}))$ is well-defined. If f is a measurable function, then this is of course guaranteed—the preimage of any measurable set on the y-axis is a measurable set on the x-axis. More generally, the x- and y-axes can be any two measure spaces (Definition 4.3), as he denoted in Figure 1 when he labeled the x-axis $(\Omega, \mathscr{A}, \mu)$. To conclude, Definition 6.1 will solve a problem that will pop up later if we don't restrict ourselves to working with measurable functions.

Example: As an example of a measurable function, let $(\Omega, \mathscr{A})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ denote the input and output measurable spaces respectively of Definition 6.1. (I.e., we are keeping the input space arbitrary and choosing the output space to be $\mathbb{R}$ along with the Borel σ-algebra on $\mathbb{R}$—see Definition 3.3.) We define the characteristic or indicator function of a set $\mathscr{A} \subseteq \Omega$ to be

$$\chi_A : \Omega \to \mathbb{R}, \qquad \chi_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It is easy to check that if A is measurable per Definition 2.1, equivalently $a \in \mathscr{A}$, then χ_A is a measurable map. Indeed, one can verify that the only possible preimages of χ_A are $\emptyset, \Omega, A, A^c$, which are all measurable either by assumption or using Definition 2.1. Indicator functions are a very important example of measurable maps because as we saw in Figure 1, they are very natural and easy to integrate.

We have the following nice results:

Theorem 6.2 (Composition of measurable functions is measurable) *Let $(\Omega_1, \mathscr{A}_1), (\Omega_2, \mathscr{A}_2), (\Omega_3, \mathscr{A}_3)$ be three measurable spaces. Let $f : \Omega_1 \to \Omega_2, g : \Omega_2 \to \Omega_3$. If f, g are measurable then so is $g \circ f : \Omega_1 \to \Omega_3$.*

Proof: It is basically a copy of the simple proof that the composition of continuous functions is continuous (which uses the topological definition of continuity). ■

Theorem 6.3 (Properties of measurable functions into the real numbers) *Let* $(\Omega, \mathscr{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *denote our input and output measure spaces. (The input measure space is arbitrary and we have made a specific choice for the output measure space.) Then if* $f, g : \Sigma \to \mathbb{R}$ *are measurable, so are* $f+g, f-g, f\cdot g, |f|$.[3]

The proof of Theorem 6.3 was not given in the lecture, but it is probably in [Rud76].

7 The Lebesgue integral (including properties)

Key takeaways: *The Lebesgue integral is first defined for piecewise-constant or simple functions. Then we extend it to general functions by approximating a general function using simple functions.*

Our setup is the following: We have a measure space (Definition 4.3) $(X, \mathscr{A}, \mu)$ as our "input space." The "output space" is the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (see Definitions 3.3, 4.1). We also have a measurable map $f : X \to \mathbb{R}$ which we would like to integrate. Recall that f being measurable (Definition 6.1) means precisely that $f^{-1}(E) \in \mathscr{A}$ for all Borel sets[4] $E \subseteq \mathbb{R}$.

However, we are not yet prepared to handle general f, so we start by looking at the case where f is of the form (2), i.e., of the form $\chi_A : X \to \mathbb{R}$ for some $A \in \mathscr{A}$. This gives the following picture, where the red line denotes the function χ_A:

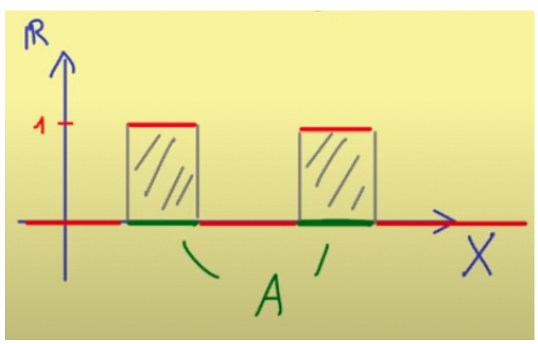

Figure 2: $\chi_A : X \to \mathbb{R}$

Since χ_A takes the value 1 on the set A and 0 elsewhere, a meaningful integral which, we will denote I, should by definition satisfy $I(\chi_A) = \mu(A)$. In other words, the integral of χ_A is just the measure of A.

We can now extend this by taking linear combinations of indicator functions:

Definition 7.1 (Simple function) *f is a simple function (also known as: step function, staircase*

[3] Note that we need the codomain of f, g to be $\mathbb{R}$ for something like $f+g$ to even be well-defined.
[4] A Borel set is just an element of the Borel σ-algebra.

function, piecewise-constant function) if it can be written in the form

$$f(x) = \sum_{i=1}^{n} c_i(x) \cdot \chi_{A_i}(x) \tag{3}$$

for some $A_1, \ldots, A_n \in \mathscr{A}$ and $c_1, \ldots, c_n \in \mathbb{R}$. Simple functions are measurable due to the fact that indicator functions are measurable (see Section 6) and Theorem 6.3.

An example:

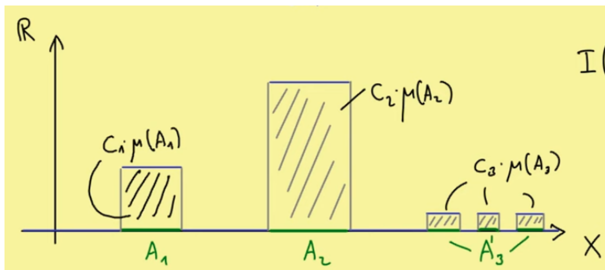

Figure 3: Simple function

Going back to our integral I, it is clear from Figure 3 and how we defined I for a single indicator function ($I(\chi_A) = \mu(A)$) that we should extend I to a simple function f of the form (3) in the following way:

$$I(f) := \sum_{i=1}^{n} c_i \cdot \mu(A_i). \tag{4}$$

However, (4) is problematic in the case where, e.g., $\mu(A_1) = \infty, \mu(A_2) = \infty$, and $c_1 = 1, c_2 = -1$. Then $c_1\mu(A_1) + c_2\mu(A_2) = \infty - \infty$, which is undefined. To solve this, we enforce an additional restriction that $c_i \geq 0$ for all i. (Note that $\mu(A_i) \geq 0$ for all i by Definition 4.2, so this indeed ensures that we will only add positive infinities, which is well-defined: $\infty + \infty := \infty$.) This leads to the following definition:

$$\mathcal{S}^+ := \{f : X \to \mathbb{R} : f \text{ is a simple function}, f \geq 0\} \tag{5}$$

Note that you can replace the condition "f is a simple function" in (5) with the equivalent condition "f is measurable and takes on only finitely many values." It is easy to see that the two are equivalent, although maybe there is one subtlety when you try to show that the latter implies the former. To take a measurable function that takes on only finitely many values and represent it in the form (3), one would naturally want to set $A_1 = f^{-1}(\{b_1\}), \ldots, A_n = f^{-1}(\{b_n\})$, where $b_1, \ldots, b_n$ are the finitely many values in the image of f. To verify that f is a simple function per Definition 7.1, we need $A_1, \ldots, A_n$ to be measurable. Since f is measurable, it is sufficient by Definition 6.1 for $\{b_1\}, \ldots, \{b_n\}$ to be measurable (or equivalently, they need to be elements of $\mathcal{B}(\mathbb{R})$). Luckily this is the case since these are closed sets (see the note at the end of Definition 3.3).

Note that $\mathcal{S}^+$ is almost a vector space since it is closed under addition of its elements (vector addition), but it is not quite a vector space because it is only closed under multiplication by nonnegative scalars. Thus, you should think of it as a half-space.

With this, we can define the Lebesgue integral for elements of $\mathcal{S}^+$:

Definition 7.2 (Lebesgue integral for nonnegative simple functions) *Let $f \in \mathcal{S}^+$ and choose a representation of the form (3) where $c_i \geq 0$ for all i. Then the Lebesgue integral with respect to the measure μ is*

$$\int_X f(x)\,\mathrm{d}\mu(x) := \int_X f\,\mathrm{d}\mu := I(f) := \sum_{i=1}^{n} c_i \cdot \mu(A_i) \in [0, \infty]. \tag{6}$$

(In other words, the first three entries (6) are different notations for the same thing.)

Note that for this definition to be rigorous, one must show that you get the same value for (6) regardless of what representation for f of the form (3) you pick. This can be proven, although it should be intuitive from Figure 3: If you split one of the A_i's into pieces and add up the measures of the pieces, you should still get $\mu(A_i)$ in the end due to Condition 2 of Definition 4.2.

Before we extend Definition 7.2 to more general functions, we first give some properties:

Theorem 7.3 (Properties of the Lebesgue integral for nonnegative simple functions) *Let $f, g \in \mathcal{S}^+$. Then*

1. *$I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for $\alpha, \beta \geq 0$. (In other words, the Lebesgue integral is almost linear, although we have to use nonnegative scalars. Actually, it is not really necessary for these scalars to be nonnegative—see Theorem 11.23 in [Rud76]. I'm just following the lecture here which so far is sticking to nonnegative functions for simplicity.)*

2. *$f \leq g \Rightarrow I(f) \leq I(g)$ (monotonicity).*

Both of these properties can be proven very easily using (6). Furthermore, they can be extended more generally to Lebesgue-integrable[5] functions—see Theorems 11.23 and 11.29 in [Rud76].

Finally, we are ready to extend to more general (albeit still nonnegative) functions:

Definition 7.4 (Lebesgue integral for nonnegative functions) *Let $f : X \to [0, \infty)$ be a measurable function. Then the Lebesgue integral of f with respect to a measure μ is*

$$\int_X f(x)\,\mathrm{d}\mu(x) := \int_X f\,\mathrm{d}\mu := I(f) := \sup\left\{I(h) : h \in \mathcal{S}^+, h \leq f\right\}, \tag{7}$$

where $h \leq f$ means that h is pointwise at most f (i.e., $h(x) \leq f(x)$ for all $x \in X$).

f is called μ-integrable if $\int_X f\,d\mu < \infty$.

This definition can be extended of course to the case where f isn't nonnegative; you simply divide f into its negative and positive parts and integrate them separately—see Definition 11.22 in [Rud76].

[5] "Lebesgue-integral" requires its own definition. See Definition 11.22 and particularly the discussion at the top of page 315 in [Rud76].

Finally, one important notational point is that if μ is the Lebesgue measure (which of course it needn't be; we won't even formally define the Lebesgue measure until Sections 13 and 14, but I'm adding this note in retroactively), then we often use the shorthand notation $\int_{\mathbb{R}} f \, \mathrm{d}x$ as opposed to, e.g., $\int_{\mathbb{R}} f(x) \, \mathrm{d}\mu(x)$. But to reemphasize, the Lebesgue integral in some sense has nothing to do with the Lebesgue measure! The Lebesgue measure is just a particular measure you can Lebesgue-integrate with respect to.

Essentially, you should think of (7) as obtaining finer and finer approximations of f from below using simple functions. Here is a picture, where the blue line denotes f:

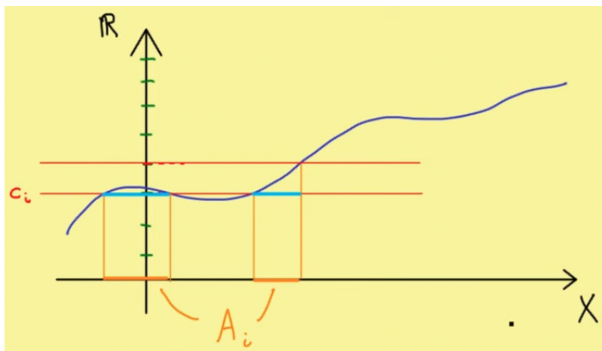

Figure 4: Image for Definition 7.4

The idea is you can split the y-axis (which is just $\mathbb{R}$) into finer and finer intervals. Then you can draw horizontal lines across the picture for each interval—this was done in Figure 4 for the interval $[c_i, c_{i+1}]$, although c_{i+1} was not labeled. Then you can look across and consider the preimage of $[c_i, c_{i+1}]$, aka $A_i := f^{-1}([c_i, c_{i+1}])$. Because $[c_i, c_{i+1}] \in \mathcal{B}(\mathbb{R})$ (it is a closed set) and f is measurable, it follows that A_i is measurable. Then $c_i \cdot \chi_{A_i}$ will become one of the terms in your approximation for f: $h = \sum_{j=1}^n c_j \cdot \chi_{A_j}$. Use finer and finer intervals on the y-axis, and you get a better approximation of f. (The fact that we choose c_i to be the coefficient of χ_{A_i} for each i ensures that $h \leq f$, as required by (7).) The fact that we are dividing the y-axis into intervals instead of the x-axis is why the Lebesgue integral is sometimes thought of as horizontal integration.

8 Properties of the Lebesgue integral and the monotone convergence theorem

Key takeaways: *The Lebesgue integral satisfies many natural properties. The monotone convergence theorem is an important result which showcases the power of the Lebesgue integral over the Riemann integral.*

We first give some properties of the Lebesgue integral. Following the lecture, we restrict our attention to nonnegative functions, but of course all of this can be extended to functions that aren't nonnegative.

Theorem 8.1 (Properties of the Lebesgue integral) *Let $f, g : X \to [0, \infty)$ be measurable. Then*

1. *$f = g$ μ-a.e. implies $\int_X f\, d\mu = \int_X g\, d\mu$, where μ-a.e. is read "μ-almost everywhere." What this means is that f, g only differ (or more generally, don't satisfy the property which appears before "μ-a.e.") on a set of measure 0:*

$$\mu\left(\{x \in X : f(x) \neq g(x)\}\right) = 0.$$

This property is analogous to the fact that if you take a continuous function f and change its value at one point and take the Riemann integral of the result, you still get the Riemann integral of f. But for the Lebesgue integral, you are allowed to change f however you want on a set of measure 0, and the integral stays the same.

2. *$f \leq g$ μ-a.e. implies $\int_X f\, d\mu \leq \int_X g\, d\mu$ (monotonicity of the Lebesgue integral).*

3. *$f = 0$ μ-a.e. if and only if $\int_X f\, d\mu = 0$. (Note that this relies on the nonnegativity of f so that we don't get cancellations.)*

He only proves the second result from Theorem 8.1 in the lecture, but the proof is highly illustrative and I would recommend watching/rewatching it. The core of the proof is the following very useful lemma:

Lemma 8.2 (Changing simple function on a set of measure zero doesn't change integral) *Let $X = \widetilde{X} \cup \widetilde{X}^c$ where $\widetilde{X}$ is measurable (implying that $\widetilde{X}^c$ is measurable too by Condition 2 of Definition 2.1). Also, suppose that $\mu(\widetilde{X}^c) = 0$.*

Now let $h : X \to [0, \infty)$ be a simple function, and let $\tilde{h} : X \to [0, \infty)$ be another simple function which only differs with h on $\widetilde{X}^c$ (a set of measure 0). (In the lecture, he has h take on some arbitrary value $a \geq 0$ on $\widetilde{X}^c$ for simplicity, although it is easy to see $\tilde{h}$ could take on finitely many different values on $\widetilde{X}^c$, and you would get the same result.) Then $I(h) = I(\tilde{h})$.

Proof: You first pick a useful representation for h in the form (3) using the fact that the image of h contains finitely many values. You can then obtain an analogous representation for $\tilde{h}$ by taking intersections with the set $\widetilde{X}$ and handling the behavior of $\tilde{h}$ on $\widetilde{X}^c$ separately. Then, one can verify that $I(h) = I(\tilde{h})$ using properties of the measure μ (Definition 4.2). ∎

One can combine Lemma 8.2 with Definition 7.4 to prove the second result from Theorem 8.1 (and the first and third one too).

We now give an important theorem:

Theorem 8.3 (Monotone convergence theorem) *Let $(X, \mathscr{A}, \mu)$ be a measure space with $f_n : X \to [0, \infty)$ measurable for all $n \in \mathbb{N}$. In addition, we have another function $f : X \to [0, \infty)$ which is also measurable. Also, $f_1 \leq f_2 \leq f_3 \leq \ldots$ holds μ-a.e.*[6] *(the sequence is monotonically increasing). And, $\lim_{n\to\infty} f_n(x) = f(x)$ holds μ-a.e. $(x \in X)$.*[7]

[6] In other words, the set of points which doesn't satisfy this property has measure 0.

[7] Since we have an x in the condition $\lim_{n\to\infty} f_n(x) = f(x)$, we can add "$(x \in X)$" after "$\mu$-a.e." to specify that this property holds μ-a.e. with respect to the set X. (He made this point in the lecture so I'm including it here, but this isn't saying anything surprising.)

The monotone convergence theorem says that we can push the limit inside the integral:

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

Convergence theorems[8] like Theorem 8.3 are where the Lebesgue integral really shines over the Riemann integral. Apparently it is also known as Beppo Levi's lemma.

Also, he begins the next lecture (but I'm putting this note here for convenience) by explaining that there is a slightly more general version of Theorem 8.3 where $f_n : X \to [0, \infty]$, meaning f_n can actually take on the value ∞. This formulation requires you to extend some previous definitions to work with the value ∞ from the extended real number line, but nothing unusual happens. The advantage of this more general formulation being that Theorem 8.3 can be stated more compactly (you don't need to introduce f—watch the lecture corresponding to Section 9 to see the details). See also the beginning of Section 10 for more on this.

9 Monotone convergence theorem: proof and applications

In this lecture, he covers the proof of Theorem 8.3:

Proof of Theorem 8.3: First he shows

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

using the second result from Theorem 8.1; this requires little effort. To prove the more difficult direction:

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

one must use simple functions and go back to the definition of the Lebesgue integral in terms of the supremum of simple functions (Definition 7.4). ■

The following is an important application of Theorem 8.3:

Theorem 9.1 (The Lebesgue integral and series) *Let $(g_n)_{n\in\mathbb{N}}$ where $g_n : X \to [0, \infty]$ are measurable for all n. Then $\sum_{n=1}^{\infty} g_n : X \to [0, \infty]$ is well-defined[9] and measurable. Then since the partial sums are monotonically increasing, Theorem 8.3 implies that*

$$\int_X \sum_{i=1}^{n} g_n \, \mathrm{d}\mu = \sum_{i=1}^{n} \int_X g_n \, \mathrm{d}\mu.$$

Thus, as long as our functions are nonnegative, Theorem 8.3 allows us to switch the order of the sum and the integral!

[8] Convergence theorems tell you when you can pull a limit into an integral.

[9] Notice that $g_n : X \to [0, \infty]$ instead of $g_n : X \to [0, \infty)$; in other words, g_n can take the value ∞. See the note

10 Fatou's lemma

Key takeaways: *Fatou's lemma is only a "half convergence theorem" because you get an inequality instead of an equality, unlike Theorem 8.3. However, it is still powerful because of how few conditions are required to apply it.*

We now go over some other important convergence theorems (and related results). From now on, we will allow our functions to map into $[0, \infty]$ instead of the more restrictive $[0, \infty)$. See the note at the end of Section 8 for more on this. Basically, it is useful because if we have functions $(f_n)_{n \in \mathbb{N}}$ such that $f : X \to [0, \infty]$, then $\lim_{n\to\infty} f_n : X \to [0, \infty]$ is always well-defined[10]; it just may take on the value ∞ at certain points. Whereas if we confine ourselves to $f : X \to [0, \infty)$, then $\lim_{n\to\infty} f_n$ may not be well-defined if we try to force it to be finite at every point. This is why in Theorem 8.3, we had to introduce $f : X \to [0, \infty)$ and specify (essentially) that $\lim_{n\to\infty} f_n = f$; this was a mechanism to enforce the fact that $\lim_{n\to\infty} f_n$ should only take on finite values.

One could of course still restrict each f_n to take finite values and just allow $\lim_{n\to\infty} f_n$ take on the value ∞ so that it always exists, but if we are going to do this we may as well let each f_n take on the value ∞ in the first place since it is a more general scenario.

Theorem 10.1 (Fatou's lemma) *Let $(X, \mathscr{A}, \mu)$ be a measure space. Let $f_n : X \to [0, \infty]$ be measurable for all $n \in \mathbb{N}$. One can show that $\liminf_{n\to\infty} f_n$ is measurable.*[11]

Then

$$\int_X \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu. \tag{8}$$

Recall that $\liminf$ *is the limit inferior. Formally,* $\liminf_{n\to\infty} f_n : X \to [0, \infty]$ *is defined via*

$$\liminf_{n\to\infty} f_n(x) := \lim_{n\to\infty} \left(\inf_{k \geq n} f_k(x) \right) \in [0, \infty]$$

for all $x \in X$. The fact that $\liminf_{n\to\infty} f_n$ is measurable follows from the fact that the infimum of measurable functions is measurable and the limit of measurable functions is measurable.

So really, this isn't quite a convergence theorem since we only get an inequality. However, the power of Theorem 10.1 arises from the fact that very few conditions are required to apply it. (It doesn't require monotonicity unlike Theorem 8.3.) Note also that if we happen to know that f_n converges, then we can replace $\liminf_{n\to\infty} f_n$ on the left side of (8) with just $\lim_{n\to\infty} f_n$. For example, this is used in the proof of Theorem 11.2.

Proof Theorem 10.1: The proof is a fairly simple application of Theorem 8.3. The key observation being that if you define $g_n(x) = \inf_{k \geq n} f_k(x)$, then clearly $g_1 \leq g_2 \leq g_3 \leq \ldots$. Thus,

---

at the end of Section 8 for more on this. The point being, this means that we don't have to worry about the case where $\sum_{n=1}^{\infty} g_n$ diverges at certain points; it is well-defined regardless.

[10] Well, this isn't quite accurate—$\lim_{n\to\infty} f_n$ may also be undefined due to cycling or similar behavior. However, this can't occur in the scenario of Theorem 8.3 since the sequence is monotonically increasing. It also can't occur in Theorem 10.1 because there we use $\liminf$, which always exists for sequences of real numbers.

[11] Also, note that $\liminf$ always exists for sequences of real numbers. Thus, $\liminf_{n\to\infty} f_n$ is well-defined (in a

$(g_n)_{n\in\mathbb{N}}$ satisfies the conditions of Theorem 8.3, so we can apply it to get our result. (It takes a few more steps including an application of the monotonicity of the Lebesgue integral, aka the second result from Theorem 8.1, but this is the basic idea.) ■

11 Lebesgue's dominated convergence theorem

Key takeaways: *Lebesgue's dominated convergence theorem is another very powerful and widely applicable convergence theorem for Lebesgue integration.*

We now go over another convergence theorem. (Recall that convergence theorems tell you when you can pull a limit into an integral.) Our setup is the following: Let $(X, \mathscr{A}, \mu)$ be a measure space. We define

Definition 11.1 (Lebesgue-integrable functions) *We let $\mathscr{L}(\mu)$[12] denote the set of Lebesgue-integrable functions with respect to μ, i.e.,*

$$\mathscr{L}(\mu) := \mathscr{L}^1(\mu) := \{f : X \to \mathbb{R} : f \textit{ is measurable}, \int_X |f| \,\mathrm{d}\mu < \infty\}. \tag{9}$$

$\mathscr{L}^1(\mu)$ is pronounced "el-one space," and comes from the fact that we are implicitly raising $|f|$ to the first power in the integral on the right side of (9). We will later extend this definition and define $\mathscr{L}^p(\mu)$ for a general $p \in \mathbb{N}$.

Note that $\mathscr{L}(\mu)$ can contain functions which aren't nonnegative, but $|f|$ is of course always nonnegative, so (9) is well-defined using what we have built up so far. (The point being that we haven't actually extended Definition 7.4 to functions that aren't nonnegative, although this can be done easily as noted in the definition.)

In fact, at this point he goes ahead and extends Definition 7.4 to functions that aren't nonnegative. As mentioned previously, the idea is to just write $f = f^+ - f^-$ where $f^+, f^- \geq 0$. f^+ is the nonnegative part[13] of f and f^- is the nonpositive part of f. Then we define

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

Although, we need to be careful as it would be problematic if both $\int_X f^+ \,\mathrm{d}\mu, \int_X f^- \,\mathrm{d}\mu$ were ∞. The lecture resolves this by requiring that $f \in \mathscr{L}(\mu)$, since $\int_X |f| \,\mathrm{d}\mu < \infty\}$ of course implies that both $\int_X f^+ \,\mathrm{d}\mu, \int_X f^- \,\mathrm{d}\mu$ are finite. [Rud76] resolves this in basically the same way, explicitly requiring that both $\int_X f^+ \,\mathrm{d}\mu, \int_X f^- \,\mathrm{d}\mu$ are finite—see Definition 11.22. There is also a nice note at the end of Definition 11.22—the typical convention is to require that both $\int_X f^+ \,\mathrm{d}\mu, \int_X f^- \,\mathrm{d}\mu$ are finite, even though $\int_X f^+ \,\mathrm{d}\mu - \int_X f^- \,\mathrm{d}\mu$ is well-defined if just one of them if finite. This being because we are mainly interested in the case where both are finite, although "in some cases it is desirable to deal with the more general situation."

Then:

pointwise manner).

[12]The same notation is used in [Rud76]—see Definition 11.22.

[13]I.e., f^+ is takes on the same value as f at x where $f(x) \geq 0$. At x where $f(x) < 0$, we have that $f^+(x) = 0$.

Theorem 11.2 (Lebesgue's dominated convergence theorem) *Let $f_n : X \to \mathbb{R}$ be measurable for all $n \in \mathbb{N}$. Let $(f_n)_{n\in\mathbb{N}}$ convergence pointwise*[14] *to f μ-a.e. (In other words, the set $\{x \in X : \lim_{n\to\infty} f_n(x) \neq f(x)\}$ has measure zero.) Also, let*[15] *$|f_n| \leq g$ for all $n \in \mathbb{N}$, for some $g : X \to \mathbb{R}$ such that $g \in \mathscr{L}^1(\mu)$. (Such a function g is called an "integrable majorant." This is where the "dominated" comes from in "Lebesgue's dominated convergence theorem.")*

Then all of the functions $f_1, f_2, f_3, \dots \in \mathscr{L}^1(\mu)$ and $f \in \mathscr{L}^1(\mu)$. Also,

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu. \tag{10}$$

Of course, (10) is where the "convergence theorem" part comes from, since convergence theorems say when you can pull a limit inside an integral.

12 Proof of Lebesgue's dominated convergence theorem

In this lecture Theorem 11.2 is proven:

Proof of Theorem 11.2: The first two implications ($f_1, f_2, \dots \in \mathscr{L}_1(\mu)$ and $f \in \mathscr{L}^1(\mu)$) follow by taking the integral of both sides of $|f_n| \leq g$ and using the monotonicity of the Lebesgue integral (second result of Theorem 8.1). (The key point being that the integral over g is finite due to it being in $\mathscr{L}_1(\mu)$ (Definition 11.1).)

Proving (10) takes a bit more work—it is a beautiful (and rather magical) application of Theorem 10.1. He actually first proves the stronger result

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

He then uses the triangle inequality for integrals: For a measurable function h, we have that

$$\left| \int_X h \, \mathrm{d}\mu \right| \leq \int_X |h| \, \mathrm{d}\mu.$$

■

13 Carathéodory's extension theorem

Key takeaways: *Carathéodory's extension theorem can be used to prove the existence and uniqueness of the Lebesgue measure!*

In this lecture, we move past convergence theorems and cover a result which doesn't require knowledge of the lectures immediately preceding this one. But first, some definitions:

Definition 13.1 (Semiring of sets) *Let X be a set, and let $\mathscr{A} \subseteq \mathcal{P}(X)$. (Even though typically used $\mathscr{A}$ to denote a σ-algebra in the past, here it is just an arbitrary subset of the power set.) $\mathscr{A}$ is a semiring of sets if*

[14] Pointwise convergence means $\lim_{n\to\infty} f_n(x) = f(x)$

[15] Of course, the inequality $|f_n| \leq g$ denotes that g dominates $|f_n|$ in a pointwise manner.

1. $\emptyset \in \mathscr{A}$ *(same as for a* σ*-algebra),*

2. $A, B \in \mathscr{A} \Rightarrow A \cap B \in \mathscr{A}$*, and*

3. *for* $A, B \in \mathscr{A}$*, there are pairwise disjoint sets* $S_1, \ldots, S_n \in \mathscr{A}$ *such that*

$$\bigcup_{j=1}^{n} S_j = A \setminus B.$$

Note that every σ*-algebra is a semiring of sets since you can write* $A \setminus B = A \cap B^c$ *(and* σ*-algebras are closed under complementation and taking intersections).*

Example: The most important example of a semiring of sets is

$$\mathscr{A} := \{[a, b) : a, b \in \mathbb{R}, a \leq b\}. \tag{11}$$

He proves that this is a semiring in the lecture. Note that this is not a σ-algebra since $\mathbb{R} \in \mathscr{A}$. But it turns out that $\sigma(\mathscr{A}) = \mathcal{B}(\mathbb{R})$ (recall Definitions 3.2, 3.3). In other words, $\mathscr{A}$ generates the Borel σ-algebra on $\mathbb{R}$.

Another definition:

Definition 13.2 (Pre-measure) *A pre-measure is almost a measure, but it is defined on a semiring instead of a* σ*-algebra. Let* $\mathscr{A}$ *be a semiring of sets. Then* $\mu : \mathscr{A} \to [0, \infty]$ *is a pre-measure if*

1. $\mu(\emptyset) = 0$,

2. $\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$ *for* $A_j \in \mathscr{A}$ *if* $A_i \cap A_j = \emptyset$ *for* $i \neq j$ *and* $\bigcup_{j=1}^{\infty} A_j \in \mathscr{A}$*. (Note that this is equivalent to* σ*-additivity from Definition 4.2, except we include the additional condition that* $\bigcup_{j=1}^{\infty} A_j \in \mathscr{A}$*. It is necessary to include this additional condition since unlike a* σ*-algebra, a* σ*-ring is not necessarily closed under a countable union.)*

Now for the main result:

Theorem 13.3 (Carathéodory's extension theorem) *Let* X *be a set and let* $\mathscr{A} \subseteq \mathcal{P}(X)$ *be a semiring of sets. Let* $\mu : \mathscr{A} \to [0, \infty]$ *be a pre-measure. Then:*

1. μ *has an extension, which we will call* $\tilde{\mu} : \sigma(\mathscr{A}) \to [0, \infty]$.[16] $\tilde{\mu}$ *is an "ordinary" measure (Definition 4.2). Furthermore,* $\tilde{\mu}$ *is an extension in the sense that it agrees with* μ *on* $\mathscr{A}$*:* $\mu(A) = \tilde{\mu}(A)$ *for all* $A \in \mathscr{A}$*.*

2. *We say that a measure or pre-measure* μ *is* σ*-finite if there is a sequence* (S_j) *with* $S_j \in A$*,* $\bigcup_{j=1}^{\infty} = X$*, and* $\mu(S_j) < \infty$*. Then an additional result contained within this theorem is the following: If* μ *(from the setup of this theorem) is* σ*-finite, then the extension* $\tilde{\mu}$ *from the previous result is unique. Furthermore,* $\tilde{\mu}$ *is also* σ*-finite.*

[16] The notation $\sigma(\mathscr{A})$ comes from Definition 3.2.

You should think of the σ-finite property as being a weaker version of a finite measure. A measure is finite if $\mu(X)$ is finite. A measure is σ-finite if you can approximate X using finite-measure sets (and you only need countably many of them).

Application: Let the semiring $\mathscr{A}$ be defined as in (11). Define the pre-measure $\mu : \mathscr{A} \to [0, \infty]$ via $\mu([a, b)) = b - a$. Then Theorem 13.3 says that there is a unique extension to $\sigma(\mathscr{A}) = \mathcal{B}(\mathbb{R})$. And, the measure which results from this extension is precisely the Lebesgue measure! Thus, a corollary of Theorem 13.3 is the existence and uniqueness of the Lebesgue measure.

14 Lebesgue-Stieltjes measures

Key takeaways: *The Lebesgue-Stieltjes measure can be formally and uniquely defined by first constructing a pre-measure which agrees with our intuitive notion of "weighted length." Then Theorem 13.3 can be applied to extract the associated Lebesgue-Stieltjes measure.*

Quoting Wikipedia: "Lebesgue-Stieltjes integration generalizes Riemann-Stieltjes and Lebesgue integration, preserving the many advantages of the former in a more general measure-theoretic framework. The Lebesgue-Stieltjes integral is the ordinary Lebesgue integral with respect to a measure known as the Lebesgue-Stieltjes measure, which may be associated to any function of bounded variation[17] on the real line." In this lecture we define Lebesgue-Stieltjes measures.

Toward this end, let $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ be monotonically increasing (equivalently, nondecreasing.) Note that it can be constant at some parts and/or discontinuous. (He draws an example in the lecture.) $\mathcal{F}$ is our "weight function" (analogous to α in Definition 6.2 in [Rud76]) which "scales" (although not exactly) the lengths of intervals. Originally the length of an interval $[a, b)$ is of course $b - a$, but under the weight function $\mathcal{F}$, the length or measure of $[a, b)$, denoted $\mu_F([a, b))$ is defined as

$$\mu_F([a, b)) := \mathcal{F}(b^-) - \mathcal{F}(a^-) \tag{12}$$

where

$$\mathcal{F}(c^-) := \lim_{\varepsilon \to 0^+} \mathcal{F}(c - \varepsilon) = \lim_{x \to c^-} \mathcal{F}(x).$$

for any $c \in \mathbb{R}$. (Recall the notation for a one-sided limit.) In other words, $\mu_F([a, b))$ is the value at which $\mathcal{F}$ approaches b from the left minus the value at which $\mathcal{F}$ approaches a from the left. Thus, you can essentially think of it as being $\mathcal{F}(b) - \mathcal{F}(a)$, with some additional changes made to deal with discontinuities in $\mathcal{F}$ at the endpoints of the interval. I'm not going to go over the intuition here as to why (12) is the correct way to deal with discontinuities in $\mathcal{F}$, but I would highly recommend watching the beginning of the lecture corresponding to this section where it is laid out very nicely. (He also notes that one can obtain an alternative definition for $\mu_F((b, a]$, where we have flipped which sides of the input interval are open and closed. However, this won't matter for our purposes.)

[17] In [Rud76], the Riemann-Stieltjes integral is defined with respect to a monotonically increasing function (typically denoted α)—see Definition 6.2. From some quick searching, it appears that functions of bounded variation generalize monotonically increasing (and bounded) functions, so this is all to say that apparently it is possible to define the Riemann-Stieltjes integral in slightly more generality than [Rud76] does. Although in these notes we will follow the same approach as [Rud76] and just have our "weight" function be monotonically increasing.

Recall from Section 13 that

$$\mathscr{A} := ([a, b) : a, b \in \mathbb{R}, a \leq b) \tag{13}$$

is a semiring of sets. Furthermore, one can check that μ_F is a σ-finite pre-measure with respect to the semiring of sets $\mathscr{A}$. Thus, we can apply Theorem 13.3 to conclude that there exists a unique extension measure $\mu_F : \mathcal{B}(\mathbb{R}) \to [0, \infty]$. (We are abusing notation here and also denoting the extension μ_F.) Note that the domain of μ_F (the measure) is $\mathcal{B}(\mathbb{R})$ because as mentioned in Section 13, $\sigma(\mathscr{A}) = \mathcal{B}(\mathbb{R})$. μ_F is called the *Lebesgue-Stieltjes measure for $\mathcal{F}$:*

Definition 14.1 (Lebesgue-Stieltjes measure) *Let $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ be monotonically increasing. Let $\mathscr{A}$ be the semiring of sets defined in (13), and let μ_F be the pre-measure on $\mathscr{A}$ defined through (12). The unique extension $\mu_F : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ (we abuse notation and also denote in μ_F) arising from an application of Theorem 13.3 is known as the Lebesgue-Stieltjes measure for $\mathcal{F}$.*

When $\mathcal{F}(x) = x$, meaning $\mu_F([a, b)) = b - a$, we recover the ordinary Lebesgue measure.

Examples: Suppose $\mathcal{F}(x) = 1$. Then $\mu_F([a, b)) = 0$, so we recover the zero measure. Now suppose

$$\mathcal{F}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

Then note that $\mu_F([-\varepsilon, \varepsilon)) = 1$ for all $\varepsilon > 0$. One can verify that the measure μ_F exactly satisfies the properties of the Dirac measure (see (1)) at 0, δ_0, and thus must actually be δ_0 due to the uniqueness aspect of Definition 14.1.

Now let's look at a very general example. Let $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ be monotonically increasing and continuously differentiable (meaning the derivative exists and is continuous). This implies that the derivative $\mathcal{F}' : \mathbb{R} \to [0, \infty)$ is nonnegative. Since $\mathcal{F}$ is continuous, we don't need to worry about one-sided limits and we have that

$$\mu_F([a, b)) = \mathcal{F}(b) - \mathcal{F}(a) = \int_a^b \mathcal{F}'(x) \, \mathrm{d}x \tag{14}$$

through application of the fundamental theorem of calculus. (The integral on the right side of (14) is with respect to the ordinary Lebesgue measure.) Then μ_F (when viewed as only a pre-measure) extends to a measure on $\mathcal{B}(\mathbb{R})$ defined via (again abusing notation and using the same symbol μ_F):

$$\mu_F : A \mapsto \int_A \mathcal{F}'(x) \, \mathrm{d}x$$

for any Borel set $A \in \mathcal{B}(\mathbb{R})$. (Recall that Borel sets are just elements of $\mathcal{B}(\mathbb{R})$, or more generally elements of $\mathcal{B}(X)$ for some set X.) Once again, we know such a measure is unique due to Definition 14.1 (and really Theorem 13.3 behind the scenes). This is a very general measure, and the function $\mathcal{F}'(x)$ is often called a density function. (I believe probability density functions are a special case!)

15 The Radon-Nikodym theorem and Lebesgue's decomposition theorem

Key takeaways: *Lebesgue's decomposition theorem allows us to decompose many important measures into a sum of two measures which satisfy nice properties. The Radon-Nikodym theorem allows us to express a certain class of important measures as a Lebesgue integral.*

In this lecture we go over two useful theorems with many applications, even outside of measure theory. For our basic setup, we have a measure space $(X, \mathscr{A}, \lambda)$. (We use λ instead of the typical μ to denote our measure.) It is good to think of it as $X = \mathbb{R}$, $\mathscr{A} = \mathcal{B}(\mathbb{R})$, and λ is the Lebesgue measure. (Recall that the Lebesgue measure λ is the unique measure such that $\lambda([a, b)) = b - a$.) In fact, we will go ahead and make this assumption for simplicity, although you can extend everything in this section to the more general case. I.e., we assume that $X = \mathbb{R}$, $\mathscr{A} = \mathcal{B}(\mathbb{R})$, and λ is the Lebesgue measure.

You should think of λ as being a "reference measure," and there is another measure, $\mu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$, which we are actually interested in.

Definition 15.1 (Absolutely continuous measure) *We say that μ is absolutely continuous (with respect to the Lebesgue measure λ) if $\lambda(A) = 0$ implies $\mu(A) = 0$ for all $A \in \mathcal{B}(\mathbb{R})$. One writes $\mu \ll \lambda$ to denote that μ is absolutely continuous with respect to λ.*

You can think of this intuitively as, "The measure μ is not finer than the Lebesgue measure λ." In other words, μ does not assign a nonzero value to something of zero measure with respect to the Lebesgue measure, although μ can assign more sets the value 0 than the Lebesgue measure.

Examples: It is always true of course that $\lambda \ll \lambda$. The zero measure is of course also absolutely continuous with respect to any measure.

Definition 15.2 (Singular measure) *μ is singular (with respect to λ, although we usually omit this if the Lebesgue measure is our reference measure) if there is $N \in \mathcal{B}(\mathbb{R})$ such that $\lambda(N) = 0$ and $\mu(N^c) = 0$. (Of course, $N^c := \mathbb{R} \setminus \mathbb{N}$.) We denote this $\mu \perp \lambda$.*

Intuitively, this says that the measure μ and the Lebesgue measure λ are in some sense disjoint or orthogonal. μ puts all of its mass in N, and λ puts all of its mass in N^c. (This is true in a formal sense due to the properties of σ-algebras and measures.)

Example: The Dirac measure at 0 (see (1)), δ_0, is singular with respect to the Lebesgue measure λ. (Recall that δ_0 assigns the value 1 to any set that contains 0 and assigns 0 to every other set.) To see this, choose $N = \{0\}$ and verify that such an N satisfies the conditions of Definition 15.2.

We can now state the two theorems:

Theorem 15.3 (Lebesgue's decomposition theorem and the Radon-Nikodym theorem) *Let $\mu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ be a σ-finite measure (see Theorem 13.3). Let λ denote the Lebesgue measure (our "reference measure"). Then:*

1. **Lebesgue's decomposition theorem:** *There are two measures (uniquely determined) $\mu_{ac}, \mu_s : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ where $\mu = \mu_{ac} + \mu_s$, $\mu_{ac} \ll \lambda$, and $\mu_s \perp \lambda$. In other words, we*

can uniquely decompose any measure μ into one absolutely continuous measure and one singular measure.

2. **Radon-Nikodym theorem:** *There is a measurable map $h : \mathbb{R} \to [0, \infty)$ with*

$$\mu_{ac}(A) = \int_A h \, \mathrm{d}\lambda \tag{15}$$

for all $A \in \mathcal{B}(\mathbb{R})$. ($\mu_{ac}$ is defined as in the first part.) h is known as a density function.

This is useful because it says that instead of working with an abstract absolutely continuous measure μ_{ac}, we can instead deal with a more concrete object: The Lebesgue integral of some measurable function. Also, note that this implies that it is actually possible to express any absolutely continuous measure in the form (15), since one could pick μ to be any absolutely continuous measure, in which case the uniqueness result of the first part says that $\mu_{ac} = \mu$ and μ_s is the zero measure. (He mentioned this in the lecture, but I'm not sure why he didn't just present the result this way.)

It is possible to extend these theorems to the case where λ is not necessarily the Lebesgue measure but is instead any σ-finite measure.

16 The image measure and the substitution formula

Key takeaways: *The image measure allows us to push a measure defined on one space to another space using a measurable map between the spaces. The substitution rule for Lebesgue integrals generalizes the substitution/change of variables rule for Riemannian integrals.*

The image measure is often called the pushforward measure, and the substitution formula is often called the change of variables formula.

We first define the image measure, which is basically a way of defining a measure on one space using a measure defined on another space and a measurable function between the spaces. Our setup is the following:

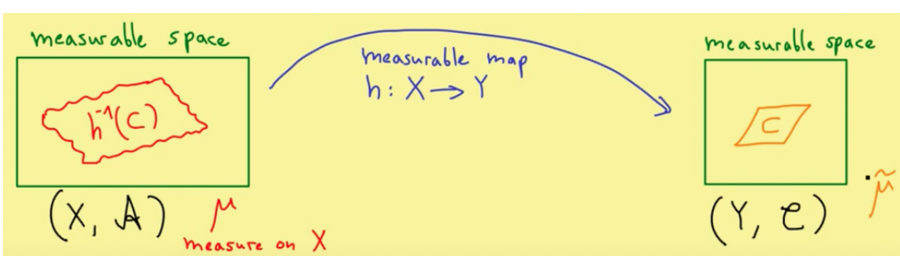

Figure 5: Image measure setup

As you can see, we have two measurable spaces (Definition 4.1): $(X, \mathscr{A})$ and $(Y, \mathscr{C})$. In fact, the former measurable space is actually a measure space (Definition 4.3) since it is equipped with the measure μ. We also have a measurable map $h : X \to Y$. Our objective is to define a natural measure $\tilde{\mu}$ on $(Y, \mathscr{C})$. With this in mind:

Definition 16.1 (Image measure or pushforward measure) *With the above setup, the image or pushforward measure $\tilde{\mu} : \mathscr{C} \to [0, \infty]$ is defined via*

$$\tilde{\mu}(C) = \mu(h^{-1}(C))$$

for all $C \in \mathscr{C}$.

Recall that $h^{-1}(C)$ denotes the preimage of C, which is measurable due to the fact that h is a measurable function. You should think of h as "pushing the measure on $(X, \mathscr{A})$ forward to $(Y, \mathscr{C})$," hence why the image measure is also known as the pushforward measure.

Typically instead of using the notation $\tilde{\mu}$ to denote the image measure, we will use $h_*\mu$ or $\mu \circ h^{-1}$ (the latter of which is just the definition).

Then:

Theorem 16.2 (Substitution or change of variables formula) *With the setup as above, define a function $g : Y \to \mathbb{R}$ which is measurable with respect to the image measure $h_*\mu$ defined in Definition 16.1. Then:*

$$\int_Y g \, \mathrm{d}(h_*\mu) = \int_X g \circ h \, \mathrm{d}\mu.$$

He recommends using the notation $\mu \circ h^{-1}$ instead of $h_*\mu$ and writing

$$\int_Y g(y) \, \mathrm{d}(\mu \circ h^{-1})(y) = \int_X g(h(x)) \, \mathrm{d}\mu(x) \tag{16}$$

to remember the formula more easily. (Recall that introducing a variable is perfectly fine per the alternate notation given in Definitions 7.2, 7.4.) Then one can go from the left side of (16) to the right side through the substitution $y = h(x)$.

You should think of this theorem as describing the "transformation of an integration between two measure spaces."

Example: We have the following setup:

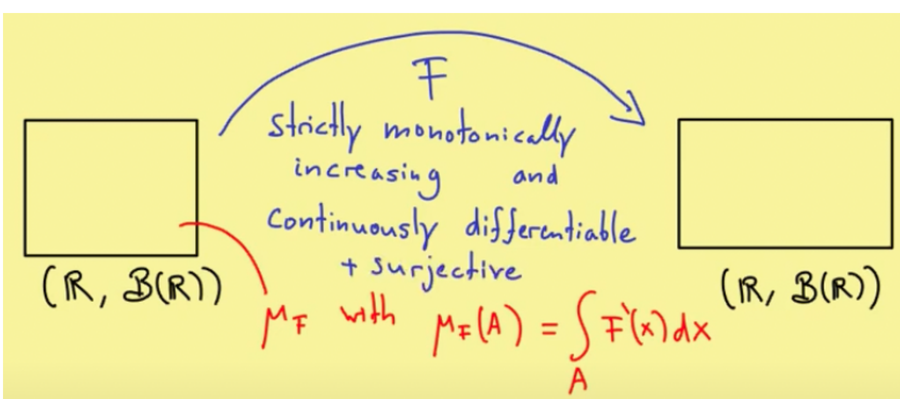

Figure 6: Image measure and substitution formula example

Note that with this definition, μ_F is a particular Lebesgue-Stieltjes measure—see the example at the end of Section 14. (Here, the integral $\int_A \mathcal{F}'(x)\,\mathrm{d}x$ in the definition of $\mathcal{F}$ is over the regular Lebesgue measure.)

Now we ask the question, "What is the image measure $\mathcal{F}_*\mu_{\mathcal{F}}$?" To answer this question, note that it is sufficient to work out $(\mathcal{F}_*\mu_{\mathcal{F}})([a,b))$. (This is because we are working on $\mathcal{B}(\mathbb{R})$, which is made up of unions, intersections, and complements of intervals of the form $[a,b)$. So using $(\mathcal{F}_*\mu_{\mathcal{F}})([a,b))$ and Property 2 of Definition 4.2, one can work out the measure of any set in $\mathcal{B}(\mathbb{R})$.) See the lecture for details, but he uses the definition of μ_F to work out that

$$(\mathcal{F}_*\mu_{\mathcal{F}})([a,b)) = \lambda([a,b)),$$

where λ is the Lebesgue measure, meaning that $\mathcal{F}_*\mu_{\mathcal{F}}$ is just the Lebesgue measure.

He then applies Theorem 16.2 to this scenario to show that it reduces to/reproves the common one-dimensional substitution rule. In other words, Theorem 16.2 contains the usual one-dimensional substitution rule from real analysis as a special case; that being

$$\int_{\mathbb{R}} g(y)\,\mathrm{d}y = \int_{\mathbb{R}} g(\mathcal{F}(x))\mathcal{F}'(x)\,\mathrm{d}x.$$

17 Proof of the substitution rule for measure spaces

In this lecture he proves Theorem 16.2.

Proof of Theorem 16.2: As per usual, he starts by proving Theorem 16.2 for the special case where g is a characteristic function, $g = X_C$ for some $C \in \mathscr{C}$. Since the Lebesgue integral of a characteristic function X_C is of course just the measure of the set C, this makes things very easy.

He then extends this to the case where g is a simple function using the linearity of the Lebesgue integral (see Theorem 7.3 and also Theorem 11.23 in [Rud76]).

He finally proves Theorem 16.2 for a general measurable function g using the fact that he has already proven Theorem 16.2 for simple functions and the definition of the Lebesgue integral (Definition 7.4). ■

18 The product measure and Cavalieri's principle

Key takeaways: *There is a natural way to combine two measure spaces two form a "product measure space." Cavalieri's principle allows you to compute the measures of sets in the product measure space.*

Suppose we have two measure spaces $(X_1, \mathscr{A}_1, \mu_1)$ and $(X_2, \mathscr{A}_2, \mu_2)$. Our goal is to define a product measure space $(X_1 \times X_2, \mathscr{A}, \mu)$. If we imagine a "rectangle" of the form $A_1 \times A_2$ for $A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2$, then clearly the product measure μ should be such that $\mu(A_1 \times A_2) = \mu(A_1) \cdot \mu(A_2)$. We can visualize this as follows:

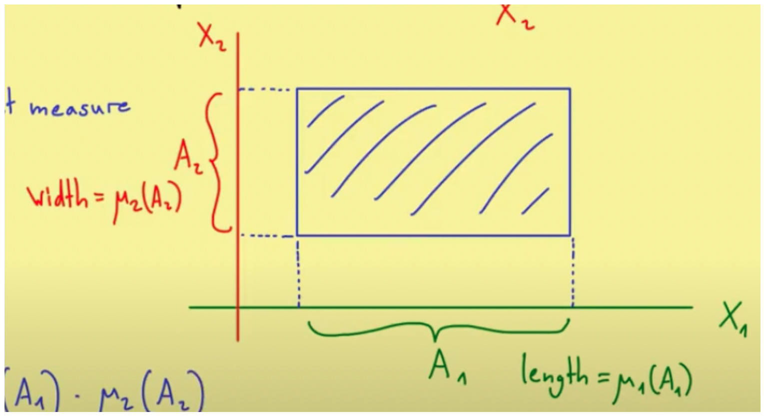

Figure 7: Product measure visualization

However, before we formalize this, we have to be careful with how we define the product σ-algebra $\mathscr{A}$. $\mathscr{A}$ cannot just be $\mathscr{A}_1 \times \mathscr{A}_2$ since the latter is not a σ-algebra. To see this, recall (Definition 2.1) first that σ-algebras are closed under union. Then if we write out two "rectangles" in $\mathscr{A}_1 \times \mathscr{A}_2$, we see that their union may not be a rectangle, implying it is not in $\mathscr{A}_1 \times \mathscr{A}_2$ (since the latter is precisely the set of "rectangles" of the form $A_1 \times A_2$ for $A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2$). Here is a picture:

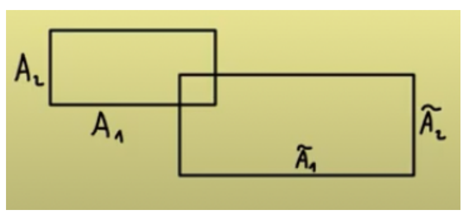

Figure 8: We need to be careful how we define the product σ-algebra $\mathscr{A}$.

However, we can of course extend $\mathscr{A}_1 \times \mathscr{A}_2$ to a σ-algebra. Thus, we make the following natural definition:

Definition 18.1 (Product σ-algebra) *With the setup as above, the product σ-algebra $\mathscr{A}$ is defined via $\mathscr{A} := \sigma(\mathscr{A}_1 \times \mathscr{A}_2)$ (see Definition 3.2).*

Now for the product measure:

Definition 18.2 (Product measure) *With the above setup, we first note that $\mathscr{A}_1 \times \mathscr{A}_2$ is in fact a semiring of sets (Definition 13.1). Thus, we can define a pre-measure $\tilde{\mu}$ via $\tilde{\mu}(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)$. We then use Theorem 13.3 to extend the pre-measure $\tilde{\mu}$ to a measure μ on the product σ-algebra $\mathscr{A}$ (Definition 18.1). We call μ the product measure.*

Note that the product measure in general is not unique.

As mentioned in Definition 18.2, the product measure is not in general unique. However, the following proposition reveals that it is unique in many cases, and in such cases, we can obtain a very nice explicit "formula" for the product measure.

Proposition 18.3 (Uniqueness of the product measure and Cavalieri's principle) *When μ_1, μ_2 are σ-finite,*[18] *then it follows from Theorem 13.3 (after maybe a bit of work) that the product measure μ from Definition 18.2 is unique.*

Furthermore, the unique product measure μ satisfies the following for any $M \in \sigma(\mathscr{A}_1 \times \mathscr{A}_2)$:

$$\mu(M) = \int_{X_2} \mu_1(M_y)\, \mathrm{d}\mu_2(y) = \int_{X_1} \mu_2(M_x)\, \mathrm{d}\mu_1(x)$$

where

$$M_y := (x_1 \in X_1 : (x_1, y) \in M)\,,$$
$$M_x := (x_2 \in X_2 : (x, x_1) \in M)\,.$$

This is known as Cavalieri's principle.

You can visualize the second half of Proposition 18.3 in the following way:

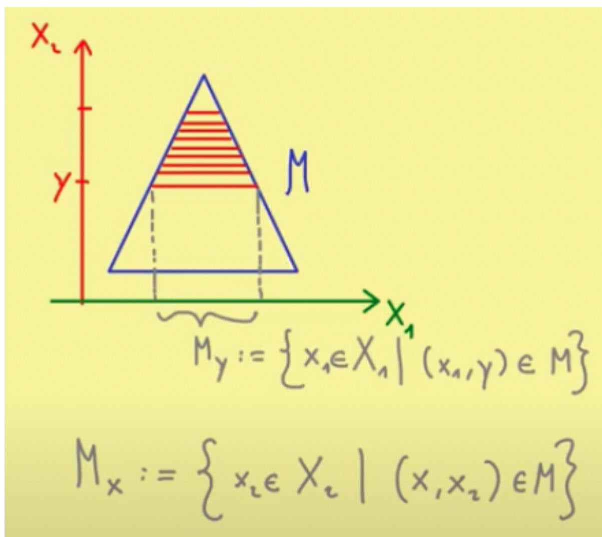

Figure 9: Cavalieri's principle visualization

Cavalieri's principle allows you to calculate the volume of a set M which is not a "rectangle" (i.e., not a set of the form $A_1 \times A_2$ for $A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2$) by dividing it into "slices" and using both measures u_1, u_2.

19 Cavalieri's principle: an example

In this lecture, he goes over a simple application of Cavaleri's principle (Proposition 18.3) to calculate the volume (measure) of a pyramid in $\mathbb{R}^3$. The only thing worth noting here is that we

[18] σ-finiteness was defined in Theorem 13.3.

define the Lebesgue measure in $\mathbb{R}^n$ using what we did in Section 18. We just treat $\mathbb{R}^n$ as a product measure space resulting from the product of $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ with itself n times. (Here, λ is of course the usual Lebesgue measure on $\mathbb{R}$.)

20 Fubini's theorem

The following result from multivariable calculus naturally can be extended to Lebesgue integration:

Theorem 20.1 (Fubini's theorem) *Let $(X_1, \mathscr{A}_1, \mu_1)$ and $(X_2, \mathscr{A}_2, \mu_2)$ be two measure spaces which we combine to form the product measure space $(X_1 \times X_2, \mathscr{A}, \mu)$ per Section 18. Furthermore, suppose μ_1, μ_2 are σ-finite so that Proposition 18.3 implies the uniqueness of the product measure μ.*

Now let $f \in \mathscr{L}^1(\mu)$ (Definition 11.1). Then

$$\int_{X_1 \times X_2} f \, \mathrm{d}\mu = \int_{X_2} \left(\int_{X_1} f(x,y) \, \mathrm{d}\mu_1(x) \right) \mathrm{d}\mu_2(y) = \int_{X_1} \left(\int_{X_2} f(x,y) \, \mathrm{d}\mu_2(y) \right) \mathrm{d}\mu_1(x). \tag{17}$$

In other words, calculating integrals with respect to the product measure is not any "harder" than calculating them with respect to the individual measures. Naturally, (17) can be extended to the case where the product measure space is the product of n measure spaces instead of just two.

He then gives an application of this theorem by calculating an integral in $\mathbb{R}^2$. (Basically, it is just a review of iterated integrals from multivariable calculus.)

21 Outer measures part 1

Key takeaways: *Outer measures, used to approximate a set from the outside, are key to proving some important results in measure theory. Not all outer measures are measures, but one can obtain a canonical measure from an outer measure.*

In this lecture we introduce outer measures. You don't actually need outer measures to "use measure theory." But if you want to prove certain important theorems in measure theory, like Carathéodory's extension theorem (Theorem 13.3), outer measures are essential. At a high level, outer measures are used to approximate a set from the outside. Note that "outer" is not an attribute for a measure in "outer measure." Outer measures are a completely new notion; an outer measure does not need to be a measure. That said, the construction of outer measures is not complicated; in fact, they may be even simpler to construct than ordinary measures.

Definition 21.1 (Outer measure) *Let X be some set. A map $\phi : \mathcal{P}(X) \to [0, \infty]$ is called an outer measure if*

1. $\phi(\emptyset) = 0$,
2. *$A \subseteq B$ implies $\phi(A) \leq \phi(B)$ (monotonicity), and*

3. for $A_1, A_2, \dots \in \mathcal{P}(X)$, we have that

$$\phi\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \phi(A_n). \tag{18}$$

This is known as "σ-subadditivity."[19] *Note that as usual, (18) also captures the union of finitely many A_i's since you can set however many A_i's you want to just be the empty set.*

It is worth comparing Definitions 4.2 (measure) and 21.1 (outer measure). The first and second properties of outer measures are also satisfied by measures. (In regards to the second property of outer measures, monotonicity, one can show that it is implied by the properties which measures satisfy.) The third property of outer measures, σ-subadditivity, is a weaker version of the second property of measures, σ-additivity, because the former has an inequality where the latter has an equality. (Actually, you need to be a bit more careful than this because the equality in σ-additivity is only for *disjoint* sets. But one can still show that σ-additivity implies σ-subadditivity.)

Another very important difference is that the domain of an outer measure is the entire power set, whereas the domain of a measure is a σ-algebra. Since outer measures are defined on the entire power set, we know from Section 5 that outer measures can't capture the Lebesgue measure. (In other words, they can't capture our intuitive notion of length on the real line.)

An important question then is, "How can we use an outer measure to obtain a related measure?" We know we will need to restrict the domain so that it is a σ-algebra, and we will also have to ensure that σ-additivity, as opposed to just σ-subadditivity, is satisfied. We now work toward accomplishing this goal:

Definition 21.2 (ϕ-measurable) *Let ϕ be an outer measure. $A \in \mathcal{P}(X)$ is called ϕ-measurable if for all $Q \in \mathcal{P}(X)$ we have*

$$\phi(Q) = \phi(Q \cap A) + \phi(Q \cap A^c). \tag{19}$$

Note that sometimes you may see this same definition except the equality (19) is replaced with $\phi(Q) \geq \phi(Q \cap A) + \phi(Q \cap A^c)$. The latter inequality is equivalent to (19) due to σ-subadditivity (Property 3 in Definition 21.1), so the definitions are equivalent.

You can visualize Definition 21.2 as follows:

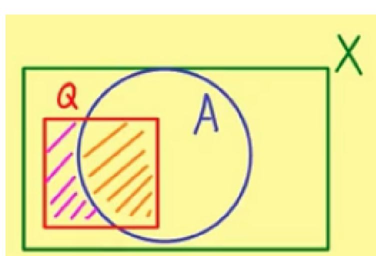

Figure 10: ϕ-measurability of A

[19]Note that σ-subadditivity is just a generalization of the union bound! In fact, they even note this at the top of

A being ϕ-measurable says that one can use A to "split ϕ over any other set,"[20] such as Q in Figure 10. (19) then says that the total measure of Q is the pink area plus the orange area. Intuitively, you should think of ϕ-measurable sets as being particularly "good" or well behaved. The next proposition makes this concrete:

Proposition 21.3 (Obtaining a measure from an outer measure) *If $\phi : \mathcal{P}(X) \to [0,\infty]$ is an outer measure, then*

$$\mathscr{A}_\phi := (A \subseteq X : A \text{ is } \phi\text{-measurable})$$

is a σ-algebra. Furthermore, the map $\mu : \mathscr{A}_\phi \to [0,\infty]$ defined via $\mu(A) = \phi(A)$ is a measure.

Proposition 21.3 gives us exactly what we were looking for: a way to obtain a natural measure from an outer measure.

22 Outer measures part 2: examples

Key takeaways: *Some outer measures are also measures but some are not. Outer measures can be used to construct the Lebesgue measure.*

In this lecture three examples of outer measures are given, the last example being particularly important because it is the outer measure which yields the Lebesgue measure after applying Proposition 21.3.

Example 1: We define an outer measure on $\mathbb{R}$, $\phi : \mathcal{P}(\mathbb{R}) \to [0,\infty]$, via

$$\phi(A) := \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{if } A \neq \emptyset. \end{cases}$$

In other words, we assign the empty set 0 and everything else 1. It is easy to check that monotonicity and σ-subadditivity (Properties 2 and 3 from Definition 21.1) hold. Additionally, note that ϕ is *not* a measure since σ-additivity (Property 2 of Definition 4.2) does not hold. This is easy to see; if you pick two disjoint sets A_1, A_2, then $\phi(A_1 \cup A_2) = 1$ but $\phi(A_1) + \phi(A_2) = 2$. Thus, this example confirms that not all outer measures are measures.

Example 2: We define an outer measure on $\mathbb{N}$, $\phi : \mathcal{P}(\mathbb{N}) \to [0,\infty]$, via

$$\phi(A) := \begin{cases} |A| & \text{if } A \text{ is a finite set,} \\ \infty & \text{otherwise.} \end{cases}$$

As we saw in Section 4, this is also a measure—the counting measure). Integration with respect to the counting measure gives you ordinary sums and series! This example shows that some outer measures are also measures.

the Wikipedia page for the union bound: "In measure-theoretic terms, Boole's inequality follows from the fact that a measure (and certainly any probability measure) is σ-subadditive." The point being that a probability measure is just a special kind of measure!

[20] More concretely, a set is ϕ-measurable if it divides all other sets into two pieces such that ϕ is additive over the

Example 3: We first define the semiring of sets (Definition 13.1) $\mathcal{I}$ as follows:

$$\mathcal{I} := \{[a, b) : a, b \in \mathbb{R}, a \leq b\}.$$

We also define the pre-measure (Definition 13.2) $\mu : \mathcal{I} \to [0, \infty]$ via $\mu([a, b)) = b - a$ so that it agrees with our usual notion of length. (Recall that we already did all of this in Section 13.)

Now define an outer measure on $\mathbb{R}$, $\phi : \mathcal{P}(\mathbb{R}) \to [0, \infty)$, via

$$\phi(A) := \inf \left\{ \sum_{j=1}^{\infty} \mu(I_j) : I_j \in \mathcal{I}, A \subseteq \bigcup_{j=1}^{\infty} I_j \right\}. \tag{20}$$

What is (20) doing? Well, if you give it a set $A \in \mathcal{P}(\mathbb{R})$, then the part inside the braces in (20) is covering A using intervals. (You can see that if A is an unbounded set, we would need a countably infinite number of intervals. At the same time, (20) also allows us to effectively cover A using finitely many intervals since we can set as many of the I_j's as we want to be empty intervals.) The infimum outside of the braces then says that we should use smaller and smaller intervals to cover A so as to better approximate A. Now you can see where the name "outer measure" comes from—we are approximating A from the outside by covering it with finer and finer intervals!

The rest of the lecture is spent checking that (20) indeed satisfies the properties of Definition 21.1. The proof that it satisfies σ-subadditivity is actually very nice—I would highly recommend (re)watching it. It feels very much like a classic real analysis proof.

As mentioned earlier, note that (20) yields the Lebesgue measure through an application of Proposition 21.3.

Finally, he notes at the end that in the case where you don't want to just calculate lengths but rather areas or n-dimensional volumes, you can use a similar definition for ϕ (and a similar proof to show that this "new" ϕ is indeed an outer measure). This way you can also construct the n-dimensional Lebesgue measure using the same ideas.

23 Outer measures part 3: proof

In this lecture he proves Proposition 21.3. The proof is basically a question of churning through the many terms (and therefore definitions) present in Proposition 21.3, so reviewing it can be a good way to get a handle on all of these definitions.

Note that Proposition 21.3 is very similar to Theorem 13.3; both have to do with pulling a measure out of a more general setup. (It might even be that Proposition 21.3 is used in the proof of Theorem 13.3, although I'm not sure.) He says at the end of the lecture that he plans to prove Theorem 13.3 in a future lecture, although this video has not come out at the time of writing.

24 The Riemann integral vs. the Lebesgue integral

A note about this lecture before starting: This is actually the oldest lecture in this video series (older even than the lecture corresponding to Section 2), and thus is in some sense separate from the other lectures. I'm including it at the end because that is what he does, but just know that it isn't really related to the lectures directly preceding it.

Key takeaways: *The main drawbacks of the Riemann integral are that it is cumbersome to extend the Riemann integral to higher dimensions, it struggles with "very discontinuous" functions, and it doesn't handle limit processes well. The Lebesgue integral solves these problems.*

In this lecture, he goes over various problems with the older Riemann integral and how the more modern Lebesgue integral resolves them. I'm not going to cover this lecture in a lot of detail here, but the main problems with the Riemann integral are the following:

1. Defining the Riemann integral of a function $f : \mathbb{R} \to \mathbb{R}$ is less cumbersome than defining the Lebesgue integral for such a function. However, the opposite is true as soon as you move into higher dimensions. Defining the Riemann integral of a function $f : \mathbb{R}^n \to \mathbb{R}$, while possible (divide the domain into hypercubes instead of intervals), is very cumbersome, particularly when the function f is not defined on all of $\mathbb{R}^n$ and you only want to integrate it over a subset of $\mathbb{R}^n$.

2. The Riemann integral cannot handle functions which are "very discontinuous," so it has some dependence on continuity.

3. The most important disadvantage of the Riemann integral has to do with its relation to limit processes, where the question is, "In what situations can I pull a limit inside an integral?" In other words, we want to know when the following is true:

$$\lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x \stackrel{?}{=} \int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{d}x. \tag{21}$$

   Recall from the examples at the beginning of Chapter 7 in [Rud76] that (21) is not in general true. (21) is true under the assumption of uniform convergence (Theorem 7.16 in [Rud76]), but this is a very strong assumption.

The Lebesgue integral solves all of these problems:

1. To show that it solves the "higher-dimensional extension" problem, he considers a function $f : \mathbb{R}^3 \to \mathbb{R}$ which we want to integrate. Now, the Riemann integral would need to figure out how to partition the domain, $\mathbb{R}^3$, to integrate this function. The Lebesge integral, on the other hand, notices that the codomain $\mathbb{R}$ is much simpler (and also "constant" in some sense, since even if the domain of our function is an abstract set X, the codomain of functions we want to Lebesgue integrate will always be $\mathbb{R}$). Thus, the Lebesgue integral partitions the codomain $\mathbb{R}$ instead of the complicated domain, as discussed earlier around Figure 4. The trade-off is that we need to be able to measure more complicated subsets of the domain, but this isn't hard to do once we have defined the notion of a measure.

2. As we saw earlier (e.g., Theorem 8.1) we don't really care how badly any function behaves (e.g., it can be wildly discontinuous) *as long as the bad behavior is confined to a set of measure zero.* This is already much more flexible than the Riemann integral.

3. As we saw in Sections 8 through 12 via various convergence theorems, the Lebesgue integral is well behaved when it comes to limit processes!

References

[Rud76] Walter Rudin. Principles of mathematical analysis. 3rd ed. International Series in Pure and Applied Mathematics. Düsseldorf etc.: McGraw-Hill Book Company. X, 342 p. DM 47.80 (1976)., 1976.

division.