Measure Theory for Analysts and Probabilists

Daniel Raban

Contents

1 Motivation

What is length? Intuitively, you might want to line up a ruler (or an interval on the real line) against an object to measure "length." This suggests that we could define length based on intervals on the real line; ignoring units, you can say that the length of the interval $[a, b]$ is $b - a$. We could fairly intuitively extend this idea to finite unions and intersections of intervals; just add the lengths of disjoint intervals. But what about other subsets of $\mathbb{R}$? Can we extend the idea of length to $\mathbb{Z}$, $\mathbb{Q}$, or complicated fractal-like subsets of $\mathbb{R}$?

What about area? This might seem even harder to define. More generally, it is a common desire to want to assign a number or a value to a set. We want to assign area or volume to subsets of $\mathbb{R}^2$ or $\mathbb{R}^3$. Maybe we have a distribution of mass in $\mathbb{R}^3$, and we want to be able to see how much mass is in any kind of set, even a fractal-like set (maybe this brings to mind an idea of integration). Or maybe we have a set of outcomes of some situation, and we want to define probabilities of subsets of these outcomes.

At the most basic level, measure theory is the theory of assigning a number, via a "measure," to subsets of some given set. Measure theory is the basic language of many disciplines, including analysis and probability. Measure theory also leads to a more powerful theory of integration than Riemann integration, and formalizes many intuitions about calculus.

2 Limitations of the theory

Before actually constructing measures, we must first devote some discussion to a fundamental limitation of the theory. In particular, **it is not always possible to measure all subsets of a given set**. We illustrate this with an example:

Example 2.1 (Vitali set)**.** Let's assume we can measure all subsets of $\mathbb{R}$, with some function $\mu : \mathscr{P}(\mathbb{R}) \to [0, \infty]$ that takes subsets of $\mathbb{R}$ and assigns them a "length." In particular, let's assume that μ has the following properties:

1. $\mu([0, 1)) = 1$.
2. If sets A_i are mutually disjoint, $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.
3. For all $x \in \mathbb{R}$, $\mu(A) = \mu(A + x)$, where $A + x := \{a + x : a \in A\}$ (invariance under translation).

Define an equivalence relation on $[0, 1)$ as follows: $x \sim y \iff x = y + q$ for some $q \in \mathbb{Q}$. This partitions $[0, 1)$ into equivalence classes of numbers that are closed under addition by rational numbers. Let N be a set containing exactly one member from each member of each equivalence class.[1] For $q \in \mathbb{Q} \cap [0,1)$, let

$$N_q := \{x+q : x \in N \cap [0, 1-q)\} \cup \{x+q-1 : x \in N \cap [1-q, 1)\}.$$

That is, we translate N over by q to the right by at most 1, and we take the part that sticks out of $[0,1)$ and shift it left by 1; you can also think of it as the translated set wrapping around the interval $[0,1)$, much like how movement works in a video game where moving past the right edge of the screen makes you enter from the left edge of the screen (e.g. PAC-MAN). This is just two translations, so $\mu(N_q) = \mu(N)$. You should note that

1. $N_q \subseteq [0,1)$ for all q.

2. $N_p \cap N_q = \varnothing$ if $p \neq q$: If $x \in N_p \cap N_q$, then $x - p$ (or $x - p + 1$) would equal $x - q$ (or $x - q + 1$), making two distinct elements of N differ by a rational number. This is impossible because then these two elements would be in the same equivalence class, so only one of them could be in N.

3. Each $x \in [0,1)$ is contained in some N_q: Given $x \in [0,1)$, let $y \in N$ be such that $x \sim y$. Then $x - y = q$ (or $q - 1$) for some $q \in [0,1)$, making $x \in N_y$.

So the N_q with $q \in \mathbb{Q} \cap [0,1)$ form a partition of $[0,1)$. We have already run into a problem: what is $\mu(N)$? Observe that

$$\mu([0,1)) = \mu\left(\bigcup_{q \in \mathbb{Q} \cap [0,1)} N_q\right) = \sum_{q \in \mathbb{Q} \cap [0,1)} \mu(N_q) = \sum_{q \in \mathbb{Q} \cap [0,1)} \mu(N).$$

If $\mu(N) = 0$, then $\mu([0,1)) = 0$. But if $\mu(N) = c > 0$, then $\mu([0,1)) = \infty$. So there is no value μ can assign N. We call N a **nonmeasurable** set.

Our assumptions in the preceding example were reasonable properties of any formalization of the concept of length. The fault here lies with our assertion that μ can measure any subset of $\mathbb{R}$. The next step, then, is to figure out what subsets of a set we *can* measure.

[1] This requires the axiom of choice, which is standard to accept in most fields of mathematics. If we reject the axiom of choice, all subsets of $\mathbb{R}$ can become measurable, but only because fewer things can be called "sets."

3 σ-algebras

3.1 Definition and examples

Definition 3.1. Let let X be any set. A **σ-algebra** (or σ-**field**)[2] $\mathcal{F} \subseteq \mathscr{P}(X)$ is a collection of subsets of X such that

1. $\mathcal{F} \neq \varnothing$.
2. If $E \in \mathcal{F}$, then $E^c \in F$.
3. If $E_1, E_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^\infty E_i \in \mathcal{F}$.

In other words, a σ-algebra is a nonempty collection that is closed under set complements and countable unions. The definition also implies closure under countable intersections because $\bigcap_{n=1}^\infty E_n = (\bigcup_{n=1}^\infty E_n^c)^c$.

Example 3.1. Let X be any set. Then $\mathscr{P}(X)$ is a σ-algebra.

Remark 3.1. You might wonder what the whole point of defining σ-algebras was, since $\mathscr{P}(X)$ is a σ-algebra. Indeed, using $\mathscr{P}(\mathbb{R})$ as a σ-algebra for our "length" measure would be no different from if we had not introduced σ-algebras at all. The key point to realize here is that some measures with desired properties (such as translation invariance) will be defined on smaller σ-algebras, while other measures may be defined on larger σ-algebras, even $\mathscr{P}(X)$.

Example 3.2. Let X be any set. The collection $\mathcal{F} = \{\varnothing, X\}$ is a σ-algebra.

In fact, this is in some sense the *minimal* σ-algebra on a set.

Proposition 3.1. *Let $\mathcal{F} \subseteq \mathscr{P}(X)$ be a σ-algebra. Then $\varnothing, X \in \mathcal{F}$.*

Proof. The collection $\mathcal{F}$ is nonempty, so there exists some $E \in \mathcal{F}$. Then $E^c \in \mathcal{F}$, and we get that $X = E \cup E^c \cup \cdots \in \mathcal{F}$. Moreover, $\varnothing = X^c \in \mathcal{F}$. $\square$

So a σ-algebra is also closed under *finite* unions, since $\bigcup_{i=1}^n E_i = \bigcup_{i=1}^n E_i \cup \varnothing \cup \varnothing \cup \cdots$. The same holds for finite intersections. Here is a nontrivial example of a σ-algebra.

Example 3.3. Let X be an uncountable set. Then the collection

$$\mathcal{F} := \{E \subseteq X : E \text{ is countable or } E^c \text{ is countable}\}$$

is a σ-algebra.

[2] These are not to be confused with algebras and fields from abstract algebra. In my experience, analysts use the term σ-algebra, and probabilists use σ-field.

3.2 Constructing σ-algebras

How do we construct σ-algebras? Often, it is difficult to just define a set that contains what we want. Sometimes, it is useful to create σ-algebras from other σ-algebras or from other collections of sets, such as a topology.

Proposition 3.2. *Let $\{\mathcal{F}_\alpha : \alpha \in A\}$ be a collection of σ-algebras on a set X. Then $\mathcal{F} = \bigcap_{\alpha \in A} \mathcal{F}_\alpha$ is a σ-algebra.*

Proof. We check the 3 parts of the definition.

1. Nonemptiness: $\mathcal{F} \neq \varnothing$ because $X \in \mathcal{F}_\alpha$ for each $\alpha \in A$.
2. Closure under complements: If $E \in \mathcal{F}$, then $E \in \mathcal{F}_\alpha$ for each $\alpha \in A$. Every $\mathcal{F}_\alpha$ is a σ-algebra and is closed under complements, so $E^c \in \mathcal{F}_\alpha$ for each $\alpha \in A$. Then $E^c \in \bigcap_{\alpha \in A} \mathcal{F}_\alpha = \mathcal{F}$.
3. Closure under countable unions: If $E_i \in \mathcal{F}$ for each $i \in \mathbb{N}$, then all the $E_i \in \mathcal{F}_\alpha$ for each $\alpha \in A$. Every $\mathcal{F}_\alpha$ is a σ-algebra and is closed under countable unions, so $\bigcup_{i=1}^\infty E_i \in \mathcal{F}_\alpha$ for each $\alpha \in A$. Then $\bigcup_{i=1}^\infty E_i \in \bigcap_{\alpha \in A} \mathcal{F}_\alpha = \mathcal{F}$. $\square$

We now introduce one of the most common ways to construct a σ-algebra. This construction closely mirrors other constructions in mathematics, such as closures of sets in topology and ideals generated by elements in ring theory.

Definition 3.2. Let $\mathcal{E}$ be a collection of subsets of a set X. The **σ-algebra generated by** $\mathcal{E}$, denoted $\sigma(\mathcal{E})$, is the smallest σ-algebra containing $\mathcal{E}$. That is, $\sigma(\mathcal{E})$ is the intersection of all σ-algebras containing E.

Why is this intersection well defined? There is always one σ-algebra containing $\mathcal{E}$, namely $\mathscr{P}(X)$. Now we can introduce possibly the most commonly used σ-algebra.

Example 3.4. Let X be a nonempty set, let $A \subseteq X$, and let $\mathcal{E} = \{A\}$. Then $\sigma(\mathcal{E}) = \{\varnothing, A, A^c, X\}$. If you take $B \subseteq X$ and let $\mathcal{E}' = \{A, B\}$, then $\sigma(\mathcal{E}')$ contains sets such as $A, B, A^c, A \cup B, A \cap B^c, (A \cup B)^c, \dots$. Constructing the σ-field generated by a collection of sets creates a large collection of sets to measure.

Example 3.5. Let X be a metric space (or topological space), and let $\mathcal{T}$ be the collection of open subsets of X. Then the **Borel σ-algebra** is given by $\mathcal{B}_X = \sigma(\mathcal{T})$. This σ-algebra contains open sets, closed sets, and more. We most commonly use $\mathcal{B}_\mathbb{R}$, which we will sometimes denote by $\mathcal{B}$. When we talk about $\mathcal{B}_{\mathbb{R}^d}$, we assume the Euclidean metric.[3]

[3] We can actually use any metric induced by a norm on $\mathbb{R}^d$, since all norms on a finite-dimensional $\mathbb{R}$-valued vector space induce equivalent metrics.

Example 3.6. What kinds of sets are in $\mathcal{B} = \mathcal{B}_{\mathbb{R}}$? By definition, $\mathcal{B}$ contains all open intervals and unions of open intervals. Since closed sets are complements of open sets, $\mathcal{B}$ contains all closed intervals, as well. For any $x \in \mathbb{R}$, $\{x\} \in \mathcal{B}$ because $\{x\} = \bigcap_{n=1}^{\infty} [x, x + 1/n]$. From this, we get that all countable subsets of $\mathbb{R}$ are in $\mathcal{B}$. In fact, most subsets of $\mathbb{R}$ you would ever care about can be found in $\mathcal{B}$, save for pathological sets we might construct, such as the Vitali set.

Here is an example motivated by probability theory.

Example 3.7. Suppose you are flipping a coin repeatedly.[4] Define the set of outcomes $\Omega = \{H, T\}^{\infty} := \{H, T\} \times \{H, T\} \times \cdots$. For each $n \geq 1$,

$$\mathcal{F}_n = \{A \times \{H, T\} \times \{H, T\} \times \cdots : A \subseteq \{H, T\}^n\}.$$

The collection $\mathcal{F}_n$ is a σ-field, and contains the events that can be determined after n flips of the coin. Note that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$. We can define $\mathcal{F}_{\infty}$, the σ-field of events in the whole random process, as $\sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$. This is actually larger than $\bigcup_{n=1}^{\infty} \mathcal{F}_n$, since all events in $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ are only determined by finitely many coin flips, while events in $\sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$ can be dependent on the results of infinitely many coin flips.

The previous example illustrates a fact about σ-algebras that you should be aware of: a union of σ-algebras need not be a σ-algebra.

We can also create σ-algebras on the Cartesian product of sets in a way that "respects" the σ-algebras on the components.

Definition 3.3. Let $\{X_{\alpha}\}_{\alpha \in A}$ be a collection of sets with corresponding σ-algebras $\{\mathcal{M}_{\alpha}\}_{\alpha \in A}$. The **product σ-algebra** $\bigotimes_{\alpha \in A} \mathcal{M}_{\alpha}$ is the σ-algebra on $\prod_{\alpha \in A} X_{\alpha}$ given by

$$\bigotimes_{\alpha \in A} \mathcal{M}_{\alpha} = \sigma(\{\pi_{\alpha}^{-1}(E_{\alpha}) : E_{\alpha} \in \mathcal{M}_{\alpha}, \alpha \in A\}),$$

where $\pi_{\alpha} : \prod_{\alpha \in A} X_{\alpha} \to X_{\alpha}$.

This definition is almost identical to the definition of the product topology in point-set topology. We leave the relationship between these two as an exercise.

Exercise 3.1. Show that $\mathcal{B}_{X \times Y} = \mathcal{B}_X \otimes \mathcal{B}_Y$.

[4]We assume that you have a lot of free time, so you flip the coin infinitely many times.

4 Measures

4.1 Definitions and examples

Now that we have introduced σ-fields, we can define measures.

Definition 4.1. Let X be a measure with a σ-algebra $\mathcal{F}$. A **measure** is a function $\mu : \mathcal{F} \to [0, \infty]$ that satisfies the following properties:

1. $\mu(\varnothing) = 0$.
2. If sets $E_1, E_2, \dots$ are mutually disjoint, then $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$.

The second condition is called **countable additivity**; the same property also holds for only finitely many sets $E_1, \dots, E_n$ (finite additivity) because you can just let $E_i = \varnothing$ for $i > n$. Note that μ can take on the value ∞; measures that do not are called **finite measures**.

Definition 4.2. A measure μ on a set X is called **σ-finite** if there exist countably many sets $E_1, E_2, \dots$ such that $\bigcup_{i=1}^{\infty} E_i = X$, and $\mu(E_i) < \infty$ for each i.

Sometimes it is easiest to prove statements for sets with finite measure. If a measure is σ-finite, we can often extend the proof to the whole space by first breaking it down into countably many finite pieces. It is generally rare to deal with measures that are not σ-finite, as they can be unruly.

Example 4.1. Let X be any set, equipped with the σ-algebra $\mathcal{F} = \mathscr{P}(X)$. The function $\mu(E) = |E|$ that returns the size of the set is a measure called **counting measure**. Counting measure is σ-finite iff X is countable.

Example 4.2. Let X be any nonempty set, equipped with the σ-algebra $\mathcal{F} = \mathscr{P}(X)$. Fix some $x \in X$. The function

$$\mu(E) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

is a measure called **point mass** at x.

Example 4.3. Let X be a countable set equipped with $\mathcal{F} = \mathscr{P}(X)$, and let μ be a measure on $(X, \mathcal{F})$. Then for any set $E \in \mathcal{F}$, $\mu(E) = \sum_{x \in E} \mu(\{x\})$.

Example 4.4. Let μ be a measure on $(X, \mathcal{F})$, and let $E \in \mathcal{F}$. Then the function $\mu_E : \mathcal{F} \to [0, \infty]$, given by

$$\mu_E(F) := \mu(E \cap F),$$

is a measure on E.

Definition 4.3. A **probability measure** $\mathbb{P}$ on a set Ω, equipped with a σ-field $\mathcal{F}$, is a measure with $\mathbb{P}(\Omega) = 1$.

Example 4.5. Imagine you flip a fair coin once. The set of outcomes is $\Omega = \{H, T\}$, and the σ-field of events is $\mathcal{F} = \mathscr{P}(\Omega)$. Define the function $\mathbb{P} : \mathcal{F} \to [0, 1]$ given by (on "rectangles" of events)

$$\mathbb{P}(E) = \begin{cases} 0 & E = \varnothing \\ 1/2 & E = \{H\} \text{ or } \{T\} \\ 1 & E = \Omega. \end{cases}$$

Later, we will learn how to formally extend such a function to a measure on any set in $\mathcal{F}$. The function $\mathbb{P}$ is a probability measure that gives the probability of each event occurring.

If we want to flip our coin n times, the set of outcomes is $\Omega_n = \{H, T\}^n$, our σ-algebra is $\mathcal{F}_n = \mathscr{P}(\Omega_n)$, and we can construct the probability measure

$$\mathbb{P}_n(E_1 \times \cdots \times E_n) = \prod_{i=1}^{n} \mathbb{P}(E_i).$$

If we flip the coin infinitely many times, the set of outcomes is $\Omega_\infty = \{H, T\}^\infty$, our σ-field is $\sigma(\bigcup_{n=1}^\infty \mathcal{F}_n)$, and we can construct the probability measure

$$\mathbb{P}_\infty(E_1 \times E_2 \times \cdots) = \prod_{i=1}^{\infty} \mathbb{P}(E_i).$$

Here is some terminology.

Definition 4.4. If X is a set, and $\mathcal{F} \subseteq \mathscr{P}(X)$ is a σ-algebra, then the pair $(X, \mathcal{F})$ is called a **measurable space**. If $\mu : \mathcal{F} \to [0, \infty]$ is a measure, then the triple $(X, \mathcal{F}, \mu)$ is called a **measure space**.

In probability theory, measure spaces are denoted by $(\Omega, \mathcal{F}, \mathbb{P})$.

4.2 Properties of measures

Here are four basic facts about probability measures. The first two properties should form your "common sense" intuition about what measures are. The third and fourth properties are very useful formal properties that allow us to determine the measure of complicated sets; they should also inform your intuition about how measures work.

Proposition 4.1. *Let $(X, \mathcal{F}, \mu)$ be a measure space. Then*

1. *(Monotonicity) If $E \subseteq F$, then $\mu(E) \leq \mu(F)$.*
2. *(Subadditivity) $\mu(\bigcup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mu(E_n)$.*
3. *(Continuity from below) If $E_1 \subseteq E_2 \subseteq \cdots$, then $\mu(\bigcup_{n=1}^\infty E_n) = \lim_{n\to\infty} \mu(E_n)$.*
4. *(Continuity from above) If $E_1 \supseteq E_2 \supseteq \cdots$, and $\mu(E_n) < \infty$ for some n, then $\mu(\bigcap_{n=1}^\infty E_n) = \lim_{n\to\infty} \mu(E_n)$.*

Proof. To prove 1, note that

$$\mu(E) \leq \mu(E) + \mu(E \setminus F) = \mu(E \cup (F \setminus E)) = \mu(F).$$

To prove 2, define $F_1 = E_1$ and $F_{n+1} = E_{n+1} \setminus \bigcup_{i=1}^n E_i$ for each $n \in \mathbb{N}$; F_{n+1} is the set of new elements in E_{n+1} that $E_1, \dots, E_n$ did not have. Then $F_n \subseteq E_n$ for each n, and the F_n are disjoint. So

$$\mu\left(\bigcup_{n=1}^\infty E_n\right) = \mu\left(\bigcup_{n=1}^\infty F_n\right) = \sum_{n=1}^\infty \mu(F_n) \leq \sum_{n=1}^\infty \mu(E_n),$$

where we used monotonicity for the last step.

To prove 3, observe that if $E \subseteq F$, $\mu(E)+\mu(E\setminus F) = \mu(F)$ implies that $\mu(F\setminus E) = \mu(F) - \mu(E)$. Then, setting $E_0 = \varnothing$, we get

$$\begin{aligned}\mu\left(\bigcup_{n=1}^\infty E_n\right) &= \mu\left(\bigcup_{n=1}^\infty (E_n \setminus E_{n-1})\right) = \sum_{n=1}^\infty \mu(E_n \setminus E_{n-1}) = \lim_{n\to\infty} \sum_{i=1}^n \mu(E_n) - \mu(E_{n-1}) \\ &= \lim_{n\to\infty} \mu(E_n).\end{aligned}$$

To prove 4, first assume (without loss of generality) that $\mu(E_1) < \infty$; otherwise, we can throw away the first finitely many E_i to make $\mu(E_1) < \infty$. Then

$$\mu\left(\bigcap_{n=1}^\infty E_n\right) = \mu\left(E_1 \setminus \bigcup_{n=2}^\infty (E_{n-1} \setminus E_n)\right) = \mu(E_1) - \sum_{n=2}^\infty \mu(E_{n-1} \setminus E_n)$$

$$= \lim_{n\to\infty} \mu(E_1) - \sum_{i=1}^{n} \mu(E_{n-1}) - \mu(E_n)$$
$$= \lim_{n\to\infty} \mu(E_n). \qquad \square$$

If μ is a probability measure, property 2 is known as the "union bound." It says that the probability of at least one out of a collection of events occurring is less than the sum of the probabilities when the events are considered separately.

Why are the continuity properties important? Suppose we had a measure corresponding to "area" in $\mathbb{R}^2$. If we wanted to find the area of a set E under a curve (as in integration), we could write $E = \bigcup_{n=1}^{\infty} E_n$, where E_n is some increasing sequence of successive approximations to the set E, such as approximations using rectangles. This is the basis for a very powerful theory of measure-theoretic integration.

5 Measurable functions and random variables

5.1 Measurable functions

How do measurable spaces interact with each other?

Definition 5.1. Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. A **measurable function** $f : X \to Y$ is a function such that for all $E \in \mathcal{N}$, $f^{-1}(E) \in \mathcal{M}$.

The idea here is that if we "pull back" measurable subsets of Y to X via f, they should still be measurable. Here is in some sense the most basic example of a measurable function.

Example 5.1. Let A be a measurable subset of X. Then the **indicator function**[5] of A,

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

is a measurable function from X to $(\mathbb{R}, \mathcal{B})$. Indicator functions are extremely important in both analysis and probability, and they often are the simplest examples of complicated ideas.[6]

This should look a lot like the topological definition of continuity. In fact, we have the following proposition.

[5] Some people call this a "characteristic function." The term "characteristic function" has multiple meanings elsewhere, but the term "indicator function" does not.

[6] Be nice to them, and they will help you in return.

Proposition 5.1. *Let X and Y be metric (or topological) spaces, equipped with the respective Borel σ-algebras $\mathcal{B}_X$ and $\mathcal{B}_Y$. Then if $f : X \to Y$ is continuous, it is measurable.*

Proof. Let $\mathcal{A} = \{E \in \mathcal{B}_Y : f^{-1}(E) \in \mathcal{B}_X\}$; we want to show that $\mathcal{A} = \mathcal{B}_Y$. For any open $U \subseteq Y$, $f^{-1}(U)$ is open in X by continuity, so $f^{-1}(U) \in \mathcal{B}_X$ because $\mathcal{B}_X$ contains all open subsets of X. This shows that $\mathcal{A}$ contains all open subsets of Y. Observe that $\mathcal{A}$ is a σ-algebra:

1. $\varnothing$ is open in Y, so $\mathcal{A}$ is nonempty.
2. If $E \in \mathcal{A}$, then $f^{-1}(E^c) = (f^{-1}(E))^c$, which is in $\mathcal{M}$ by the closure under complements of σ-algebras. So $E^c \in \mathcal{A}$.
3. If $E_1, E_2, \dots \in \mathcal{A}$, then $f^{-1}(\bigcup_{i=1}^{\infty} E_i) = \bigcup_{i=1}^{\infty} f^{-1}(E_i)$, which is in $\mathcal{M}$ by the closure under countable unions of σ-algebras.

So $\mathcal{A}$ is a σ-algebra containing the open sets of Y, which generate $\mathcal{B}_Y$; this gives $\mathcal{B}_Y \subseteq \mathcal{A}$. By definition, $\mathcal{A} \subseteq \mathcal{B}_Y$, so $\mathcal{A} = \mathcal{B}_Y$. $\square$

Actually, we have shown the following, more general fact.

Proposition 5.2. *If $f : X \to Y$ with σ-algebras $\mathcal{M}$ and $\sigma(\mathcal{A})$, and $f^{-1}(E) \in \mathcal{M}$ for each $E \in \mathcal{A}$, then f is measurable.*

So it is sufficient to check measurable sets in some collection that generates the σ-algebra of the codomain.

This proof, like many proofs in point-set topology, relies on the fact that inverse images of functions commute with unions and intersections. This is a strictly set-theoretic property you should be intimately familiar with.[7]

Exercise 5.1. **(a)** Let $f : X \to Y$ be a function. Show that for arbitrary unions and intersections,

$$f^{-1}\left(\bigcup_{\alpha} A_\alpha\right) = \bigcup_{\alpha} f^{-1}(A_\alpha),$$

$$f^{-1}\left(\bigcap_{\alpha} A_\alpha\right) = \bigcap_{\alpha} f^{-1}(A_\alpha),$$

[7] This is one of those things that it is publicly acceptable to be intimately familiar with. Relish this fact, and take it as your motivation to complete the associated exercise.

$$f^{-1}(A^c) = (f^{-1}(A))^c.$$

(b) Show that

$$f\left(\bigcup_\alpha A_\alpha\right) = \bigcup_\alpha f(A_\alpha),$$

and find a counterexample to show that this property does not hold for intersections.

While the previous exercise provides a motivation for the definition of measurable functions regarding formal manipulations, the following constructions provide much more satisfying motivation.

Example 5.2. Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(Y, \mathcal{N})$ be a measurable space, and let $f : X \to Y$ be a measurable function. We can define the **push-forward** measure ν on Y by setting $\nu(E) = \mu(f^{-1}(E))$. Check yourself that this is indeed a measure.

Example 5.3. Here is another way to construct σ-algebras. Given a set X, a measurable space $(Y, \mathcal{N})$, and a function $f : X \to Y$, we can construct the **pull-back σ-algebra** $f^{-1}(\mathcal{N})$ on X (or the **σ-algebra generated by** f) as

$$f^{-1}(\mathcal{N}) = \{f^{-1}(E) \subseteq X : E \in \mathcal{N}\}.$$

This σ-algebra is also sometimes denoted $\sigma(f)$. The σ-algebra generated by f is the smallest σ-algebra on X for which the function f is measurable.

If ν is a measure on Y, we can construct a **pull-back measure** μ on X by setting $\mu(f^{-1}(E)) = \nu(E)$. Check yourself that this is indeed a measure. Note, however, that f cannot define a pull-back measure on any σ-algebra on X; this only works for σ-algebras that are smaller than $f^{-1}(\mathcal{N})$.

Remark 5.1. In general, a composition of measurable functions need *not* be measurable. More explicitly, if $f : (X, \mathcal{M}) \to (Y, \mathcal{N})$, and $g : (Y, \mathcal{A}) \to (Z, \mathcal{C})$, where $\mathcal{N}$ and $\mathcal{A}$ are two different σ-algebras on Y, then $g \circ f : (X, \mathcal{M}) \to (Z, \mathcal{C})$ may not be measurable. This is very important to remember, as even with $\mathbb{R}$, there are multiple useful σ-algebras that are common to consider. However, if $\mathcal{N} = \mathcal{A}$, then the composition $g \circ f$ is measurable.[8] Check this yourself; the proof is the same as that for preservation of continuity under composition of functions.

[8]For the reader familiar with the language of category theory, this says that measurable functions are morphisms in the category of measurable spaces. The important distinction here is that in this category, the morphism carries the information of the σ-algebras of the domain and codomain.

5.2 Random variables and distributions

Measurable functions have an alternative, yet very important, interpretation in probability theory.

Definition 5.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a **random variable** X is a measurable function with domain Ω.

We view Ω as a space of events with some inherent "randomness," encapsulated by the probability measure $\mathbb{P}$.

Example 5.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to S$ be a constant function (i.e. $X(\omega) = c$ for some $c \in S$). Then for any σ-field $\mathcal{S}$ on S, X is a measurable function. We call X a **constant random variable**. We often interpret this situation as "deterministic" or having "no randomness."

Example 5.5. Let X be a real-valued random variable (we implicitly assume the Borel σ-algebra on $\mathbb{R}$). Since $f(x) = x^2$ is continuous, it is measurable (from $(\mathbb{R}, \mathcal{B})$ to $(\mathbb{R}, \mathcal{B})$). So X^2 is also a random variable. Similarly, $aX + b$ (for $a, b \in \mathbb{R}$), $\sin(X)$, e^X, etc. are random variables.

Often, random variables are specified by their distributions.

Definition 5.3. Let X be a random variable. The **distribution** of X is its push-forward measure.

To measure the probability of an event on the codomain of a random variable, the distribution "pushes forward" the value from the measure $\mathbb{P}$. In other words, we have

$$\mu(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

We suppress the ω notation and just write

$$\mu(A) = \mathbb{P}(X \in A).$$

Usually, we take $(\Omega, \mathcal{F}, \mathbb{P})$ to be a certain canonical "reference" measure space (with a measure we will discuss in depth later). For now, take it for granted that there is a space Ω with a sufficiently rich measure $\mathbb{P}$ that can generate all distributions on $\mathbb{R}$ as push-forward measures of random variables.

Example 5.6. Let μ be a probability measure on $(\{-1, 1\}, \mathscr{P}(\{-1, 1\}))$ given by $\mu(\{-1\}) = \mu(\{1\}) = 1/2$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be our canonical measure space. If X is a measurable function from Ω to $\{-1, -1\}$ with push-forward measure μ, we call X a **Rademacher random variable**. In particular, we have

$$\mathbb{P}(X = 1) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \{1\}\}) = \mu(\{1\}) = 1/2,$$
$$\mathbb{P}(X = -1) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \{-1\}\}) = \mu(\{-1\}) = 1/2.$$

Example 5.7. Let's construct a **Poisson random variable**. Let μ be a probability measure on $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$ given by

$$\mu(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!},$$

for some real-valued constant $\lambda > 0$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be our canonical measure space, and let X be a measurable function from Ω to $\mathbb{N}$ with push-forward measure μ. In particular,

$$\mathbb{P}(X = k) = \mu(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!},$$

and the probability of any subset of $\mathbb{N}$ can be specified by computing a countable sum of $\mathbb{P}(X = k)$ for different k.

One of the amazing aspects of measure theory is that is unifies the ideas of discrete and continuous probability (and even allows for mixing of the two). We have covered a few examples of discrete probability spaces above. Here is an example of a non-discrete case.

Example 5.8. A random variable with **uniform distribution** on $[0, 1]$ (also denoted as $U[0, 1]$) is a random variable X with codomain $([0, 1], \mathcal{B}_{[0,1]})$ and distribution μ that satisfies

$$\mu([a, b]) = \mathbb{P}(X \in [a, b]) = b - a.$$

for all $0 \leq a \leq b \leq 1$. The existence of such a measure is non-obvious, and we shall prove its existence in the next section.

Example 5.9. Here is a distribution on $([0, 1], \mathcal{B}_{[0,1]})$ that is not discrete but also has no continuous probability density over the real numbers. Let

$$\mu(\{0\}) = 1/2, \qquad \mu([a, b]) = \begin{cases} 0 & 0 < a \leq b < 1/2 \\ b - a & 1/2 \leq a \leq b \leq 1. \end{cases}$$

That is, μ is the uniform distribution but with all the probability in the interval $[0, 1/2)$ "concentrated" onto the value 0. Taking the existence of the uniform distribution for granted (moreso the fact that we can define a measure on $([0, 1], \mathcal{B}_{[0,1]})$ by defining its value on all subintervals), such a measure is well-defined.

To say that a random variable X has distribution μ, we write $X \sim \mu$.

5.3 Properties of real- and complex-valued measurable functions

In this section, we show that sums, products, and limits of real- and complex-valued measurable functions are measurable. Here, we always assume that the σ-algebra on $\mathbb{R}$ or $\mathbb{C}$ is $\mathcal{B}_{\mathbb{R}}$ or $\mathcal{B}_{\mathbb{C}}$, respectively. The most important parts of this section are not the results (which are not surprising) but rather the techniques used in the proofs.

Proposition 5.3. *Let $f, g : X \to \mathbb{R}$ be measurable functions. Then the functions $f + g$ and fg are measurable.*

Proof. By the proposition we proved when we introduced the idea of measurable functions, it suffices to show that $(f+g)^{-1}((a,b))$ is measurable for $a, b \in \mathbb{R}$; this is because every nonempty open set in $\mathbb{R}$ can be expressed as a countable union of open intervals, so these sets generate $\mathcal{B}$. The key property here is that $\mathbb{R}$ is separable (i.e. it contains the countable dense set $\mathbb{Q}$).

We have

$$\begin{aligned}(f+g)^{-1}((a,\infty)) &= \bigcup_{q\in\mathbb{Q}} \{x : f(x) > q\} \cap \{x : g(x) > a - q\} \\ &= \bigcup_{q\in\mathbb{Q}} f^{-1}((q,\infty)) \cap g^{-1}((a-q,\infty)),\end{aligned}$$

$$\begin{aligned}(f+g)^{-1}((-\infty,b)) &= \bigcup_{q\in\mathbb{Q}} \{x : f(x) < q\} \cap \{x : g(x) < b - q\} \\ &= \bigcup_{q\in\mathbb{Q}} f^{-1}((-\infty,q)) \cap g^{-1}((-\infty,b-q)),\end{aligned}$$

which are both measurable as countable unions of finite intersections of measurable sets. So

$$(f+g)^{-1}((a,b)) = (f+g)^{-1}((a,\infty)) \cap (f+g)^{-1}((-\infty,b))$$

is measurable as a finite intersection of measurable sets.

To show that fg is measurable, recall that continuous functions from $(\mathbb{R},\mathcal{B})$ to $(\mathbb{R},\mathcal{B})$ are measurable and that compositions of measurable functions (with σ-algebras that match up) are measurable. In particular, the functions $x \mapsto x^2$, $x \mapsto -x$ and $x \mapsto x/2$ are measurable. So, noting that

$$fg = \frac{1}{2}\big((f+g)^2 - f^2 - g^2\big),$$

we see that fg is a composition of such measurable functions and is consequently measurable. $\square$

Exercise 5.2. Show that if f, g are $\mathbb{R}$-valued measurable functions, then $\max(f, g)$ and $\min(f, g)$ are measurable.

To extend to the case of $\mathbb{C}$-valued measurable functions, we provide a more general framework for checking measurability of functions on product spaces.

Proposition 5.4. *Let $(X, \mathcal{M})$ and $(Y_\alpha, \mathcal{N}_\alpha)$ for each $\alpha \in A$ be measurable spaces, and let $\pi_\alpha : \prod_{\alpha \in A} Y_\alpha \to Y_\alpha$ be the projection maps. Then $f : X \to \prod_{\alpha \in A} Y$ is measurable iff $f_\alpha = \pi_\alpha \circ f$ is measurable for each $\alpha \in A$.*[9]

$$\begin{tikzcd}
 & & X \arrow[ddll, "f_1"'] \arrow[ddl, "f_2"'] \arrow[d, "f"] \arrow[ddr, "f_3"] \arrow[ddrr, "f_4"] & & \\
 & & (\prod_{n=1}^4 Y_n, \bigotimes_{n=1}^4 \mathcal{N}_n) \arrow[dll, "\pi_1"'] \arrow[dl, "\pi_2"] \arrow[dr, "\pi_3"'] \arrow[drr, "\pi_4"] & & \\
(Y_1, \mathcal{N}_1) & (Y_2, \mathcal{N}_2) & & (Y_3, \mathcal{N}_3) & (Y_4, \mathcal{N}_4)
\end{tikzcd}$$

Proof. ($\Longrightarrow$): Note that each π_α is measurable. So if f is measurable, so is every f_α because the composition of measurable functions is measurable (if the σ-algebras match up).

($\Longleftarrow$): Suppose each f_α is measurable. Then for all $E_\alpha \in \mathcal{N}_\alpha$, $f^{-1}(\pi^{-1}(\alpha)) = f_\alpha^{-1}(E_\alpha) \in \mathcal{M}$. Since $\{\pi^{-1}(E_\alpha) : E_\alpha \in \mathcal{N}_\alpha, \alpha \in A\}$ generates $\bigotimes_{\alpha \in A} \mathcal{N}_\alpha$ by definition, our proposition for determining measurability by checking a generating set gives us that f is measurable. $\square$

Applying this theorem to the complex numbers, we can separate out the requirement for measurability of a function into requirements for the real and imaginary parts of the function.

Corollary 5.1. *A function $f : X \to \mathbb{C}$ is measurable iff* $\mathrm{Re}(f)$ *and* $\mathrm{Im}(f)$ *are measurable.*

Proof. Since $\mathbb{R}^2$ and $\mathbb{C}$ are homeomorphic as topological spaces, we may view $\mathbb{C}$ as $\mathbb{R}^2$. Then

$$\mathcal{B}_{\mathbb{C}} = \mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$$

by the result of a previous exercise, and we can use the preceeding proposition to finish the rest. $\square$

[9] In the language of category theory, this proposition says that $(\prod_{\alpha \in A} Y_\alpha, \bigotimes_{\alpha \in A} \mathcal{N}_\alpha)$ is the product of $(Y_\alpha, \mathcal{N}_\alpha)$ in the category of measurable spaces.

Corollary 5.2. *Let $f, g : X \to \mathbb{C}$ be measurable functions. Then the functions $f+g$ and fg are measurable.*

Proof. By the previous corollary, it is sufficient to show that $\mathrm{Re}(f+g)$, $\mathrm{Im}(f+g)$, $\mathrm{Re}(fg)$, and $\mathrm{Im}(fg)$ are measurable. Moreover, the same corollary gives us that $\mathrm{Re}(f)$, $\mathrm{Im}(f)$, $\mathrm{Re}(g)$, and $\mathrm{Im}(g)$ are all measurable. We have

$$\mathrm{Re}(f+g) = \mathrm{Re}(f) + \mathrm{Re}(g), \qquad \mathrm{Im}(f+g) = \mathrm{Im}(f) + \mathrm{Im}(g),$$

$$\mathrm{Re}(fg) = \mathrm{Re}(f)\,\mathrm{Re}(g) - \mathrm{Im}(f)\,\mathrm{Im}(g), \qquad \mathrm{Im}(f+g) = \mathrm{Re}(f)\,\mathrm{Im}(g) + \mathrm{Im}(f)\,\mathrm{Re}(g),$$

which are measurable as sums and products of measurable real-valued functions. □

In probability theory, this gives us an often taken-for-granted result: that sums and products of random variables are indeed random variables.

What if we have a random process (i.e. a sequence of random variables)? Are limits of sequences of measurable functions measurable? To talk about limits on the real line, we must be prepared to have functions take values in $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. Equip $\overline{\mathbb{R}}$ with $\mathcal{B}_{\overline{\mathbb{R}}}$, the σ-algebra generated by $\{E \subseteq \overline{\mathbb{R}} : E \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}$.[10]

Proposition 5.5. *Let $\{f_j\}_{j\in\mathbb{N}}$ be a sequence of $\overline{\mathbb{R}}$-valued measurable functions. Then the functions*

$$g_1(x) := \sup_j f_j(x), \qquad g_2(x) := \inf_j f_j(x),$$

$$g_3(x) := \limsup_j f_j(x), \qquad g_4(x) := \liminf_j f_j(x),$$

are all measurable. Additionally, the set $\{x : \lim_{j\to\infty} f_j(x) \text{ exists}\}$ is measurable.

Proof. For g_1, it is sufficient to show that $g_1^{-1}((a,\infty))$ is a measurable set because $\{(a,\infty) : a \in \mathbb{R}\}$ generates $\mathcal{B}$; you can check this by checking that countable unions and intersections of sets in this collection gets you all of the open intervals in $\mathbb{R}$. We have that

$$g_1^{-1}((a,\infty)) = \{x : \sup_j f_j(x) > a\} = \bigcup_{j\in\mathbb{N}} \{x : f_j(x) > a\} = \bigcup_{j\in\mathbb{N}} f_j^{-1}((a,\infty)),$$

which is measurable as a countable union of measurable sets.

For measurability of g_2, note that

$$g_2(x) = \inf_j f_j(x) = -\sup_j -f_j(x),$$

[10] As the notation suggests, this is actually a Borel σ-algebra, induced by a metric on $\overline{\mathbb{R}}$. The metric is $\rho(x,y) = |\arctan(x) - \arctan(y)|$.

which is measurable because the inside is g_1 but with the functions $-f_j$.

For measurability of g_3, let $h_n(x) := \sup_{j\geq n} f_j(x)$; h_n is measurable by the same reasoning used for g_1. Then $g_3 = \inf_n h_n$, so g_3 is measurable since g_2 is measurable.

For measurability of g_4, note that

$$g_4(x) = \liminf_j f_j(x) = -\limsup_j -f_j(x),$$

which is measurable because the inside is g_3 but with the functions $-f_j$.

Finally, note that

$$\begin{aligned}\{x : \lim_{j\to\infty} f_j(x) \text{ exists}\} &= \{x : \limsup_j f_j(x) = \liminf_j f_j(x)\} \\ &= \{x : \limsup_j f_j(x) - \liminf_j f_j(x) = 0\} \\ &= (g_3 - g_4)^{-1}(\{0\}),\end{aligned}$$

which is measurable. □

The above proof relied heavily on the countability of the sequence $\{f_j\}_{j\in\mathbb{N}}$. In general, the result is not true for uncountable collections of functions.

Corollary 5.3. *If $\{f_j\}_{j\in\mathbb{N}}$ is a sequence of $\mathbb{R}$-valued measurable functions, and $\lim_{j\to\infty} f_j(x)$ exists for every x, then $f(x) := \lim_{j\to\infty} f_j(x)$ is measurable.*

Proof. If $\lim_{j\to\infty} f_j(x)$ exists for every x, then $f(x) = \limsup_j f_j(x)$, which is measurable by the previous proposition. □

Corollary 5.4. *If $\{f_j\}_{j\in\mathbb{N}}$ is a sequence of $\mathbb{C}$-valued measurable functions, and $\lim_{j\to\infty} f_j(x)$ exists for every x, then $f(x) := \lim_{j\to\infty} f_j(x)$ is measurable.*

Proof. If $\lim_{j\to\infty} f_j(x)$ exists for every x, then $\mathrm{Re}(f) = \lim_{j\to\infty} \mathrm{Re}(f_j)(x)$ and $\mathrm{Im}(f) = \lim_{j\to\infty} \mathrm{Im}(f_j)(x)$ exist for every x and are measurable functions. □

We conclude this section with a result that will be important later, when we discuss distribution functions in additional depth.

Proposition 5.6. *If $f : \mathbb{R} \to \mathbb{R}$ is monotone, it is measurable.*

Proof. Without loss of generality, we may assume f is increasing (otherwise consider the function $-f$). For each $a \in \mathbb{R}$, let $x_a = \sup\{x \in \mathbb{R} : f(x) < a\}$. We have two cases:

1. If $f(x_a) \leq a$, then $f^{-1}((a,\infty)) = \{x : f(x) > a\} = (x_a, \infty)$.
2. If $f(x_a) > a$, then $f^{-1}((a,\infty)) = \{x : f(x) > a\} = [x_a, \infty)$.

So $f^{-1}((a,\infty)) \in \mathcal{B}$ for every $a \in \mathbb{R}$. Hence, f is measurable. □

6 Construction of measures

6.1 Outer measure

Until now, we have been vague about how to construct measures, especially the more complicated measures on $\mathbb{R}$. We now develop tools for doing so. These constructions are sometimes skipped by people who wish to assume the existence of such measures and treat them as "black boxes." However, if you go on to do work involving measure theory, you will invariably run into issues involving measurability, and in such times, knowledge of outer measure will save the day.

Here is some motivation. We have had two main issues so far in constructing measures:

1. complicated non-measurable sets,

2. how to actually specify values for a large class of subsets of a measurable space.

Outer measure solves both these issues by "approximating complicated sets using simpler sets." For example, in $\mathbb{R}^2$, you can approximate the area under a curve by successively finer coverings of the area by rectangles (as in Riemann integration); you might call this approximation by "outer area." As an analogy, consider the relationship between the limit and $\limsup$ of a real-valued sequence; the $\limsup$ approximates the limit from above, and we can define the limit as the value of the $\limsup$ under certain conditions (in this case, the $\limsup$ equalling the $\liminf$).

The construction takes two steps and is summarized in the following diagram:

$$\text{premeasure } (\mu_0) \xrightarrow{\text{outer approx.}} \text{outer measure } (\mu^*) \xrightarrow{\text{Carathèodory}} \text{measure } (\mu)$$

$$\text{algebra } (\mathcal{E}) \xrightarrow{\text{outer approx.}} \mathscr{P}(X) \xrightarrow{\mu^*\text{-measurability}} \sigma\text{-algebra } (\mathcal{A})$$

Let's start in the middle, since outer measure is the most important of these.

Definition 6.1. An **outer measure** on a nonempty set X is a function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ such that

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

2. $\mu^*(A) \leq \mu^*(B)$ if $A \subseteq B$,

3. $\mu^*(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu^*(A_i)$.

Here is how we construct outer measure using "upper approximation." Take a function μ_0 which gives us the desired value of a measure on some relatively simple class of sets $\mathcal{E}$. Then we can define the outer measure of a set A by taking the best approximation of the value of μ_0 on coverings of A by simpler sets in $\mathcal{E}$.

Proposition 6.1. *Let $\mathcal{E} \subseteq \mathscr{P}(X)$ with $\varnothing, X \in \mathcal{E}$, and let $\mu_0 : \mathcal{E} \to [0,\infty]$ be a function such that $\mu_0(\varnothing) = 0$. For $A \subseteq X$, define the function*

$$\mu^*(A) = \inf\left\{\sum_{i=1}^\infty \mu_0(E_i) : E_i \in \mathcal{E}, A \subseteq \bigcup_{i=1}^\infty E_i\right\}.$$

Then μ^* is an outer measure.

Proof. We verify the three parts of the definition:

1. The set $\varnothing \in \mathcal{E}$, so we can set $E_i = \varnothing$ for all i to get $\mu^*(\varnothing) \leq \sum_{i=1}^\infty \mu_0(\varnothing) = 0$. So $\mu^*(\varnothing) = 0$.

2. If $A \subseteq B$, then $B \subseteq \bigcup_{i=1}^\infty E_i$ implies $A \subseteq \bigcup_{i=1}^\infty E_i$. So

$$\begin{aligned}\mu^*(A) &= \inf\left\{\sum_{i=1}^\infty \mu_0(E_i) : E_i \in \mathcal{E}, A \subseteq \bigcup_{i=1}^\infty E_i\right\} \\ &\leq \inf\left\{\sum_{i=1}^\infty \mu_0(E_i) : E_i \in \mathcal{E}, B \subseteq \bigcup_{i=1}^\infty E_i\right\} \\ &= \mu^*(B)\end{aligned}$$

because the infimum is being taken over a smaller collection of coverings.

3. Let $\varepsilon > 0$. For each A_i, choose $E_{i,j} \in \mathcal{E}$ for each $j \in \mathbb{N}$ such that $A_i \subseteq \bigcup_{j=1}^\infty E_{i,j}$ and $\sum_{j=1}^\infty \mu_0(E_{i,j}) \leq \mu^*(A_i) + \varepsilon 2^{-i}$. Then $\bigcup_{i=1}^\infty A_i \subseteq \bigcup_{i,j \geq 1} E_{i,j}$, and

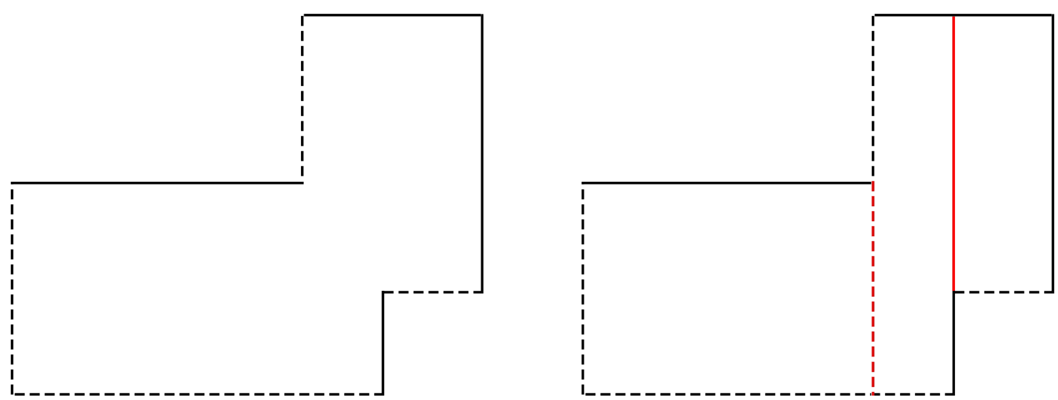
$$\mu^*\left(\bigcup_{i=1}^\infty A_i\right) \leq \sum_{i=1}^\infty \sum_{j=1}^\infty \mu_0(E_{i,j}) \leq \sum_{i=1}^\infty \mu^*(A_i) + \varepsilon 2^{-i} = \sum_{i=1}^\infty \mu^*(A_i) + \varepsilon.$$

This holds for every $\varepsilon > 0$, so $\mu^*(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu^*(A_i)$. □

The verification of the last part of the definition uses a very valuable[11] technique: if you wanst to establish inequalities (or equalities) involving an infimum or supremum, consider an element that almost achieves the infimum (or supremum) but misses it by at most ε. If you're wondering why an inequality holds (and are struggling to prove it), sometimes the ε comes in and solves everything like magic. In cases like this, taking a step back and viewing the problem from a a vaguer viewpoint of "things approximating other things" should provide you with the intuition you were missing.

What kind of function is μ_0? Defining an outer measure out of any function may lead to issues when we try to make the outer measure into a measure. The following example provides some intuition for what nice properties we need.

Example 6.1. Let $\mathcal{E}$ be the set of countable unions of half-open rectangles in $\mathbb{R}^2$; that is $\mathcal{E} = \{\bigcup_{i=1}^{\infty}(a_i, b_i] \times (c_i, d_i] : a_i, b_i, c_i, d_i \in \mathbb{R}\}$. Define the function $\mu_0 : \mathcal{E} \to [0, \infty]$ by $\mu_0((a, b] \times (c, d]) = (b-a)(d-c)$. So μ_0 just gives the area of a rectangle. For unions of disjoint rectangles, add the values of μ_0 on the different parts; and if two rectangles intersect, we can split the union into several disjoint half-open rectangles.[12]



We want to make an "outer area" outer measure μ^* that will behave nicely on complicated sets. What property of μ_0 makes it possible to determine the area of a complicated region?

A "good covering" of a complicated subset of $\mathbb{R}^2$ will probably consist of countably many tiny rectangles as to not overestimate the area of the region by too much. We

[11]It's also just super cool.

[12]We don't use closed rectangles because you can't split the union of two intersecting closed rectangles into disjoint closed rectangles. The boundaries of the rectangles end up intersecting.

have built in countable additivity into the function μ_0, so to approximate the area of the region, we add up the areas of countably many disjoint rectangles in our covering. This countable additivity condition is the condition we need.

When we define μ_0, we don't necessarily have a σ-algebra, but we should still be able to talk about how much measure we want to assign to complements of sets and unions of sets we are already dealing with. This is a restriction on $\mathcal{E}$.

Definition 6.2. An **algebra**[13] (or **field**) of subsets of X is a nonempty collection closed under complements and finite unions.

This is like a σ-algebra but without closure under *countable* unions. Defining μ_0 on such a collection is generally much easier than doing so on a σ-field. In fact, the whole construction of outer measure could be thought of as extending a measure from an algebra to a σ-algebra.

Example 6.2. Let X be a metric space (or a topological space), and let $\mathcal{E}$ be the collection of open and closed sets. Then $\mathcal{E}$ is an algebra.

We can now explicitly state what μ_0 should be.

Definition 6.3. Let $\mathcal{E}$ be an algebra. Then a **premeasure** $\mu_0 : \mathcal{E} \to [0, \infty]$ is a function such that

1. $\mu_0(\varnothing) = 0$,
2. If $(E_i)_{i \in \mathbb{N}}$ are disjoint, $E_i \in \mathcal{E}$, and $\bigcup_{i=1}^\infty E_i \in \mathcal{E}$, then $\mu_0(\bigcup_{i=1}^\infty E_i) = \sum_{i=1}^\infty \mu_0(E_i)$.

Note that countable additivity implies finite additivity by setting all but finitely many E_i equal to $\varnothing$. Since algebras are closed under finite unions, finite additivity always holds for premeasures.

Exercise 6.1. Let $\mu_0 : \mathcal{E} \to [0, \infty]$ be a premeasure, and let μ^* be the outer measure constructed from μ_0. Show that $\mu^*|_{\mathcal{E}} = \mu_0$.

Now that we have premeasures and outer measures, we can finally construct measures.[14] The next definition essentially does the work for us. It defines sets whose outer and "inner" measures are the same.

[13] This is not to be confused with an algebra or field in the abstract algebraic sense. The terminology is regrettable.

[14] Premeasures are not important to remember in detail; they are essentially an artifact of this construction. Outer measures, by contrast, are still useful when you can't guarantee the measurability of a set.

Definition 6.4. Let $A \subseteq X$. Then A is μ^*-**measurable** if for all $E \subseteq X$,

$$\mu^*(E) = \mu(E \cap A) + \mu(E \cap A^c).$$

If we rearrange this equation, we get

$$\mu^*(E) - \mu^*(E \cap A^c) = \mu^*(E \cap A).$$

In the example of outer area, this says that the outer and inner areas of A are equal.[15] Note that the inequality

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

always holds by the subadditivity of μ^*.

Proposition 6.2. *Let $\mu_0 : \mathcal{E} \to [0, \infty]$ be a premeasure, and let μ^* be the outer measure constructed from μ_0. Then every $A \in \mathcal{E}$ is μ^*-measurable.*

Proof. We need to show that

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

for every $E \subseteq X$. Let $E_i \in \mathcal{E}$ such that $E \subseteq \bigcup_{i=1}^\infty E_i$. Then, by the finite additivity of E,

$$\begin{aligned}
\sum_{i=1}^\infty \mu_0(E_i) &= \sum_{i=1}^\infty (\mu_0(E_i \cap A) + \mu_0(E_i \cap A^c)) \\
&= \sum_{i=1}^\infty \mu^*(E_i \cap A) + \sum_{i=1}^\infty \mu^*(E_i \cap A^c) \\
&\geq \mu^*(E \cap A) + \mu^*(E \cap A).
\end{aligned}$$

Only the left hand side is dependent on the covering $\bigcup_{i=1}^\infty \supseteq E$. If we take the infimum over all such coverings, the left hand side becomes $\mu^*(E)$. $\square$

Finally, we can construct our measures.

Theorem 6.1 (Carathèodory's Extension Theorem)**.** *Let μ^* be an outer measure, and let $\mathcal{A}$ be the collection of μ^*-measurable sets. Then $\mathcal{A}$ is a σ-algebra, and $\mu := \mu^*|_{\mathcal{A}}$ is a measure.*

[15] The definition of μ^*-measurability might seem unnatural. It is. This is the best interpretation I know of.

Proof. The collection $\mathcal{A}$ contains $\varnothing$, and it is closed under complements since the definition of μ^*-measurability is symmetric in A and A^c. So to prove that $\mathcal{A}$ is a σ-algebra, we need to show that it is closed under countable unions.

We first show that $\mathcal{A}$ is closed under finite unions. Let $A, B \in \mathcal{A}$. Then, for $E \subseteq X$,

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\ &= (\mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c)) + (\mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c))\end{aligned}$$

The first three terms partition the set $E \cap (A \cup B)$. The last set is equal to $E \cap (A \cup B)^c$. So by the subadditivity of outer measure,

$$\geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).$$

The reverse inequality follows from the subadditivity of outer measure, and we get that $A \cup B$ is μ^*-measurable. Now note that $A_1 \cup \cdots \cup A_n = (A_1 \cup \cdots \cup A_{n-1}) \cup A_n$. So by induction on the number of sets in the union, $\mathcal{A}$ is closed under finite unions.

We now extend to countable unions. Let A_n be μ^*-measurable for each n, and let $B_n = A_n \setminus \bigcup_{\ell=1}^{n-1} A_\ell$, the set of new elements in A_n). Then $A := \bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$, and

$$\mu^*(E) = \mu^*\left(E \cap \bigcup_{\ell=1}^n B_\ell\right) + \mu^*\left(E \cap \bigcup_{\ell=1}^n B_\ell^c\right)$$

Since $\bigcup_{\ell=1}^n B_\ell^c \supseteq A^c$, we can use monotonicity to get

$$\begin{aligned}&\geq \mu^*\left(E \cap \bigcup_{\ell=1}^n B_\ell\right) + \mu^*(E \cap A^c) \\ &= \sum_{\ell=1}^n \mu^*(E \cap B_\ell) + \mu^*(E \cap A^c).\end{aligned}$$

Only the right hand side depends on n, so we may let $n \to \infty$ on the right. We get

$$\mu^*(E) \geq \sum_{\ell=1}^\infty \mu^*(E \cap B_\ell) + \mu^*(E \cap A^c)$$

By subadditivity, since $E \cap A = \bigcup_{\ell=1}^\infty (E \cap B_\ell)$,

$$\geq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

As before, the reverse inequality is given by the subadditivity of μ^*, so $A = \bigcup_{n=1}^\infty A_i$ is μ^*-measurable. So we have shown that $\mathcal{A}$, the collection of μ^*-measurable sets, is a σ-algebra.

Note that $\mu^*(\varnothing) = 0$ by definition, so to show that $\mu := \mu^*|_{\mathcal{A}}$ is a measure, we need only show that μ is countably additive on disjoint sets. Let $(B_\ell)_{\ell \in \mathbb{N}}$ be disjoint μ^*-measurable sets. Recall that in proving closure under countable unions of μ^*-measurable sets, we had the inequality $\mu^*(E) \geq \sum_{\ell=1}^\infty \mu^*(E \cap B_\ell) + \mu^*(E \cap (\bigcup_{\ell=1}^\infty B_\ell)^c)$ when the B_ℓ are disjoint. We actually showed that this is an equality, since the right hand side is sandwiched between $\mu^*(E)$ and $\mu^*(E \cap \bigcup_{\ell=1}^\infty B_\ell) + \mu^*(E \cap (\bigcup_{\ell=1}^\infty B_\ell)^c)$. This holds for any $E \subseteq X$, so setting $E = \bigcup_{\ell=1}^\infty B_\ell$ gives us

$$\mu\left(\bigcup_{\ell=1}^\infty B_\ell\right) = \mu^*\left(\bigcup_{\ell=1}^\infty B_\ell\right) = \sum_{k=1}^\infty \mu^*\left(\bigcup_{\ell=1}^\infty B_\ell \cap B_k\right) + \mu^*(\varnothing) = \sum_{k=1}^\infty \mu(B_\ell).$$

So μ is a measure on the σ-algebra of μ^*-measurable sets. $\square$

Corollary 6.1. *Let $\mu_0 : \mathcal{E} \to [0, \infty]$ be a premeasure. Then there exists a measure μ such that $\mu|_{\mathcal{E}} = \mu_0$.*

Proof. Let μ^* be the outer measure constructed from μ_0, and let μ be the measure constructed from μ^*. The exercise about premeasures shows that $\mu^*|_{\mathcal{A}} = \mu_0$. Since every $A \in \mathcal{A}$ is μ^*-measurable, the domain of μ contains $\mathcal{A}$. $\square$

6.2 Distribution functions and Lebesgue measure

[coming to a set of notes near you]

References

[Bas13] R.F. Bass. *Real Analysis for Graduate Students*. Createspace Ind Pub, 2013.

[Dur10] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.

[Fol13] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2013.