A Primer on Analysis, Topology, and Measure Theory

Aditya Sengupta

October 31, 2019

This is a document with multiple purposes:

1. To relate measure theory content from Math 202A to probability content from EE 126.
2. To act as an easy reference for the Math 104 content that's used in Math 202A.
3. As a replacement for lecture notes for Math 202A, because I don't understand what's going on there right now.

I'm going to try and keep this compact. This means that sometimes, I'll have a definition that's somewhat opaque, or dependent on a lot of terms that were just introduced (otherwise this would be 200 pages and it'd be scary to even start reading.) I'll try and illustrate those with examples, or provide a restatement in simpler words, whenever that happens.

Section 1 consists of mathematical preliminaries: basic notation, and the definition of an equivalence relation (because it comes up in some surprising places). Sections 2 and 3 runs the reader through the essential parts of real analysis to understand metric spaces, in what I hope are relatively easy terms. Section 4 is an introduction to metric spaces, Section 5 covers topology, and Section 6 covers measure theory.

Contents

1 Preliminaries

1.1 Philosophizing

I like talking about math in accessible terms. It's very easy for notation and terminology and the details of proofs to obscure ideas that are inherently not that complicated — and this happens far too often in real analysis and measure theory. No one would care that a sequence is convergent if and only if it is Cauchy, if it's just presented as statements in a vacuum, without any context for why it's a question that anyone would care about. But any piece of mathematical theory was once invented and put down on paper by someone, and that person had reasons for caring about the question, and reasons why they came up with the answer that they did. Math is much more fun when you take this perspective — instead of a sequence of absolute truths, it's a collection of aesthetic decisions.

To that end, I'm going to do my best to present the thinking behind why we care about specific concepts, and why definitions are the way they are. I'm also going to try and keep this document as self-contained and jargon- and notation-free as possible. However, there are some things that it helps to know upfront, which I present just below. Additionally, there will be places where machinery is introduced before it's immediately obvious why it's necessary, just to maintain the flow of reasoning without having to pause in the middle of a proof to define a whole new framework for something. Sometimes a justification will be "this will make sense when we get to (some later topic)" — I don't like doing that, but sometimes it's true, so bear with me when that happens. I also hate the phrase "it turns out that" preceding by some very convenient result, and a proof of that result that makes the larger goal trivial. Wherever possible, I'll fill in gaps like that with some indication of how you could think of this yourself.

1.2 Notation

$\forall$	for all / for every
$\exists$	there exists
s. t.	such that
$P \implies Q$	P implies Q: if P is true, Q is true
$P \iff Q$	P if and only if Q: equivalent to $P \implies Q$ and $Q \implies P$
$x \in S$	x is an element of the set S
$x \notin S$	x is not an element of the set S
$\varnothing = \{\}$	the empty set
$A \subseteq B$	A is a subset of B: $\forall x \in A, x \in B$
$A \subset B$	A is a proper subset of B ($A \subseteq B$ and $A \neq B$)
$\|A\|$	cardinality or size of A (usually, the number of elements in a finite set)

1.3 The Axiom of Choice

The Zermelo-Fraenkel axioms with the Axiom of Choice (ZFC) are basic facts about sets that we consider to be true and build on further. Most of them are intuitive and seem reasonable as a starting point (which is why I won't list them here), but the Axiom of Choice is somewhat controversial. It states that for any collection of sets C, we can define a function f that takes in any set in the collection, and returns one of its elements. That is, there exists a function f such that for each $S \in C$, $f(S) \in S$. This seems obviously true: for example, I can say "take the minimum element" on the collection $\{\{1,2\},\{4,5\},\{6,7\}\}$ and get a function that returns $\{1,4,6\}$. However, we're going to run into some surprising results that are equivalent to the AoC, and some that seem intuitively false but hold because of the AoC!

1.4 Equivalence Relations

Consider a set S and elements $a, b, c \in S$. An equivalence relation on S is some statement of the form "a is related to b", denoted $a \sim b$, that satisfies these properties:

1. Reflexive: every element of S is related to itself, $a \sim a$.
2. Symmetric: if a is related to b then b is related to a, $a \sim b \implies b \sim a$.
3. Transitive: if a is related to b and b is related to c, then a is related to c, $a \sim b, b \sim c \implies a \sim c$.

As an example, consider equality: $a = a$, if $a = b$ then $b = a$, if $a = b$ and $b = c$ then $a = c$. $a \leq b$ is another equivalence relation that we can use when we've defined the idea of an order on a set (a level of theory that I'm going to skip). We're going to see a few different examples of equivalence relations, whenever we define some property that holds on a set but not uniquely, so we need to pick a representative of each set of elements that satisfy the equivalence relation, called an <u>equivalence class</u>.

2 The construction of the real numbers

I've skipped some content on precise definitions of certain sets and operations, subsequential limits, sums and products of sequences, convergence testing, and calculus due to a lack of relevance or because those things work like we'd expect: instead, I presented only the parts necessary for metric spaces, topology, and measure theory. If I come back to this and add Math 214, I may edit this to include calculus.

2.1 Motivation

I like to think of analysis as (at least some of) the math that arises when you unleash your inner annoying three-year-old and ask "why?" to everything. Specifically, it arises when you do this to calculus. There are strong reasons to ask questions about calculus, as it's possibly the type of mathematics that's used the most in other fields.

For example, let's start with the fact that to maximize or minimize a function, we take a derivative and set it to zero.
Why? Because a zero derivative is a flat point on the curve of a function.
Why? Because a derivative represents the local infinitesimal change in a function.
Why? Because a derivative is the limit $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$; some change in a function over a corresponding change in the input.
Why? Because as we go arbitrarily close to a point, all local changes look close enough to straight lines.
Why? Because we can get as close as we want to any point and make that true.
Why? Because of how we construct the real line, so that it doesn't have any gaps.
Why? Look below!

Pretty quickly, we get to fundamental questions about what real numbers actually are, why our definitions (such as the limit definition of a derivative) are the way they are, what things like "arbitrarily close" mean, and so on. The part of real analysis I'm going to cover is the first of these: how we construct $\mathbb{R}$ through a sequence of closing certain sets under operations (making it so that the result of an operation is always within the set.)

2.2 Construction of common sets

We're going to start building things up from the idea that I can look at a set of things and say "there are three things here" (I'm not going to get into the details of what a set is.) I know there are three because there's two and one more, so there's one and one and one more. We can formalize this idea as a successor function and define the natural numbers that way.

Natural numbers $\mathbb{N} = \{0, 1, 2, 3, \dots\}$; starting from zero, the set of all successors of elements of $\mathbb{N}$. The precise construction of $\mathbb{N}$ doesn't matter for understanding analysis, but you can look up the Peano axioms if you're interested. (This is the foundation for why mathematical induction works!) Addition and multiplication are defined rigorously according to this, but let's assume we know how those work. Further, we can formalize subtraction as inverse addition, which gives us the idea of the additive inverse $-n$ of n. We can extend the natural numbers to include these inverses.

Integers $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$; $\mathbb{N}$ along with its negation. $\mathbb{Z} = \mathbb{N} \cup \{-n \mid n \in \mathbb{N}\}$. This closes $\mathbb{N}$ under subtraction, i.e. creates the set of everything you can reach by taking the difference of natural numbers. This idea of "completing" a space will come in soon, when we try to construct the real numbers! We can also define multiplication on the integers, but we find that not every product has an inverse in $\mathbb{Z}$, so we can extend this further.

Rational numbers $\mathbb{Q}$ can be thought of as the set of pairs (a, b) where $a, b \in \mathbb{Z}$ and $b \neq 0$, and we define division accordingly. This is almost but not quite correct, because two fractions can be the same and have different representations; for example, $\frac{1}{2} = \frac{2}{4}$. This is where we get to use equivalence relations: we say the pair $(a, b) \sim (c, d)$ if $\frac{a}{b} = \frac{c}{d}$. Further, since we only know what multiplication is so far, we rewrite this as $a \cdot d = b \cdot c$. $\mathbb{Q}$ is the set of equivalence classes under this equivalence relation.

We've got some idea that $\mathbb{Q}$ has "gaps" in it. For example, I can define the polynomial $x^2 - 2$ and show that I can't make that zero for any choice of $x \in \mathbb{Q}$. So, we want to define the completion of $\mathbb{Q}$ that fills in these gaps. It'll turn out that "completion" is actually a formal term: $\mathbb{R}$ is the smallest complete metric space that contains $\mathbb{Q}$. The problem here is we haven't defined what we mean by "smallest", "complete", or "metric space" yet, so let's get on that. For now we can think of $\mathbb{R}$ as the set that fills in the gaps in $\mathbb{Q}$, and add rigor to that when we've built up the machinery to do this.

2.3 Tools to Build $\mathbb{R}$ from $\mathbb{Q}$

Infimum and supremum

The infimum of a set $\inf S$ is its maximal (largest possible) lower bound; for all $s \in S$, $\inf S \leq s$, and if there is another lower bound m such that for all $s \in S, m \leq s$ then $\inf S \geq m$. Similarly, the supremum of a set $\sup S$ is its minimal (smallest possible) upper bound; for all $s \in S, \sup S \geq s$ and if there is another upper bound M such that for all $s \in S, M \geq s$ then $\sup S \leq M$.

Think of the infimum and supremum as versions of the minimum and maximum that work for infinite sets. When a set is finite, $\inf S = \min S$ and $\sup S = \max S$. But when the set is infinite, we often can't talk about a minimum or maximum, so we use infimum or supremum instead. For example, the open interval $(0, 1) \in \mathbb{R}$ doesn't have a minimum element — if you suggest any element as the minimum, I can divide it by 2 and it'll be smaller but still in the set — but its infimum is 0 and supremum is 1 (try to prove this!)

The Triangle Inequality on $\mathbb{R}$

For $a, b \in \mathbb{R}$, $|a + b| \leq |a| + |b|$; the sum of two real numbers is less than or equal to the sum of their magnitudes. We can prove this by cases. We'll use this a lot, even when we move past the reals!

Denseness of $\mathbb{Q}$ in $\mathbb{R}$

For $a, b \in \mathbb{R}$, there's some $q \in \mathbb{Q}$ such that $a < q < b$.

Proof. $b - a > 0 \implies$ there's some $n \in \mathbb{N}$ such that $n(b - a) > 1 \implies nb > na + 1$ (this is the Archimedean property, which we don't need for anything but this proof). We want $m \in \mathbb{N}$ such that $nb > m > na \implies b > \frac{m}{n} > a$.

Intuitively, m should be the smallest integer greater than an. This can be shown explicitly: there's at least one natural number k greater than $|an|$ by the Archimedean property, so let's make a set of all the integers between an and k. Let $S = \{z \in \mathbb{Z} \mid |an| \leq z \leq k\}$. Because $k \in S, S \neq \varnothing$. Also, S is finite, because $-k \leq an \leq k$. Moreover, $\min S$ exists because $S \subset \mathbb{Z}$ and it is finite. Let $m = \min S$. Then $m \in S \implies m > an$, but $m - 1 \leq a_n$, because otherwise $m - 1$ would be $\min S$. So $an < m \leq an + 1 < bn$. Therefore m exists, meaning $\frac{m}{n}$ exists, as required. $\square$

Denseness is nice, because it gives us some idea that even if we can't write out a rational form for every real number, we can make a seqeuence of rational numbers that get closer and closer to it. For example, we can define a sequence $(1, 1.4, 1.41, 1.414, \dots)$ that gets closer and closer to $\sqrt{2}$. Let's formalize this idea.

Convergence of a sequence of reals / definition of a limit

We say $(s_n) \to s$, or $\lim_{n \to \infty} s_n = s$, if for every $\epsilon > 0$ there exists an N such that for any $n > N$ we know that $|s_n - s| < \epsilon$. (I'll often say "s_n is ϵ-close to s", or if s_n gets arbitrarily small (close to 0), "s_n is ϵ−small.")

This is a nice way to define convergence, because it gives us a quantitative way of saying "s_n gets close to s". $|s_n - s|$ is a measure of how close s_n is to s, so the above statement guarantees that at some finite point in the sequence, $|s_n - s| < \epsilon$, that is, s_n is separated from s by at most ϵ. Since we require that this is true of any $\epsilon > 0$, this means that we can get as close as we want to s by going sufficiently far in the sequence (s_n). Further, since we require that all $n > N$ have $|s_n - s| < \epsilon$, we can't have the sequence moving further away: it can only get closer and closer to s, like we want for convergence.

2.4 Building $\mathbb{R}$ from $\mathbb{Q}$

Now that we know what convergence is, let's try and build up the set of limits that we can construct with sequences of rational numbers, since we intuitively know that we should be able to reach any real number with a sequence of rationals. This requires that we build a mechanism to check whether a sequence is going to converge without us knowing its limit upfront. If we're going to define a real number by the limit of a certain sequence, we can't require that we know the limit of a convergent sequence to say that it converges.

An observation that could help us here is that when a sequence converges, the difference between terms gets arbitrarily small as you go further in the sequence. For example, take $s_n = \frac{1}{n}$ and consider the differences between successive terms: we get $1 - \frac{1}{2} = \frac{1}{2}, \frac{1}{2} - \frac{1}{3} = \frac{1}{6}, \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$, and so on getting smaller and smaller. We'll soon prove that this is an if-and-only-if, therefore we can characterize convergence based on this property of successive differences, or general differences between any two terms that are sufficiently far in the seqeuence. Let's formalize this:

Cauchy sequences

A sequence is Cauchy if for every $\epsilon > 0$ there exists an N such that for any two $m, n > N$ we know that $|s_m - s_n| < \epsilon$.

To show that this is an equivalent property to being convergent, we'll need the folowing intermediate steps.

Boundedness A set is bounded if $-\infty < \inf S \leq \sup S < \infty$, that is, it can be lower- and upper-bounded by finite real numbers. A sequence is bounded if the set of all its values is bounded. If a sequence converges it is bounded.

Bolzano-Weierstrass theorem Every bounded sequence has a convergent subsequence (a selection of infinitely many of its elements such that it converges). For example, $s_n = (-1)^n$ has the convergent subsequences of $s_{2n} = (1, 1, 1, \ldots)$ and $s_{2n+1} = (-1, -1, -1, \ldots)$. Note that subsequences have to be infinite.

Cauchy sequences are bounded If (s_n) is Cauchy, then it is bounded.

Proof. Pick $\epsilon = 1$ and find the corresponding N, and by taking its ceiling if required we can say it's an integer. Then for any $n > N$, we get that $|s_n| - |s_{N+1}| \leq |s_n - s_{N+1}| < 1$ so $|s_n| < |s_{N+1}| + 1$ for all $n > N$. Then every element of the sequence is either in the first N^+ or is in this upper bound, so $(|s_n|)$ is bounded by $M = \sup\{|s_1|, \ldots, |s_N|, |s_{N+1}| + 1\}$. (you can replace sup with max here as it's a finite set.) Therefore (s_n) is at most M and is at least $-M$, so it's bounded. □

Cauchy $\iff$ convergent

Any convergent sequence is Cauchy and any Cauchy sequence is convergent.

Proof. In the direction convergent $\implies$ Cauchy: suppose $(s_n) \to s$. Then for a fixed $\epsilon > 0$ there's an N such that $n > N \implies |s_n - s| < \frac{\epsilon}{2}$ and $m > N \implies |s_m - s| < \frac{\epsilon}{2}$. Then we use the triangle inequality:

$$|s_m - s_n| = |(s_m - s) - (s_n - s)| \leq |s_m - s| + |s_n - s| < \frac{\epsilon}{2} + \frac{\epsilon}{2} < \epsilon$$

so (s_n) is Cauchy.

In the direction Cauchy implies convergent: we know a Cauchy sequences is bounded, so by the Bolzano-Weierstrass theorem, it has a convergent subsequence $(s_{n_k}) = (s_{n_1}, s_{n_2}, \ldots)$, so if we believe that it converges, then it has to converge to the limit of the subsequence. Call this subsequential limit s. Fix $\epsilon > 0$ and take N_1 such that $m, n > N_1 \implies |s_m - s_n| < \frac{\epsilon}{2}$. Then, pick some s_{n_k} in the subsequence; because it converges, we know that for the same choice of ϵ as before, there exists some N_2 such that $n_k > N_2 \implies |s_{n_k} - s| < \frac{\epsilon}{2}$.

Finally, we choose $N = \max\{N_1, N_2\}$ and choose some $n_k > N$ (since subsequences are infinite, we're allowed to do this). Then,

$$\begin{aligned} n > N \implies |s - s_n| &= |(s - s_{n_k}) + (s_{n_k} - s_n)| && \\ &\leq |s - s_{n_k}| + |s_{n_k} - s_n| && \text{triangle inequality} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon && n_k, n > N \text{ and } (s_n) \text{ Cauchy, and subsequence converges} \end{aligned}$$

Therefore $n > N \implies |s - s_n| < \epsilon$, i.e. the sequence converges. □

This completes the machinery we need to build $\mathbb{R}$: we identify each element of $\mathbb{R}$ as a Cauchy sequence consisting of elements of $\mathbb{Q}$. We can see that with this definition, $\mathbb{Q} \subset \mathbb{R}$, because we can just take the sequence $(q, q, q, \ldots)$. This isn't the unique way to construct $\mathbb{R}$ (look up Dedekind cuts for an equivalent one), but this is the construction of $\mathbb{R}$ that scaffolds the theory of metric spaces.

3 Real Analysis

3.1 Proving $\mathbb{R}$ Is What We Expect

We've constructed something out of rational numbers that seems to fill in the gaps in $\mathbb{Q}$, but there are reasons not to be convinced that this is $\mathbb{R}$ as we know it, or that we've constructed something worth constructing. For one thing, I justified the "gaps in $\mathbb{Q}$" argument by providing a polynomial that didn't have a zero on $\mathbb{Q}$, but by the same token, there are gaps in $\mathbb{R}$, because x^2+1 doesn't have any zeros in $\mathbb{R}$. This is admittedly a slightly different situation, because for x^2-2 we could numerically approximate a rational zero, i.e. for any $\epsilon > 0$ we could find some q such that $|q^2 - 2| < \epsilon$ (choose $q = s_n, n >$ the N corresponding to ϵ, where s_n is the Cauchy sequence of rationals that we identify as $\sqrt{2}$). However, we can't do that in this case, because $|x^2+1| \geq 1$, so for any $\epsilon < 1$, we can't get $\epsilon-$close to the zero. Even so, we still have some notion that $\mathbb{R}$ might need to be completed or closed under some operation, so we don't know that there's anything special about $\mathbb{R}$ yet.

Further, I haven't shown you that we can fill all the gaps on the real line through Cauchy sequences of rational numbers; how do we know that there aren't any holes in this extension? How do we know that the zero for x^2+1 isn't a real number that we just can't get to through this method?

To address these, I'll prove two things: that $\mathbb{R}$ with this definitions forms a field (addressing the "nothing special" concern) and that $\mathbb{R}$ is complete, i.e. every Cauchy sequence of elements of $\mathbb{R}$ converges to something in $\mathbb{R}$.

Fields A full treatment of groups, rings, and fields would require a completely different note, so I'll just present all of the field axioms. This isn't as compact as defining it in terms of abelian groups, but it's more straightforward when we don't really need to know what an abelian group is.

A field is any set F on which there are two operations that satisfy the following axioms.

The first operation is addition, satisfying

1. associativity, $(a+b)+c = a+(b+c)\ \forall a,b,c \in F$;
2. existence of an additive identity, usually called 0, such that $a+0=0+a=a\ \forall a \in F$;
3. existence of additive inverses for each element in the field, i.e. $\forall a \in F, \exists -a \in F \text{ s. t. } a+(-a) = (-a)+a = 0$;
4. commutativity $(a+b=b+a\ \forall a,b \in F)$.

The second operation is multiplication, satisfying

5. associativity, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
6. existence of a multiplicative identity, usually called 1, such that $a \cdot 1 = 1 \cdot a = a\ \forall a \in F$;
7. existence of multiplicative inverses for each element in the field except possibly 0, i.e. $\forall a \in F/\{0\}, \exists a^{-1} \in F/\{0\} \text{ s. t. } a^{-1} \cdot a = a \cdot a^{-1} = 1$;
8. commutativity $(a \cdot b = b \cdot a\ \forall a,b \in F)$.

Finally, the combination of addition and multiplication has to satisfy

9. left and right distributivity: $\forall a,b,c \in F, a \cdot (b+c) = (a \cdot b) + (a \cdot c)$ and $(a+b) \cdot c = (a \cdot c) + (b \cdot c)$.

This might look similar to the axioms for a vector space if you're familiar with those, and it is: it's almost exactly the same, just that associativity of multiplication in a vector space shows an equivalence between field-field and field-vector multiplication whereas this is just a fact about the field.

So we've introduced all this theory to say that we can show that $\mathbb{R}$ is a field. This should help you intuitively think about the construction we just did from $\mathbb{Q}$, i.e. that we've done something nontrivial and made an object that we can do algebra over in more generality than with $\mathbb{Q}$ (which is also a field.)

$\mathbb{R}$ is a field

If we define addition and multiplication by adding and multiplying termwise the rational Cauchy sequences that represent each real number, i.e.,

$$\begin{aligned} ((s+t)_n) &= (s_n) + (t_n) = (s_n + t_n) \\ ((st)_n) &= (s_n)(t_n) = (s_n t_n) \end{aligned}$$

and we select $(0,0,0,\dots)$ as the additive identity and $(1,1,1,\dots)$ as the multiplicative identity, then $\mathbb{R}$ is a field.

Proof. We can show addition satisfies associativity, because addition of rational numbers is associative and we're essentially just doing that infinitely many times. We defined a zero element and we can show that works, because it works termwise. There are additive-inverse sequences, which we get just by inverting every term, i.e. $-(s_n) = (-s_n)$. Addition commutes, because addition of every rational term commutes.

Similarly, we can show multiplication satisfies associativity because it holds termwise. We defined a unit element and we can show that works, because it works termwise. There are multiplicative-inverse sequences, which we get just by inverting every term, i.e. $(s_n)^{-1} = (s_n^{-1})$. Multiplication commutes, which it does because multiplication of every rational term commutes.

Finally, distributivity holds because it holds termwise for rational terms. □

A fun note here is that the additive identity is the sequence that's all zeros, and the multiplicative identity is the sequence that's all ones. We said above that if a sequence is all one constant rational, we can identify it as that element of $\mathbb{Q}$, which means the 0 and 1 from $\mathbb{Q}$ carry over here. This is a nice property that we'd expect from something that contains $\mathbb{Q}$.

Of course, all this is predicated on the rational numbers being a field, which we could also prove from first principles (going back to the successor function and more on set theory). But the major takeaway is that if we define this construction from a field, we'll get another field, which makes it believable that the real numbers are fundamental in some sense, because they're closed under some natural operations.

Closure doesn't necessarily mean the set is special. For example, the set of integers that divide some n are also closed under addition and multiplication (although they're not a field as there's no multiplicative identity unless $n = 1$). Let's make a stronger claim: the reals are special among fields because we've filled in all the gaps; it's the unique complete field that extends the rationals. If we used another construction of the reals, we would be able to show that it is isomorphic to this one. (Technically, it's the unique complete ordered field, but order matters more from a set-theoretic than an analytic perspective, so we'll disregard this.) We've seen it's a field, so now let's show it's complete.

Completeness of the reals

Every Cauchy sequence of real numbers converges to a real number.

This definition will make sense when we talk about a generalization of the reals to metric spaces. For now, note that this definition means that for every Cauchy sequence of real numbers, there exists a Cauchy sequence of rational numbers converging to the same limit. In other words, when we took the set of Cauchy sequences of rational numbers, we didn't leave any gaps: we could get as precise as we wanted by taking a sequence of real numbers converging to whatever we wanted, and we'd still have been able to do it with just the rationals. This is kind of a crazy idea when you think about it: there's uncountably many real numbers (arguments like Cantor diagonalization show this and also clarify what is meant by countably vs. uncountably infinite), and yet we can reach them with just the countable number of rationals.

To prove this, we introduce the concepts of limit superior (lim sup) and limit inferior (lim inf); these are essentially the supremum and infimum of a sequence for arbitrarily large n.

Limit inferior and limit superior

$\liminf s_n$ is the infimum of the set of values a sequence takes on as n goes to infinity. Let $S_n = \{s_m \mid m \geq n\}$; then $\liminf s_n = \lim\limits_{n\to\infty} \inf S_n$. Similarly, $\limsup s_n$ is the supremum of that same set: $\limsup s_n = \lim\limits_{n\to\infty} \sup S_n$.

These will be useful in proving the reals are complete by bounding a sequence of real numbers below and above, and showing that the bounding sequence (the ones whose limits are the lim inf and lim sup of the sequence) converge to the same limit, so the actual limit of the sequence must be the same as both of those. Let's clarify why this is true.

If a sequence converges, then we expect that under a limit the maximum and minimum attainable values should both converge to the actual limit of the sequence: as you go further and further in the sequence, the supremum of all the sequence values from that point onwards should get closer and closer to the limit, and so should the infimum. For example, $s_n = \frac{1}{2^n}$ has a lim inf of 0 because it's bounded below by 0, and a lim sup of 0 because for every $\epsilon > 0$, I can give an N such that $n > N \implies s_n < \epsilon$. So I can make the sequence elements arbitrarily small positive numbers, meaning that in the limit the sequence is at most 0. Since it's at least 0 and at most 0, it must be exactly 0. We can formalize this:

$\liminf s_n = \limsup s_n = \lim s_n$ iff it converges

If a sequence converges, the lim inf and lim sup are the same and both equal the limit. However, if a sequence doesn't converge, the lim sup and lim inf are not the same (consider $s_n = (-1)^n$). This is an if-and-only-if statement, so we can also say that if $\liminf s_n = \limsup s_n$ then s_n converges to both of them. (I'm skipping the proof for now, but I may come back and add it later.)

Now, we're ready to prove that every Cauchy sequence of real numbers converges to a real number.

Proof of completeness of the reals

Proof. Consider a Cauchy sequence of real numbers (s_n). Because the sequence is Cauchy, we know it must be bounded (its lim sup and lim inf are both finite), so consider the sequences $(l_n) = (\inf\{s_m \mid m \geq n\}), (u_n) = (\sup\{s_m \mid m \geq n\})$. (l_n) is bounded below by the infimum of (s_n), and (u_n) is bounded above by the supremum of (s_n), both of which are finite because (s_n) is bounded. Further, the sequence of infima can only increase and the sequence of suprema can only decrease (i.e. they are both monotonic). Therefore (l_n) converges to $\liminf s_n \in \mathbb{R}$ and (u_n) converges to $\limsup s_n \in \mathbb{R}$.

Fix $\epsilon > 0$; because s_n is Cauchy, we can find $N \in \mathbb{N}$ such that $m \geq N$ implies $|s_m - s_N| < \frac{\epsilon}{2}$, that is, $s_N - \frac{\epsilon}{2} < s_m < s_N + \frac{\epsilon}{2}$. The sequence of infima (l_n) is strictly less than the sequence itself, and the sequence of suprema (u_n) is strictly greater, but both are the tightest possible bounds which means each infimum must be greater than this bound we've placed on s_n, and each supremum must be less than the bound. That is, for all $m \geq N$, $l_m > s_N - \frac{\epsilon}{2}$ and $u_m < s_N + \frac{\epsilon}{2}$. Therefore, consider the difference between the inf and sup:

$$|u_m - l_m| = u_m - l_m < s_N + \frac{\epsilon}{2} - s_N + \frac{\epsilon}{2} = \epsilon$$

i.e. for all $\epsilon > 0$ there exists an N such that $m \geq N \implies |u_m - l_m| < \epsilon$. Therefore $(u_n - l_n)$ is a sequence converging to 0, so $\lim(u_n - l_n) = 0 \implies \limsup s_n - \liminf s_n = 0$. Therefore the sequence converges to $\lim s_n = \limsup s_n = \liminf s_n$. □

This should convince you that $\mathbb{R}$ is special and is a useful thing to have constructed. We could still come up with an extension of $\mathbb{R}$ based on the previous argument that $x^2 + 1$ doesn't have any zeros in $\mathbb{R}$, but we've shown that we can't get arbitrarily close to a zero, so this does not represent a "gap" to be filled in $\mathbb{R}$. (We can still construct $\mathbb{C}$ from $\mathbb{R}$ using the idea of a field extension, but again, that's a topic for the Algebra version of this document.)

3.2 Continuous Functions on $\mathbb{R}$

We have a formal definition of what it means to be complete, and we know that $\mathbb{R}$ satisfies it, which is enough for us to be able to do calculus. To start with this, let's define what a function is, and define what it means for a function to be continuous.

Functions A function from set S to set T is a choice for every $s \in S$ of some $f(s) \in T$. This is denoted $f : S \to T$. (Sometimes, a function is denoted along with its action on an arbitrary element of S, as in $f : S \to T, s \to f(s)$.)

We call S the domain of f and T the codomain of f. T is sometimes also referred to as the range or image. Terminology is confusing here; personally, I like codomain as the full space in which the output values live that is somehow analogous to the domain (for example, if S is a field then the codomain of $f : S \to T$). The image is the set of values that are actually attained, i.e. $\{f(s) \mid s \in S\}$. The range is used as either of these. I'll try not to use range, but usually I like to use it interchangeably with image.

Note that this is only partially an aesthetic choice: once you've defined a function and define the sets it acts between, the codomain and image are well defined and not up to interpretation. The choice comes in setting what T is. For example, consider the two functions

$$f : \mathbb{R} \to \mathbb{R}, x \to x^2$$
$$g : \mathbb{R} \to \mathbb{R}^+, x \to x^2$$

where $\mathbb{R}^+ = [0, \infty) \subset \mathbb{R}$. The codomain of f is $\mathbb{R}$, but the image is $\mathbb{R}^+$ because squaring a real number can only get you a positive real number. The codomain of g is $\mathbb{R}^+$, and the image is also $\mathbb{R}^+$. The "aesthetic choice" part of this comes in choosing whether you want to consider f or g as the function that squares real numbers. I prefer f because you have the terminology to deal with both the field where the output values live, and the actually-attained output values, but g is also valid.

Continuous functions on $\mathbb{R}$

We say a function $f : S \to \mathbb{R}, x \to f(x)$ is continuous at some point $a \in S$ if either of the following (equivalent) properties hold:

- for all sequences (a_n) such that $a_n \in S$ for all n, we can say $\lim f(a_n) = f(a) = f(\lim a_n)$.
- for all $\epsilon > 0$, there exists a $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

We say f is continuous if it is continuous at a for all $a \in S$.

The first property (the sequence definition of continuity) is probably the more fundamental one, because it comes clearly out of the idea of convergence of a sequence that we established. Essentially, whenever we choose a sequence that converges to the input point we care about, applying the function termwise to the sequence should give us another sequence that converges, and we say that sequence converges to $f(a)$. This is a reasonable notion of what it means to be continuous: you could imagine a geometric mapping of the real line to some curve in space such that you could draw it without lifting your pen. This would have the property that if two sequences represent real numbers that are arbitrarily close together, then the points on the curve they map to also represent real numbers whose closeness is related to that. Put more simply, a continuous function should preserve the idea that there are no gaps in the real line when we map every point in its domain to something in its range.

The second property (the epsilon-delta definition of continuity) is more algebraically useful, and we'll use it more when we move from the real line to metric spaces. We claim that the two definitions are equivalent.

Proof. First, we show that if the sequence definition holds, then the epsilon-delta definition holds. We do this by contraposition, that is, by showing that if the epsilon-delta definition does not hold, then the sequence definition does not hold either.

If the epsilon-delta definition doesn't hold then there's some $\epsilon > 0$ such that no valid choice of δ exists; that is, for every $\delta > 0$ there's some $x \in S$ such that $|x - a| < \delta$, but $|f(x) - f(a)| \geq \epsilon$. For each $n \in \mathbb{N}^+$ (new notation I'm introducing for the natural numbers without zero), choose $\delta = \frac{1}{n}$ and set a_n equal to the corresponding $x \in S$ that's $\delta-$close to a but whose function value is not $\epsilon-$close to $f(a)$. Then $(a_n) \to a$ because $\lim \frac{1}{n} = 0$, so a_n gets arbitrarily close to a for sufficiently high n. However, $f(a_n)$ does not converge to $f(a)$ because $|f(a_n) - f(a)| < \epsilon$ is never true for any n. Therefore the sequence definition does not hold.

Second, we show that if the epsilon-delta definition holds, then the sequence definition holds. We do this by identifying δ as playing the usual role of ϵ in the convergence of a sequence. Pick an arbitrary sequence $(a_n) \to a$. If for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\forall x \in S, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$, then we can say there exists some N such that $n > N$ implies $|a_n - a| < \delta$. That is, we can eliminate δ in the definition:

$$\begin{aligned}
&\forall \epsilon > 0, \exists \delta > 0 \,\text{s.t.}\; \forall x \in S, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon \\
&\forall \epsilon > 0, \exists N \in \mathbb{R} \,\text{s.t.}\; n > N \implies |a_n - a| < \delta \\
\implies &\forall \epsilon > 0, \exists N \in \mathbb{R} \,\text{s.t.}\; n > N \implies |f(a_n) - f(a)| < \epsilon
\end{aligned}$$

What we've ended up with is the definition of the convergence of a sequence, which is what we wanted! □

Analogous to the idea that a sequence being Cauchy meant we didn't need to know what a sequence converged to in order to say it was convergent, we can define the idea of uniform continuity, a way of saying a function is continuous at some a without having to write down what $f(a)$ is.

Uniform continuity

$f : S \to \mathbb{R}$ is uniformly continuous on a subset $T \subseteq S$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in T$,

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

This isn't completely the same idea as the relationship between being Cauchy and convergent: in the case of the reals, uniform continuity is only the same as continuity if the subset T is a closed interval. (If we take T to be an open interval, we can define a unique extension of the function to the corresponding closed interval such that it is continuous.)

Equivalence of continuity and uniform continuity

Let $f : [a, b] \to \mathbb{R}$, where $[a, b]$ represents the set $\{r \in \mathbb{R} \mid a \leq r \leq b\}$. Then f is uniformly continuous (abbreviated u.c.) if and only if f is continuous.

Proof. The direction "u.c. implies continuous" can be proved by translating the u.c. condition to a statement about sequences. Take an arbitrary point $c \in [a, b]$. Suppose f is u.c.; then, for all $\epsilon > 0$, if x and c are $\delta-$close, we can look at the Cauchy sequence converging to x and say it approaches being $\delta-$close to c. Call this (x_n); then there exists some N such that $n > N \implies |x_n - c| < \delta \implies |f(x_n) - f(c)| < \epsilon$. This is the sequence definition of continuity,therefore a u.c. function must be continuous.

In the direction "continuous implies u.c.', suppose, towards a contradiction, that f is continuous but not u.c. Then there exists some $\epsilon > 0$ such that for every δ, there exist $x, y \in [a, b] \,\text{s.t.}\, |x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$.

For all n, take $\delta = \frac{1}{n}$. Then, there exist $x_n, y_n \in [a,b]$ such that $|x_n - y_n| < \frac{1}{n}$, but $|f(x_n) - f(y_n)| \geq \epsilon$. (x_n) is bounded, so by the Bolzano-Weierstrass theorem, there exists (x_{n_k}) convergent. Let $z = \lim_{k \to +\infty} x_{n_k}$. Then (y_{n_k}) converges to z as well because it's $\delta-$close to (x_{n_k}). $x_{n_k}, y_{n_k} \in [a,b] \forall k$ and $z \in [a,b]$. So f is continuous at z, which implies

$$\lim f(x_{n_k}) = f(z) = \lim f(y_{n_k})$$
$$\therefore \lim f(x_{n_k}) - f(y_{n_k}) = 0$$

But $|f(x_{n_k}) - f(y_{n_k}| \geq \epsilon$ for all k, which is a contradiction. Therefore f must be uniformly continuous. $\square$

4 Metric Spaces

4.1 Motivation

So far, we've built up a lot of tools that allowed us to define limits and continuity on the real line, which opens up a ton of possibilities. Some of the cool results you can prove are the intermediate, mean, and extreme value theorems; the definitions of a derivative and integral (Darboux and Riemann, and their equivalence); why the chain rule works; and Taylor's theorem and the theoretical foundation for being able to expand analytic functions into a power series. But instead of just building this theory for the reals, we could look at more general spaces, derive more powerful results, and talk about more general spaces along with the reals all at once. For example, suppose we'd like to talk about how functions, random variables, or binary strings converge when we operate on them somehow. The study of metric spaces asks the question: what are the essential components of the reals that allowed us to build up our ideas of convergence and continuity, and how can we make those more general?

4.2 Defining a metric space

If you're given a set, it seems pretty clear how to define a sequence of points in that set: just pick one of them for each $n \in \mathbb{N}^+$. Proving whether it converges might be harder. We've built up an intuitive idea that convergence means we get closer and closer to some point, but what's meant by "close" when we don't know what distance looks like outside of the real line? For example, how far away is one function $f : \mathbb{R} \to \mathbb{R}$ from another?

We can define our own idea of what distance means if we look at the essential properties we've been using of our distance function on $\mathbb{R}$ (the absolute value of the difference). It's always positive, and if two points are the same then they have distance zero (and vice versa). It's symmetric: it doesn't matter if I do $|x-y|$ or $|y-x|$. And the property we've used most of all is the triangle inequality. Based on this, we define the requirements for a function to be a distance function on a set, or metric.

Definition of a metric

A metric on a set X is a function $d : X \times X \to \mathbb{R}^+$ such that for all $x, y, z \in X$,

1. $d(x, y) = 0$ if and only if $x = y$: no distinct points can be zero distance apart.
2. $d(x, y) = d(y, x)$: distance is the same whether I start with x or y.
3. $d(x, y) \leq d(x, z) + d(z, y)$: the triangle inequality.

One reason why this is a nice definition is the conceptual symmetry with an equivalence relation: both of them have conditions relating to reflexivity, symmetry, and transitivity. (As far as I know, this symmetry is only conceptual, and it doesn't really make sense to talk about equivalence classes under a metric as some sort of subspace, because, for example, transitivity won't hold like you want it to.) Further, we've encapsulated all the properties we implicitly or explicitly liked about the absolute value distance. Finally, we can use our metric to convert questions about how close points in an arbitrary space are to questions about how close the real-number outputs of the metric are.

Definition of a metric space

A metric space (X, d) is any set X paired with a metric d on X.

4.3 $\mathbb{R}$ as a metric space

Since we built this definition of a metric space with the motivation that we want to generalize properties of the reals, a first sanity check would be to verify that $\mathbb{R}$ with $d_1(x, y) = |x - y|$ is a metric space. We can easily see that the absolute value is a valid metric. What's interesting, though, is that this isn't the only possible metric on $\mathbb{R}$. For example, consider the trivial-looking metric

$$d_0 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+, d_0(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

This is reminiscent of a Dirac delta (just that this is nonzero everywhere the Dirac delta would be zero, and this is more well-defined.) Note that we can't say this is $1 - \delta_{xy}$ (the Kronecker delta) as we're working in a continuous space. This satisfies all of the requirements for being a metric, so we could use it. The tradeoff for the relative simplicity is that we lose the ability to talk about convergence in many cases. For example, I can't show that $s_n = \frac{1}{n}$ converges to 0 under this norm, because I can't say any particular term gets $\epsilon-$small unless I actually reach 0 (which I don't). The sequence of distances is $(1, 1, 1, 1, \dots)$, so if I pick any $\epsilon < 1$ I can't say there exists an N such that $n > N \implies d_0(s_n, 0) < \epsilon$.

Another interesting norm on the reals is the squared distance: $d_2(x, y) = (x - y)^2$. This has essentially the same behaviour as the absolute value — any convex even function applied to $x - y$ would have the same characteristics. Nonetheless, it's interesting to look at these because they scaffold the idea of the L_p norms that we'll look at next, which is in turn going to let us show that $\mathbb{R}^n$ is a valid metric space.

4.4 Norms

There are other properties of the absolute value, such as idempotence ($||a|| = |a|$) or multiplicativity ($|ab| = |a||b|$), so it's natural to ask why those aren't part of the definition of a metric. The answer is that generalizing these to a metric doesn't really make sense. The absolute value mapped pairs of real numbers to other real numbers, and it had an implicit idea that you could take the absolute value of one point in $\mathbb{R}$ by looking at its distance from zero. We could generalize the second by saying that conventionally $d(x) = d(x, 0)$ for some definition of zero in each space, but if the space we're looking at isn't the reals, then either idempotence or multiplicativity would need us to look at d of a real number, which isn't defined. Some properties of the absolute values therefore can't be generalized to a metric, but we didn't really use these when we were establishing real analysis anyway, so this is fine.

Nonetheless, the idea of looking at the magnitude of a real number $|a|$ by looking at its distance from zero was useful, and we'd like to make a conceptual generalization of this way of looking at the absolute value, just without necessarily extending all the individual properties. This will give us a different angle on the triangle inequality that'll be easier to deal with, and will give us some more general results for dealing with metric spaces so that we can establish some more examples.

4.5 $\mathbb{R}^n$ as a metric space

A natural space to look at is the multidimensional space of reals $\mathbb{R}^n$. A natural choice of metric is the Euclidean distance, and now that we've built up the theory of norms, we can show this works.

The Euclidean distance is a metric on $\mathbb{R}^n$

The Euclidean distance $d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ between points $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$ is a metric on $\mathbb{R}^n$.

Proof.

1. If the Euclidean distance between x and y is zero, then $x_i = y_i$ for all $1 \leq i \leq n$, so $x = y$. (This works in both directions.)
2. Symmetry holds because $d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} = d_2(y, x)$.
3. The triangle inequality is geometrically obvious, but its proof requires that we build some more machinery first.

□

Oops. The triangle inequality needs the Cauchy-Schwartz inequality, which means we need some idea of a norm. I'll come back to this.

5 Topology (TBD)

5.1 Motivation

5.2 Construction of Topologies

5.3 Strength of a Topology

6 Measure Theory

This section is very much a work in progress, because I'm learning the intuition behind measure theory as I write it, so I'm not going to guarantee that it's at all intuitive or understandable.

6.1 Motivation

In the spirit of wanting to generalize the ways in which we reason about the real line and related easily-visualizable metric spaces, let's try and generalize the idea of length. In a way, we already did this for the real line with the norm, but the norm doesn't completely answer the question of "how large is an interval (a, b)?" We have the length of the vectors to a and b (the norms of a and b), and we can geometrically reason through this to say that (a, b) should have a length $b - a$, so the length of this interval (or to use a term that we'll be using a lot soon, the measure of this interval) is a clear concept and it makes sense that we would want to ask about it. In $\mathbb{R}^2$, the analogous concept is area, and in $\mathbb{R}^3$ it's volume, so we can ask about the concept of a subset of a space and how "large" it is in some sense.

6.2 Measures

6.2.1 Motivation for Sigma-Algebras: The Vitali Set

Let's start with a set X. We're interested in the question: "how big is a subset of X?". A reasonable definition for a measure would be to say that it's a function that assigns a positive real number to any subset of X, but we can show that this doesn't perfectly hold with the following example on the real numbers, called the Vitali set.

We start with the measure that we would intuitively want on the real line. Let's say this has the following properties:

- An interval (a, b) (or its closed or half-open version) has measure $b - a$.
- Arbitrary sets can be expressed as the union of arbitrarily many disjoint intervals and their measures can be added.

We're going to construct a subset of the reals for which the idea of a measure doesn't hold like we want it to.

Take the interval $[0, 1]$ under the equivalence relation $x \sim y \iff x - y \in \mathbb{Q}$: two elements are related if their difference is rational. This subdivides $[0, 1]$ into equivalence classes of elements that are either rational or offset from the rationals by some irrational number that's the same for all the elements in the class.

Next, using the Axiom of Choice, let's define V_0 to be the set constructed by taking one element from each equivalence class. Because the rationals are countable, let's enumerate the subset $\mathbb{Q} \cap [-1, 1]$. That is, for each $n \in \mathbb{N}^+$, assign a unique q_n by a process like Cantor enumeration.

Then, let $V_n = \{v + q_n \mid v \in V_0\}$, the set generated by taking every element of V_0 and shifting it by some rational q that's fixed for the set and determined by the set index.

We can say that the V_is cover $[0, 1]$: for every $r \in \mathbb{R}$ there is some $n \in \mathbb{N}^+$ and some $v \in V_0$ such that $r = q_n + v$. This is the case because V_0 covers $[0, 1]$ up to some rational factor that can't be less than -1 and can't be greater than 1, and we reach every such rational factor with some q_n. Therefore $\bigcup_{i=0}^{\infty} V_i = [0, 1]$, so we want the union of all the sets $\bigcup_{i=0}^{\infty} V_i$ to have measure 1.

However, since we got V_i just by shifting V_0, each V_i should individually have the same measure. So what could this measure be? It can't be 0, because otherwise the overall measure would be the sum of infinitely many 0s and couldn't add to 1. And it can't be any finite ϵ, because it's added together infinitely many times, so it would sum to infinity and therefore still can't add to 1.

Based on this, one of the following has to be false:

1. The Axiom of Choice.
2. Shifting a set keeps its measure the same (measures are translation-invariant).
3. Measures are additive under unions of disjoint sets.
4. We can measure all subsets of the real line.

We're going to keep the Axiom of Choice. Translation-invariance and being additive under unions are nice properties that we'd like our measure to have, and a measure that doesn't have them is not likely to be very useful. Therefore, it must not be the case that we can measure all subsets of the real line.

This leads us to the question: what is the structure of the subsets of the real line, or of any measure space, that we can measure?

6.2.2 Sigma-Algebras

Let Σ represent the family of subsets of X that we're able to measure. Then, a measure is a function $\mu : \Sigma \to [0, \infty)$. This choice makes sense, as we're essentially saying that the size of a subset must be at least zero. To construct the subsets that we can measure, let's list some properties that we'd like the family to have.

- Let's say that a set for which it makes sense to have size zero (measure zero) is the null set, so we can always measure the null set.
- We can also measure the entire set at once (even though this measure might be infinite).
- If we can measure a set, let's say we can measure everything not in the set (its complement): essentially, we're taking some known measure (the original set) out of the whole set, which we've said is measurable.
- If we add (take the union) two sets that don't have any overlap, we'd like the measure to be the sum of the two individual measures.
- If we intersect two sets, then we're measuring their overlap, so it makes sense to require that to be measurable.

This makes the family of subsets into a formal structure that we call a $\sigma-$algebra (sigma-algebra).

Definition of a σ-algebra

A σ-algebra over a set X is a collection Σ of subsets of X, with the property that if $A \in \Sigma$ then A is measurable, and the following closure properties:

1. The null set is measurable, and the entire space all at once is measurable. ($\varnothing \in \Sigma, X \in \Sigma$)
2. If we can measure a subset, we can measure its complement. ($A \in \Sigma \implies A^{\mathsf{c}} \in \Sigma$.)
3. If we can measure some countable number of sets, then we can measure their union and intersection. ($A_1, \ldots, A_n \in \Sigma \implies \cup_{i=1}^{n} A_i, \cap_{i=1}^{n} A_i \in \Sigma$.)

We can sidestep the problem of the Vitali set by simply saying that none of the V_is are in the $\sigma-$algebra we use on $\mathbb{R}$. The choice of $\sigma-$algebra is not unique: I can just say that $\varnothing$ and $\mathbb{R}$ are measurable, and that's a complete $\sigma-$algebra, but not a very useful one.

Note also that the $\sigma-$algebra is only closed under countable unions (and through the complement property and De Morgan's laws, this is equivalent to only being closed under countable intersections as well). This is because if we allowed an uncountable number of intersections, we'd have two possibilities:

1. We can construct the Vitali set by union-ing an uncountable number of singleton sets of the form $\{r\}$, which we don't want, or;
2. We can't measure a singleton set of the form $\{r\}$, in which case by countable intersections we can't measure any proper nontrivial subsets of $\mathbb{R}$ and we're just left with the trivial $\sigma-$algebra I defined above.

6.2.3 Defining a Measure Space

This gives us the complete definition of a measure space.

Definition of a measure space

A measure space is the combination of a set X, a $\sigma-$algebra Σ over X, and a measure μ, i.e. a function $\mu : \Sigma \to [0, \infty]$ such that

1. $\mu(\varnothing) = 0$
2. $A, B \in X$ disjoint $\implies \mu(A \cup B) = \mu(A) + \mu(B)$

These are properties that you might recognize from probability theory: the probability of the null set of no events happening is 0, and if A and B are independent disjoint events then the probability of both happening is just the sum of the probabilities that both happen individually.

We can extend the last property to arbitrary numbers of elements in X.

Countable additivity

Suppose $E_1, E_2, \ldots, E_n$ are a collection of sets in Σ that are pairwise disjoint, i.e. $E_i \cap E_j = \varnothing$ for $i \neq j$. Then for a valid measure μ,

$$\mu\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mu(E_i)$$

As a consequence of this, we can relax the constraint that all the E_is must be disjoint, if we're also willing to relax additivity being a perfect equality.

Countable subadditivity

For any countable sequence of $E_1, E_2, \dots \in \Sigma$,

$$\mu\left(\cup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mu(E_i)$$

Again, you might recognize this from probability as the union bound.

Further, we can disjointize sets if they're not already disjoint, which will allow us to use countable additivity.

Disjointizing

Given sets A and B, not necessarily disjoint, we can take $A' = A$ and $B' = B \setminus (A \cap B)$ and create disjoint sets with the same union as before. Then, by additivity of disjoint sets, we have

$$\mu(A') + \mu(B') = \mu(A \cup B)$$
$$\mu(A) + \mu(B \setminus (A \cap B)) = \mu(A \cup B)$$

Further, because $B \setminus (A \cap B)$ and $A \cap B$ are disjoint (because we made it like that), we can apply additivity of the measure again and get

$$\mu(B \setminus (A \cap B)) + \mu(A \cap B) = \mu((B \setminus (A \cap B)) \cup (A \cap B)) = \mu(B)$$

This gives us

$$\mu(A) + \mu(B) - \mu(A \cap B) = \mu(A \cup B)$$

We can recognize this as the inclusion-exclusion principle in probability.

6.3 Probability as a Measure

It's not surprising that we're seeing these similarities between general properties that we'd like a measure to have and the actual properties of a probability on a sample space.

Defining a probability on a sample space involves assigning probabilities to all possible events or combinations of events, which is essentially defining a function $\mathbb{P}$ from subsets of the sample space to $[0, 1] \subset \mathbb{R}^+$. This gives us relatively easily-understood examples of measure spaces.

Example of a probability measure space: flipping a coin

Suppose we flip a coin. The disjoint outcomes are $\Omega = \{H, T\}$, so the full sample space (the term for the $\sigma-$algebra of a probability measure space) to which we can assign probabilities is the power set (the set of all subsets) of these disjoint outcomes.

$$\mathcal{F} = \{\varnothing, \{H\}, \{T\}, \{H, T\}\}$$

Then, the probability measure assigns probabilities to each of these:

$$\mathbb{P}(\omega) = \begin{cases} 0 & \omega = \varnothing \\ \frac{1}{2} & \omega = \{H\} \\ \frac{1}{2} & \omega = \{T\} \\ 1 & \omega = \{H, T\} \end{cases}$$

In the discrete case, where Ω has finite order, the natural sample space is the set of all subsets, with order $2^{|\Omega|}$. If all the elements of Ω are disjoint, then additivity of the sample space means $\mathbb{P}$ can be completely specified by defining it on Ω.

Example of countable additivity of probabilities: rolling a die

If we roll a six-sided die, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. Since these outcomes are all disjoint, it suffices to say that $\mathbb{P}(\omega') = \frac{1}{6}$ for all $\omega' \in \Omega$. (Notation to be updated.) Then, if we're given any $\omega \in \mathcal{F}$, we can use additivity to find its probability. For example, if $\omega = \{2, 4, 6\}$,

$$\mathbb{P}(\omega) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

In words, the probability of getting an even result is $\frac{1}{2}$, as we want.
In general,

$$\mathbb{P}(\omega) = \frac{|\omega|}{6}.$$

I'll have a separate section (or maybe build it later into this one) on how to deal with probability-specific concepts like expectation and variance.

6.3.1 Constructing a Measure: the Lebesgue Measure on the Real Line

It's relatively easy to state in a vacuum some nice properties that we want a measure to have, but actually constructing a measure is more difficult. We're going to work through the construction of a measure on the reals, because we've already motivated why doing that is hard, and it's easier to do specifically for the reals.

In addition to the properties we already listed in the definition of a measure, let's say there are a few more we want specifically for the reals:

1. Extending length: any interval (a, b) has measure $b - a$.
2. Monotonicity: $A \subset B \subset \mathbb{R}$ implies $0 \leq \mu(A) \leq \mu(B) \leq \infty$.
3. Translation-invariance: this was a key part of the Vitali sets argument. If A is measurable in $\mathbb{R}$, define $A' = \{a + x_0 \mid a \in A\}$ for some fixed $x_0 \in \mathbb{R}$. Then $\mu(A) = \mu(A')$.

This is in addition to countable additivity and the null set having measure zero.

We've already built the idea of a $\sigma-$algebra, but for the most powerful theory, we want to build the maximal $\sigma-$algebra on the reals, that is, the largest possible $\sigma-$algebra satisfying the two properties of a measure and the three properties specific to the reals listed above. Of course, "largest" is sort of a weird word when we're defining the thing that determines relative size, but don't think about it too hard yet.

6.3.2 Constructing a Measure for General Spaces

7 Appendix

References

1. Mathematics 104 at UC Berkeley, Summer 2019, taught by Michael Christianson.
2. Mathematics 202A at UC Berkeley, Fall 2019, taught by Professor Marc Rieffel.
3. Real Analysis for Graduate Students, by Richard F. Bass. `http://bass.math.uconn.edu/rags010213.pdf`
4. Elementary Analysis, by Kenneth A. Ross. (Zero is a natural number, change my mind.)
5. Cauchy's Construction of $\mathbb{R}$, by Todd Kemp. `http://www.math.ucsd.edu/~tkemp/140A/Construction.of.R.pdf`
6. Renzo's Math 490: Introduction to Topology Lecture Notes. `https://www.math.colostate.edu/~renzo/teaching/Topology10/Notes.pdf`
7. Many answers on Math StackExchange, but particularly this complete construction of the reals (including the algebraic approach): `https://math.stackexchange.com/questions/11923/completion-of-rational-numbers-via-cauchy-sequences`
8. `https://www.math.unl.edu/~gmeisters1/papers/Measure/measure.pdf`
9. Stylistically inspired by Evan Chen's An Infinitely Large Napkin: `https://venhance.github.io/napkin/Napkin.pdf`

Acknowledgements

Thanks to Charlie Cummings, Abhinav Kejriwal, Shaunak Modak, Grant Posner, Shamay Samuel, Kanyes Thaker, and Ellen Thompson for their feedback, and to Kanyes Thaker for some LaTeX code (the coloured boxes). Thanks also to Shivani Kishnani for thinking this was a Math 104 textbook.