

1 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.

1.1 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.

1.2 Measures

1.2.1 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_i s 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?

1.2.2 Sigma-Algebras

Let Σ represent the family of subsets of X that we're able to measure. Then, a measure is a function $\mu : \Sigma \rightarrow [0, \infty)$. 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 σ -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. ($\emptyset \in \Sigma, X \in \Sigma$)
2. If we can measure a subset, we can measure its complement. ($A \in \Sigma \implies A^c \in \Sigma$)
3. If we can measure some countable number of sets, then we can measure their union and intersection. ($A_1, \dots, A_n \in \Sigma \implies \bigcup_{i=1}^n A_i, \bigcap_{i=1}^n A_i \in \Sigma$.)

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

Note also that the σ -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 σ -algebra I defined above.

1.2.3 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 σ -algebra Σ over X , and a measure μ , i.e. a function $\mu : \Sigma \rightarrow [0, \infty]$ such that

1. $\mu(\emptyset) = 0$
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, \dots, E_n are a collection of sets in Σ that are pairwise disjoint, i.e. $E_i \cap E_j = \emptyset$ 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_i s must be disjoint, if we're also willing to relax additivity being a perfect equality.

Countable subadditivity

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

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

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

$$\begin{aligned}\mu(A') + \mu(B') &= \mu(A \cup B) \\ \mu(A) + \mu(B \setminus (A \cap B)) &= \mu(A \cup B)\end{aligned}$$

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.

1.3 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 σ -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.

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

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

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

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.

1.3.1 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 σ -algebra, but for the most powerful theory, we want to build the maximal σ -algebra on the reals, that is, the largest possible σ -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.

1.3.2 Constructing a Measure for General Spaces