Measure Theory:
Why Intuitive Probability Isn't Enough

Laker Newhouse and Josh Zeitlin

July 12, 2024

Abstract

While an intuitive understanding of probability theory typically suffices, it fails to model many important scenarios. For example, what is the chance a dart thrown at the unit interval will land on a rational number? The traditional notion of a probability density function breaks down here. Mathematicians developed measure theory to resolve cases like this by giving a more rigorous notion of density. In this paper, we lay out the foundations of measure theory, discussing rings, algebras, and the Lebesgue measure, in order to solve the dart problem with our newfound understanding.

1 Why Measure Theory?

The traditional way to introduce probability theory in an undergraduate class, besides the straightforward finite case, revolves around the idea of a probability density function (PDF). For many cases, PDFs suffice. They let us calculate the chance that a plant will grow to have a height greater than one standard deviation above the mean, or the chance that a radioactive particle will have decayed after a thousand years. But consider the problem of throwing a dart at an interval:

Question 1.1. What is the chance a dart thrown randomly in $[0, 1]$ lands on a rational number?

Normally to resolve such a probability question boils down to integrating the probability density function, but what PDF can we use? The indicator function $1_{\mathbb{Q}}$ on the rationals, which maps $x \mapsto 1$ if $x \in \mathbb{Q}$ and $x \mapsto 0$ otherwise, captures what we mean, but how to integrate it? In other words, how can we find the size of the rational numbers compared to the real interval $[0, 1]$?

Perhaps the rationals occupy essentially the entire interval, because they are dense in the reals. Maybe they actually take up no space. Maybe the rationals account for precisely $\pi 2^{-10}$ percent of the unit interval. Without a rigorous sense of what length is, we cannot say. We need measure theory to answer these questions.

As we develop measure theory from the ground up, we take a more abstract approach than directly necessary to answer the question about the dart. We detour into sets and algebras in order to appreciate the generality measure theory can offer. Our goal is the so-called Lebesgue measure, which once defined will allow us to declare with certainty the size of many subsets of $\mathbb{R}$. By then, we'll be able to measure the size of $\mathbb{Q}$ within the unit interval $[0, 1]$ to answer how often the dart will land on a rational number.

2 Rings and Algebras

Measure theory is all about measures, which are functions from the power set of a set M to $\mathbb{R}$. To solve the dart problem, we eventually define the so-called Lebesgue measure on $\mathbb{R}^n$, but at first we consider measures in the more abstract setting of pure set theory. That is, we seek to define a sensible concept of measure for all sets, not just $\mathbb{R}$ and its variants.

The path to doing so begins with two structures called rings and algebras, which despite the strong connotation have nothing to do with ring theory in modern algebra. In this section, we develop the notion of a ring and of an algebra as it relates to measures.

2.1 Preliminary Measures

To begin, let M be an arbitrary, non-empty set and S be a family of subsets of M. A measure, typically written μ, is a function that maps members of S to nonnegative real numbers. The measure is the new notion of size that motivates measure theory. One useful property a measure μ can have is to be *additive*.

Definition 2.1. A functional $\mu : S \to \mathbb{R}_+$ is called a *σ-additive measure* if

$$\mu(A) = \sum_{k=1}^{N} \mu(A_k)$$

for any set $A \in S$ that is a disjoint union of a countable sequence $\{A_k\}_{k=1}^{N}$, where N is either finite or $N = \infty$. If the property holds but not for $N = \infty$, then we say μ is a *finitely additive measure*.

In words, a measure is finitely additive when the expected occurs: taking two disjoint parts of the container set and combining them yields a measure equal to the sum of the measures of the two parts. A measure is σ-additive when we can do the same for countably large disjoint unions. Sometimes a measure is not quite σ-additive, but it still has a useful, less strict property.

Definition 2.2. A functional $\mu : S \to \mathbb{R}_+$ is called a *σ-subadditive measure* if

$$\mu(A) \leq \sum_{k=1}^{N} \mu(A_k)$$

whenever $A \subset \bigcup_{k=1}^{N} A_k$, where $A \in S$ and all $A_k \in S$ and N is finite or infinite.

Example 2.3. Consider the family of bounded intervals in $\mathbb{R}$. If we define a functional l to be the length $b - a$ of any such interval $[a, b]$, then l is a σ-additive measure. The proof requires topological tools we do not develop here, so we just outline it.

Proof Sketch. Consider a disjoint union $I = \bigsqcup_{k=1}^{\infty} I_k$, where each I_k is a bounded interval. We show that

$$l(I) \geq \sum_{k=1}^{\infty} l(I_k) \quad \text{and} \quad l(I) \leq \sum_{k=1}^{\infty} l(I_k).$$

For the first inequality, note that for any $n \in \mathbb{N}$, we have $I \supset \bigsqcup_{k=1}^{n} I_k$. By finite additivity, we therefore get

$$l(I) \geq \sum_{k=1}^{n} l(I_k).$$

Letting n tend to ∞ yields the first inequality. To show the second inequality, which is equivalent to showing that l is σ-subadditive, is the tricky part. Without getting too far into the details, we construct an open cover of I using open intervals $\{J_k\}_{k=1}^{\infty}$. Fix $\epsilon > 0$. For each $k \in \mathbb{N}$, we define $J_k \supset I_k$, where

$$l(J_k) < l(I_k) + \frac{\epsilon}{2^k}.$$

We can do this assuming that each I_k has a finite length, which is sensible, because otherwise clearly

$$l(I) = \infty = \sum_{k=1}^{\infty} l(I_k).$$

The Heine-Borel theorem from topology guarantees that we can find a *finite* subset of $\{J_k\}_{k=1}^{\infty}$ that still covers I. Renumber the I_k so that the first N of them correspond to the N open intervals J_k given by the Heine-Borel theorem. Pick some closed interval $[a, b] \subset I$ such that $b - a > l(I) - \epsilon$. (If $l(I) = \infty$, choose any large length interval $[a, b]$.) Now the finite subadditivity of l gives us

$$\begin{aligned} b - a &\leq \sum_{k=1}^{N} l(J_k) \\ &\leq \sum_{k=1}^{N} l(I_k) + \frac{\epsilon}{2^k} \\ &= \sum_{k=1}^{N} l(I_k) + \sum_{k=1}^{N} \frac{\epsilon}{2^k} \\ &\leq \sum_{k=1}^{\infty} l(I_k) + \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} \\ &= \epsilon + \sum_{k=1}^{\infty} l(I_k). \end{aligned}$$

Because $l(I) - \epsilon \leq b - a$, the transitive property gives

$$l(I) - \epsilon \leq \epsilon + \sum_{k=1}^{\infty} l(I_k).$$

As ϵ tends to 0, we get $l(I) \leq \sum_{k=1}^{\infty} l(I_k)$, completing the proof sketch. □

2.2 Semi-rings and Rings

Currently, our notion of measure is based quite generally on an arbitrary family S of subsets of some nonempty set M. It is now the time to narrow our focus. We introduce two structures on sets called semi-rings and rings, to which we will limit our family of subsets S. Contrary to what one might expect from the names, these structures have little or no relation to the rings and semi-rings of abstract algebra. Our goal will then become to extend an arbitrary σ-additive measure on a semi-ring to an equivalent one on a ring.

Definition 2.4. A family S of subsets of a nonempty set M is called a *semi-ring* if

1. S contains $\emptyset$.
2. For $A, B \in S$, the intersection $A \cap B$ is in S.
3. For $A, B \in S$, the difference $A \setminus B$ is a disjoint union of a finite subset of S.

For example, if $M = \mathbb{R}$, then the family of subsets S corresponding to all intervals is a semi-ring. The intersection of two intervals is an interval, even if one or both are half-open. In addition, the difference of two intervals is either one new interval or the union of two intervals now separated by a gap. Therefore S is a semi-ring. For another example, if M is arbitrary and nonempty, then its power set forms a semi-ring. In fact, the power set of M is more than just a semi-ring.

Definition 2.5. A family of subsets S of M is called a *ring* if

1. S contains $\emptyset$.
2. For $A, B \in S$, the union $A \cup B$ is in S.
3. For $A, B \in S$, the difference $A \setminus B$ is in S.

Unlike in a semi-ring, the difference of two sets in a ring S must be in S. Additionally, the union of two sets in S must be in S. Although the intersection of $A, B \in S$ is not explicitly stated to be in S, the fact follows from writing

$$A \cap B = A \setminus (A \setminus B).$$

Thus, every ring is a semi-ring. Returning to our example from earlier, the power set of an arbitrary nonempty set M is a ring. Just as we extended additivity of measures to a more general σ-additivity, we can extend the notion of a ring to include closure under countable unions of members of S.

Definition 2.6. A ring S is called a *σ-ring* if the union of any countable family $\{A_k\}_{k=1}^{\infty}$ is in S.

Not only is the power set $\mathcal{P}(M)$ of M a ring, it is in fact a σ-ring. Indeed, every subset of M is contained in $\mathcal{P}(M)$, so any arbitrary union must fall within $\mathcal{P}(M)$, too. It follows as well that countable intersections are closed in a σ-ring. For any $A = \bigcap_k A_k$ for $A_k \in S$, choose $B = A_1$. Then $B \supset A$, so

$$A = B \setminus (B \setminus A) = B \setminus \left(B \setminus \bigcap_k A_k \right) = B \setminus \left(\bigcup_k (B \setminus A_k) \right) \in S.$$

Soon we will see how to extend a σ-additive measure μ from a ring to a σ-ring, but first we must show how to extend μ from a semi-ring to a ring. The idea of a minimal ring will be crucial.

Definition 2.7. For a nonempty set M and a family of subsets S, the *minimal ring containing S* is the intersection of all rings containing S. Similarly, the *minimal σ-ring containing S* is the intersection of all σ-rings containing S.

Why must such an intersection even yield a ring? For any family $\{S_\alpha\}$ of rings (or σ-rings), call the intersection $S = \bigcap_\alpha S_\alpha$. Clearly $\emptyset \in S$ because $\emptyset \in S_\alpha$ for each ring S_α. Suppose $A, B \in S$. Then $A \cup B \in S_\alpha$ for each α because $A, B \in S_\alpha$. Similarly, we see that $A \setminus B \in S_\alpha$ for each α because $A, B \in S_\alpha$. Given that $A \cup B$ and $A \setminus B$ are in S_α for each α, they are in the intersection, too. Furthermore, there is always at least one ring that contains S, namely the power set $\mathcal{P}(M)$. Thus, the minimal ring containing S always exists.

For an example, suppose S is the semi-ring of all intervals on $\mathbb{R}$. Then the minimal ring $R(S)$ consists of all finite disjoint unions of intervals on $\mathbb{R}$. The concept of length extends to all these finite unions, because for any disjoint union of intervals I composed of $\{I_k\}_{k=1}^{n}$, we have

$$l(I) = \sum_{k=1}^{n} l(I_k).$$

We can use the sum above to extend l on S to a new measure on $R(S)$. In fact, this extension is not limited to intervals on $\mathbb{R}$. Using the minimal ring, we can extend an arbitrary σ-additive measure uniquely from a semi-ring to a ring. The following theorem makes the result explicit.

Theorem 2.8. *Let S be a semi-ring.*

a) The minimal ring $R(S)$ consists of all finite disjoint unions of sets from S.

b) If μ is a finitely additive measure on S, then μ extends uniquely to a finitely additive measure on $R(S)$.

c) If μ is σ-additive, then its extension to $R(S)$ is also σ-additive.

Proof. a) Let S' be the family of sets consisting of all finite disjoint unions of sets of S, meaning

$$S' = \left\{ \bigsqcup_{k=1}^{n} A_k : A_k \in S, n \in \mathbb{N} \right\}.$$

Our goal is to show $S' = R(S)$. It suffices to show that S' is a ring. If that were so, then because S' contains S, it would also contain $R(S)$. In addition, because $R(S)$ is a ring, we already know that it will contain all finite disjoint unions of elements of S. The preceding containments prove that $S' = R(S)$ if S' is a ring. We break down the proof into four steps.

Step 1. If $A, B \in S'$ and A, B are disjoint, then $A \sqcup B \in S'$. We know so because we can write $A = \bigsqcup_{k=1}^{n} A_k$ and $B = \bigsqcup_{j=1}^{m} B_j$ for mutually disjoint $A_k, B_j \in S$, and then we can write $A \sqcup B$ as a disjoint union of all the A_k with the all B_j.

Step 2. If $A, B \in S'$, then $A \cap B \in S'$. This follows from writing

$$A \cap B = \bigsqcup_{k,j} A_k \cap B_j.$$

Because S' is a semi-ring, it is closed under intersection. Thus, the finite disjoint union of the intersections is also in S'.

Step 3. If $A, B \in S'$, then $A \setminus B \in S'$. We reduce the claim to simpler components in order to prove it. First, write

$$A \setminus B = \bigsqcup_{k} A_k \setminus B$$

so that, by Step 1, we need only prove that $A_k \setminus B \in S'$. Then write

$$A_k \setminus B = A_k \setminus \bigsqcup_{j} B_j = \bigcap_{j} A_k \setminus B_j.$$

so that, by Step 2, we need only prove that $A_k \setminus B_j \in S'$. Because A_k and B_j are in S, which is a semi-ring, by definition $A_k \setminus B_j$ is a disjoint union of a finite family of sets from S, proving that $A \setminus B \in S'$.

Step 4. If $A, B \in S'$, then $A \cup B \in S'$. We get the result by decomposing

$$A \cup B = (A \setminus B) \sqcup (B \setminus A) \sqcup (A \cap B),$$

each part of which is in S' by Steps 2 and 3, and the disjoint union of which is in S' by Step 1.

By Steps 3 and 4, we conclude that S' is a ring, which completes the proof.

b) Now let μ be a finitely additive measure on S. We seek the extend the measure uniquely to the minimal ring $R(S)$. Indeed, for $A \in R(S)$, we know that $A = \bigsqcup_{k=1}^{n} A_k$ for some $A_k \in S$. Define

$$\mu(A) = \sum_{k=1}^{n} \mu(A_k).$$

Let us prove that μ as defined is finitely additive. Before anything, we show that μ does not depend on a choice of how to split $A = \bigsqcup_{k=1}^{n} A_k$. Suppose we have another splitting $A = \bigsqcup_{j=1}^{m} B_j$. Then

$$A_k = \bigsqcup_{j} A_k \cap B_j,$$

and because $A_k \cap B_j \in S$ and μ is finitely additive on S, we get

$$\mu(A_k) = \sum_j \mu(A_k \cap B_j).$$

Summing over k, we obtain

$$\sum_k \mu(A_k) = \sum_k \sum_j \mu(A_k \cap B_j).$$

Similarly,

$$\sum_j \mu(B_j) = \sum_j \sum_k \mu(B_j \cap A_k).$$

But then clearly $\sum_k \mu(A_k) = \sum_j \mu(B_j)$, so the way we split A does not matter.

Now we prove the finite additivity of μ over $R(S)$. Let A and B be two disjoint sets from $R(S)$. Assume that $A = \bigsqcup_{k=1}^n A_k$ and $B = \bigsqcup_{j=1}^m B_j$, where $A_k, B_j \in S$. By choosing another way to split $A \sqcup B$, we get

$$\mu(A \sqcup B) = \mu\left(\left(\bigsqcup_{k=1}^n A_k\right) \sqcup \left(\bigsqcup_{j=1}^m B_j\right)\right) = \sum_k \mu(A_k) + \sum_j \mu(B_k) = \mu(A) + \mu(B).$$

For any finite family of disjoint sets $C_1, \ldots, C_n \in R(S)$, we can use induction in n along with the property that unions of sets in $R(S)$ are in $R(S)$ to show that

$$\mu\left(\bigsqcup_k C_k\right) = \sum_k \mu(C_k).$$

c) Suppose that $A = \bigsqcup_{j=1}^\infty B_j$, where $A, B_j \in R(S)$. We need to prove that

$$\mu(A) = \sum_{j=1}^\infty \mu(B_j).$$

We define finite counterparts. Write $A = \bigsqcup_k A_k$ and $B_l = \bigsqcup_m B_{jm}$, where the sums in k and m are finite and where A_k and B_{jm} are in S. Define

$$C_{kjm} = A_k \cap B_{jm},$$

whereupon $C_{kjm} \in S$. Then we have

$$A_k = A_k \cap A = A_k \cap \bigsqcup_{j,m} B_{jm} = \bigsqcup_{j,m} (A_k \cap B_{lm}) = \bigsqcup_{lm} C_{kjm},$$
$$B_{jm} = B_{jm} \cap A = B_{jm} \cap \bigsqcup_k A_k = \bigsqcup_k (A_k \cap B_{lm}) = \bigsqcup_k C_{kjm}.$$

By the σ-additivity of μ on S, we get

$$\mu(A_k) = \sum_{l,m} \mu(C_{kjm}),$$
$$\mu(B_{jm}) = \sum_k \mu(C_{kjm}).$$

It follows that

$$\sum_k \mu(A_k) = \sum_{k,l,m} \mu(C_{kjm}) = \sum_{j,m} \mu(B_{jm}).$$

On the other hand, by the definition of μ on $R(S)$, we have

$$\mu(A) = \sum_k \mu(A_k),$$
$$\mu(B_l) = \sum_m \mu(B_{jm}).$$

Then, summing over j,

$$\sum_j \mu(B_j) = \sum_{j,m} \mu(B_{jm}).$$

Combining the above results, we complete the proof with

$$\mu(A) = \sum_k \mu(A_k) = \sum_{j,m} \mu(B_{jm}) = \sum_j \mu(B_j).$$

□

2.3 Algebras

Having extended a σ-additive measure from a semi-ring to a ring, our principal goal now is to extend a σ-additive measure from a ring to a σ-ring. After all, a σ-ring would be a natural domain for a σ-additive measure. The majority of the work occurs in the next sections, although we begin here by defining the notion of an algebra.

So far we have seen only a few examples of σ-rings: $S = \{\emptyset\}$ and $S = \mathcal{P}(M)$. We can obtain implicit σ-rings with an analogy of the minimal ring.

Definition 2.9. For any family S of subsets of M, denote by $\Sigma(S)$ the intersection of all σ-rings containing S.

Because $\mathcal{P}(M)$ is always a σ-ring containing S, and because the intersection of any number of σ-rings remains a σ-ring, the set $\Sigma(S)$ is in fact a σ-ring. We call $\Sigma(S)$ the *minimal σ-ring containing S*.

Most rings we have seen are not σ-rings. For example, the ring of finite disjoint unions of intervals is not a σ-ring, because it does not contain the countable union

$$\bigcup_{k=1}^{\infty} (k, k+1/2).$$

In general, an explicit description of the minimal σ-ring is difficult to produce. Counterexamples such as the *Cantor set* invalidate the most natural guesses, for example that $\Sigma(S)$ would consist of countable disjoint unions of sets from S. The actual construction requires the notion of *transfinite induction*, whereby we perform an inductive step uncountably many times. We do not digress into the details of that procedure here. We would be right to imagine that extending a measure on a ring to a measure on a σ-ring is quite complicated. We succeed in this task in the coming sections, after one final definition here to distinguish an important kind of ring.

Definition 2.10. If a ring contains the entire set M as an element, then we refer to it as an *algebra*. Similarly, if a σ-ring contains the entire set M, we call it a *σ-algebra*.

3 Measures

In this section, we will develop the notion of a measure further. Going forward, let R be an algebra and let M be a random nonempty set.

3.1 Outer Measure

Assuming that R is an algebra and μ is a σ-additive measure on R, for any subset $A \subset M$, we can now define its outer measure.

Definition 3.1. The outer measure of A is defined as:

$$\mu^*(A) = \inf \left\{ \sum_{k=1}^{\infty} \mu(A_k) : A_k \in R \text{ and } A \subset \bigcup_{k=1}^{\infty} A_k \right\}$$

For a moment, let us diverge from our progression to discuss an important theme not only in measure theory but throughout mathematics. The outer measure being just one example, definitions are carefully chosen. The reason the outer measure cannot suffice as our ultimate destination is that it has certain flaws. For example, there is a set $C \cap [0,1]$ with outer measure 1 whose complement $C^C \cap [0,1]$ also has outer measure 1, a phenomenon that our intuition says shouldn't occur. But it does. We do not delve into the details of the construction here, because it would take us too far off track. The interested reader may consult [2]. Nonetheless, we need the outer measure to build up better measures. Suffice it to say that though the modern definitions given throughout are clean, impeccable, and sometimes inscrutable, they have good reason for existing. An in-depth study of the history and of the relevant counterexamples can reward newfound appreciation for why definitions take the form they do. Unfortunately, that study is out of the scope of the discussion here.

Returning to our analysis, we can see that μ^* satisfies $A \subset B \to \mu^*(A) < \mu^*(B)$. Now, we are going to show that the outer measure is finite.

Theorem 3.2. *If $A \subset M$ then $\mu^*(A) < \infty$.*

Proof. Consider a sequence $A_k = \{M, \emptyset, \emptyset, \emptyset, \dots\}$. Then it must be so that $\mu^*(A) \leq \mu(M) + \mu(\emptyset) + \cdots = \mu(M) < \infty$ because this is the "maximal covering." □

Building off of this, we can also see that if A is in an algebra, then $\mu(A) = \mu^*(A)$ because if we take the covering $\{A_k\} = \{A, \emptyset, \emptyset, \emptyset, \dots\}$, we get that $\mu^*(A) \leq \mu(A)$ and by sub-additivity we can take the infimum over all other sequences to get that $\mu^*(A) = \mu(A)$. One final property that is important to know is that it is σ-subadditive over the power set and likewise, for any two sets $A, B \in M$ we get that the difference of the outer measures of those two sets is bounded by the outer measure of the symmetric difference of those two sets. The symmetric difference of A and B is

$$A \triangle B = (A \cup B) \setminus (A \cap B),$$

which is sometimes called the XOR operation.

3.2 Measurable Sets

Before we move on to Carathéodory's Extension Theorem, we are going to introduce the concept of a measurable set.

Definition 3.3. Given a subset $A \subset M$, we say A is *measurable* if, for any $\epsilon > 0$, there exists $B \in R$ such that $\mu^*(A) - \mu^*(B) < \epsilon$.

What this definition ostensibly says is that a set is measurable if we can approximate sets in an algebra infinitely closely.

3.3 Carathéodory's Extension Theorem

In this section we will talk about Carathéodory's Extension Theorem which is quite an important theorem in measure theory. It lets us extend measures and outer measures on algebras from a given ring. Let's write out the extension theorem:

Theorem 3.4. *Let R be an algebra on a set M and μ be a σ-additive measure on R. Let $\mathcal{M}$ denote the family of all measurable subset of M. Then:*

1. *$\mathcal{M}$ is a σ-algebra and $R \subset \mathcal{M}$.*
2. *The restriction of μ^* on $\mathcal{M}$ is a σ-additive measure (that extends measure μ from R to $\mathcal{M}$).*
3. *If $\tilde{\mu}$ is a σ-additive measure defined on a σ-algebra Σ such that $R \subset \Sigma \subset \mathcal{M}$, then $\tilde{\mu} = \mu^*$ on Σ.*

Before I prove this, I will parse this information so we can get a more conceptual understanding of this theorem. The first two statements of this theorem suggest that a σ-additive measure can be extended from the algebra R to the σ-algebra of all measurable sets $\mathcal{M}$. Furthermore, part 3 shows that if $\Sigma = \mathcal{M}$, the extension/restriction is unique. Now let's do the proof.

Proof. We'll prove this, by separating this into several claims.

1. The family of all measurable sets $\mathcal{M}$, is an algebra containing R.

   Proof. If $A \in R$, then A is measurable, because

   $$\mu^*(A \triangle A) = 0 \text{ as } \mu^*(A) = \mu(A)$$

   if A is in an algebra. Thus, $R \subset \mathcal{M}$ and thus M is a measurable set. In our quest to show that $\mathcal{M}$ is an algebra, we can show that if $A_1, A_2 \in \mathcal{M}$ then $A_1 \cup A_2$ and $A_1 \ A_2$ are also measurable sets. Looking at A_1 and A_2 we know that there must be sets H_1 and B_2 such that

   $$\mu^*(A_1 \triangle B_1) < \epsilon \text{ and } \mu^*(A_2 \triangle B_2) < \epsilon$$

   If we let $B = B_1 \cup B_2 \in R$, we can get that

   $$A \triangle B \subset (A_1 \triangle B_1) \cup (A_2 \triangle B_2)$$

   and because μ^* also has the property of subadditivity, we get that

   $$\mu^*(A \triangle B) \leq \mu^*(A_1 \triangle B_1) + \mu^*(A_2 \triangle B_2) < 2\epsilon$$

   Because ϵ is arbitrary and $B \in R$ this means that A satisfies the definition of a measurable set. □

2. μ^* is σ-additive on $\mathcal{M}$

   Proof. As we have just established that $\mathcal{M}$ is an algebra and μ^* is σ-subadditive, so to prove that it is σ-additive, we can just proof that μ^* is finitely additive on $\mathcal{M}$. Now, we want to show that if $A = A_1 \sqcup A_2$ we get that

   $$\mu^*(A) = \mu^*(A_1) + \mu^*(A_2)$$

   We know that

   $$\mu^*(A) \leq \mu^*(A_1) + \mu^*(A_2)$$

so that means we have to show the inequality in the opposite direction, so

$$\mu^*(A) \geq \mu^*(A_1) + \mu^*(A_2)$$

We again know that for any $\epsilon > 0$ there are sets $B_1, B_2 \in R$ such that

$$\mu^*(A_1 \triangle B_1) < \epsilon \text{ and } \mu^*(A_2 \triangle B_2) < \epsilon$$

is true. Letting $B = B_1 \cup B_2 \in R$ we can use our inequality about outer measures and symmetric differences to get that

$$|\mu^*(A) - \mu^*(B)| \leq \mu^*(A \triangle B) < 2\epsilon \text{ and thus } \mu^*(A) \geq \mu^*(B) - 2\epsilon$$

Then, since $B \in R$ we get that

$$\mu^*(B) = \mu(B) = \mu(B_1) + \mu(B_2) - \mu(B_1 \cap B_2)$$

so for both $k = 1, 2$ we get that

$$|\mu^*(A_k) - \mu^*(B_k)| \leq \mu^*(A_k \triangle B_k) < \epsilon \text{ because } \mu(B_k) \geq \mu^*(A_k) - \epsilon.$$

Looking at $B_1 \cap B_2$, we get that

$$\mu(B_1 \cap B_2) = \mu^*(B_1 \cap B_2) \leq \mu^*(A_1 \triangle B_1) + \mu^*(A_2 \triangle B_2) < 2\epsilon$$

so combining all these terms,

$$\mu^*(A) \geq \mu^*(A_1) + \mu^*(A_2) - 6\epsilon \text{ and as } \epsilon \to 0$$

the proof is completed and thus $\mu^*(A) = \mu^*(A_1) + \mu^*(A_2)$ and thus $\mu^*(A)$ is σ-additive. □

3. $\mathcal{M}$ is a σ-algebra.

Proof. We know that $\mathcal{M}$ is an algebra and μ^* is σ-additive, so let's now show that $\mathcal{M}$ is a σ-algebra. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of measurable sets, let's show that $\bigcup_{n=1}^{\infty} A_n$ is measurable to complete our proof. To do this, let

$$\tilde{A_n} = A_n \backslash A_{n-1} \backslash \dots \backslash A_1 \in \mathcal{M}$$

and we can see that

$$A = A_1 \backslash (A_2 \backslash A_1) \cup (A_3 \backslash A_2 \backslash A_1) \dots = \sqcup_{n=1}^{\infty} \tilde{A_n}$$

Thus, renaming $\tilde{A_n}$ to be A_n, we see that it suffices to treat the case of a disjoint union: given $A = \bigcup_{n=1}^{\infty} A_n$ where $A_n \in \mathcal{M}$, prove that $A \in \mathcal{M}$. We get that for any fixed N,

$$\mu^*(A) \geq \mu^*(\sqcup_{n=1}^{N} A_n) = \sum_{n=1}^{N} \mu^*(A_n)$$

by additivity and monotonicity. Thus, letting $N \to \infty$, we get that

$$\sum_{n=1}^{\infty} \mu^*(A_n) \leq \mu^*(A) < \infty$$

so

$$\sum_{n=1}^{\infty} \mu^*(A_n)$$

converges. Thus, for any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that

$$\sum_{n=N+1}^{\infty} \mu^*(A_n) < \epsilon$$

By letting $A^{'} = \sqcup_{n=1}^{N} A_n$ and $A^{''} = \sqcup_{n=N+1}^{\infty} A_n$ we obtain by the σ-subadditivity of μ^* that

$$\mu^*(A^{''}) \leq \sum_{n=N+1}^{\infty} \mu^*(A_n) < \epsilon$$

Using our first claim that $\mathcal{M}$ is an algebra, we get that $A^{'}$ is measurable and can be written as a finite union of measurable sets. Thus, there exists some $B \in R$ such that

$$\mu^*(A^{'} \triangle B) < \epsilon$$

Because $A = A^{'} \cup A^{''}$, we have that

$$A \triangle B \subset (A^{'} \triangle B) \cup A^{''}$$

Now, if $x \in A^{'}$ then with $x \notin B$, it gives

$$x \in A^{'} \triangle B \subset (A^{'} \triangle B) \cup A^{''}$$

If $x \in A^{''}$ then the inclusion is quite obvious. Moving on, we get that

$$\mu^*(A \triangle B) \leq \mu^*(A^{'} \triangle B) + \mu^*(A^{''}) < 2\epsilon$$

This thus implies that $A \in \mathcal{M}$ and thus $\mathcal{M}$ is a σ-algebra. □

Now, for our final claim, we will talk about the extensions and restrictions.

4. Let Σ be a σ-algebra such that $R \subset \Sigma \subset \mathcal{M}$ and let $\tilde{\mu}$ be a σ-additive measure on Σ such that $\tilde{\mu} = \mu$ on R. Then, $\tilde{\mu} = \mu^*$ on Σ.

   Proof. We want to prove that $\tilde{(\mu)}(A) = \mu^*(A)$ for any $A \in \Sigma$. By the way in which we defined the outer measure μ^*, we can apply the σ-subadditivity of $\tilde{\mu}$ and obtain that

   $$\tilde{\mu}(A) \leq \sum_{n=1}^{\infty} \tilde{\mu}(A_n) = \sum_{n=1}^{\infty} \mu(A_n)$$

   Taking the infimum over all such sequences $\{A_n\}$, we obtain that

   $$\tilde{\mu}(A) \leq \mu^*(A)$$

   However, because A is measurable, for any $\epsilon > 0$ there is some $B \in R$ such that

   $$\mu^*(A \triangle B) < \epsilon$$

   By our early inequality, we can then get that

   $$|\mu^*(A) - \mu^*(B)| \leq \mu^*(A \triangle B) < \epsilon$$

   Combining these inequalities, we can use that

   $$\tilde{\mu}(B) = \mu(B) = \mu^*(B)$$

   where we get that

   $$|\tilde{\mu}(A) - \mu^*(A)| \leq |\mu^*(A) - \mu^*(B)| + |\tilde{\mu}(A) - \tilde{\mu}(B)| < 2\epsilon$$

   Letting ϵ go to 0, we can conclude that $\tilde{\mu}(A) = \mu^*(A)$, which finishes the proof. □

□

3.4 σ-finite Measures

In this next section, we will discuss σ-finite measures. Let's begin with the abstract definition of a measure.

Definition 3.5. Let M be a non-empty set and S be a family of subsets of M. A functional $\mu : S \to [0, +\infty]$ is called a *measure* if, for all $A, A_k \in S$ such that $A = \sqcup_{k=1}^{N} A_k$ (where N is either finite or infinite), we have $\mu(A) = \sum_{k=1}^{N} \mu(A_k)$

Moving on, let's define another concept relating to a measure.

Definition 3.6. A measure μ on a set S is called *finite* if $M \in S$ and $\mu(M) < \infty$. A measure μ is called σ-finite if there is a sequence $\{B_k\}_{k=1}^{\infty}$ of sets from S such that $\mu(B_k) < \infty$ and $M = \cup_{k=1}^{\infty} B_k$

We get that any finite measure is also a σ-finite measure. Finally, we'll introduce one more definition before we go to our main theorem for this section.

Definition 3.7. A set $A \in M$ is called measurable if $A \cap B_k \in \mathcal{M}_{B_k}$ for any k. For any measurable set A, set $\mu_M(A) = \sum_k \mu_{B_k}(A \cap B_k)$

Now, for our main theorem.

Theorem 3.8. *Let μ be a σ-finite measure on a ring R and $\mathcal{M}$ be the family of all measurable sets defined above. Then the following is true. Here, R_{B_k} is in algebra in B_k.*

1. *$\mathcal{M}$ is a σ-algebra containing R.*
2. *The functional μ_M which we defined in the definition of a measurable set, is a measure on $\mathcal{M}$ that extends measure μ on R.*
3. *If $\tilde{\mu}$ is a measure defined on a σ-algebra Σ such that $R \subset \Sigma \subset \mathcal{M}$, then $\tilde{\mu} = \mu_M$ on Σ.*

Proof. 1. To prove our first, claim, we see that if $A \in R$, then $A \cap B_k \in R$ because $B_k \in R$ as well. Hence, $A \cap B_k \in R_{B_k}$ whence $A \cap B_k \in \mathcal{M}_{B_k}$ and $A \in \mathcal{M}$. Hence, $R \subset \mathcal{M}$ Then, we can show that if $A = \cup_{n=1}^{N} A_n$ where $A_n \in \mathcal{M}$ then also $A \in \mathcal{M}$ (where N can be ∞). Then, we have

$$A \cap B_k = \cup_n (A_n \cap B_k) \in \mathcal{M}_{B_k}$$

because

$$A_n \cap B_k \in \mathcal{M}_{B_k}$$

and $\mathcal{M}_{B_k}$ is a σ-algebra. Thus, $A \in \mathcal{M}$. By the same manner, if

$$A^{'}, A^{''} \in \mathcal{M}$$

then the difference

$$A = A^{'} \backslash A^{''}$$

belongs to $\mathcal{M}$ because

$$A \cap B_k = (A^{'} \cap B_k) \backslash (A^{''} \cap B_k) \in \mathcal{M}_{B_k}$$

Lastly, because $M \in \mathcal{M}$ because

$$M \cap B_k = B_k \in \mathcal{M}_{B_k}$$

Thus, $\mathcal{M}$ must be a σ-algebra.

2. If $A \in R$ then $A \cap B_k \in R_{B_k}$ whence

$$\mu_{B_k}(A \cap B_k) = \mu(A \cap B_k)$$

Since

$$A = \sqcup(A \cap B_k)$$

and μ is a measure on R, we obtain

$$\mu(A) = \sum_k \mu(A \cap B_k) = \sum_k \mu_{B_k}(A \cap B_k)$$

Now, from our definition above, we can get that $\mu_M(A) = \mu(A)$. Hence, μ_M on R coincides with μ. Now, we can show that μ_M is a measure. Let

$$A = \sqcup_{n=1}^N A_n$$

where

$$A_n \in \mathcal{M}$$

and N is either finite or infinite. We want to show that

$$\mu_M(A) = \sum_n \mu_M(A_n)$$

Then, we get that

$$\sum_n \mu_M(A_n) = \sum_n \sum_k \mu_{B_k}(A_n \cap B_k)$$

and this then becomes

$$\sum_k \mu_{B_k}(\sqcup_n(A_n \cap B_k))$$

and then this becomes $\mu_M(A)$, where we use our identity of

$$A \cap B_k = \sqcup_n(A_n \cap B_k)$$

along with the σ-additivity of the measure μ_{B_k}.

3. Let $\tilde{\mu}$ be another measure defined on a σ-algebra Σ such that

$$R \subset \Sigma \subset \mathcal{M}$$

and $\tilde{\mu} = \mu$ or R. Let us prove that $\tilde{\mu}(A) = \mu_M(A)$ for any $A \in \Sigma$. Then, we can get that given $A \in \Sigma$ we get that $A = \sqcup_k(A \cap B_k)$ and $A \cap B_k \in \Sigma_{B_k}$ whence

$$\tilde{\mu}(A) = \sum \tilde{\mu}(A \cap B_k) = \sum \mu_M(A)$$

and thus we are done and the proof is finished.

□

3.5 Lebesgue Measure

In this section, we are going to try and define a new measure. Let S_1 be the semi-ring of all bounded intervals in $\mathbb{R}$, then define S_n to be $S_n = S_1^n$. Thus, any set $A \in S_n$ takes the form $I_1 \times \cdots \times I_n$, where I_k are bounded intervals in $\mathbb{R}$. Now, in order to introduce our measure, let's define the generalized notion of the product measure.

Lemma 3.9. *Let $M = M_1 \times M_2 = \{(x,y) : x \in M_1, y \in M_2\}$ and the family S of subsets of M, defined $S = S_1 \times S_2 = \{A \times B : A \in S_1, B \in S_2\}$. Now, if S_1 and S_2 are semi-rings, then S is also a semi-ring.*

Now, let's define a product measure

Definition 3.10. If μ_1 is a finitely additive measure on the semi-ring S_1 and μ_2 is a finitely additive measure on the semi-ring S_2, then the *Product Measure* $\mu = \mu_1 \times \mu_2$ on S is defined as follows: given $A \in S_1$ and $B \in S_2$, $\mu(A \times B) = \mu_1(A)\mu_2(B)$

Now, we can also show that μ is a finitely additive measure on S. We clearly get that S_n is a semi-ring as the product of semi-rings. Now, the product measure λ_n on S_n can be defined by $\lambda_m = l \times \cdots \times l$ where l is the length on S_1. Thus, $\lambda_n(A) = l(I_1) \cdots l(I_n)$. Thus, λ_n is a finitely additive measure on the semi-ring $S_n \in \mathbb{R}^n$. We can see that this is σ-additive.

Lemma 3.11. *The product measure λ_n as defined above is a σ-additive measure on S_n.*

The proof is left to the reader, however, it just requires unpacking a few definitions. We also know that λ_n is a σ-finite measure on S_n. We know from earlier that λ_n can be uniquely extended as an σ-additive measure to the minimal ring $R_n = R(S_n)$, which consists of a finite union of bounded boxes. Calling the extension λ_n, as well, we get that λ_n is a σ-finite measure on R_n. Hence, λ_n is a measure on the σ-algebra $\mathcal{M}_n$ that contains R_n. Now, we get the definition we've been working towards.

Definition 3.12. The measure λ_n on $\mathcal{M}_n$ is called the *Lebesgue Measure* (in $\mathbb{R}^n$). The measurable sets in $\mathbb{R}^n$ are also called *Lebesgue Measurable.*

4 Probability

First, let's define the notion of a probability space, and probability measures in the abstract and then we'll get into some of their properties.

Probability theory is a generalization of measure theory.

Definition 4.1. A *Probability Space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where the following properties hold:

1. Ω is a non-empty set. which is called the *sample space.*
2. $\mathcal{F}$ is a σ-algebra of subsets of Ω, whose elements are called *events.*
3. $\mathbb{P}$ is a *probability measure* on $\mathcal{F}$, that is, $\mathbb{P}$ is a measure on $\mathcal{F}$ and $\mathbb{P}(\Omega) = 1$ (in particular, $\mathbb{P}$ is a finite measure). For any even $A \in \mathcal{F}, \mathbb{P}(A)$ is called the probability of event A.

When we look at probability spaces, here are few more aspects of a measure that we need to recall.

1. $\mu(\emptyset) = 0$
2. σ-additivity: $\mu\Bigg(\bigcup_{i=1}^{\infty} A_i\Bigg) = \sum_{i=1}^{\infty} \mu(A_i)$

4.0.1 Properties of Probability Measures

Let's now run through a few properties of probability measures. For this section, let $(\Omega, \mathcal{F}, \mathbb{P})$ be our probability space.

Theorem 4.2. *1. Suppose A is a subset of Ω such that $A \in \mathcal{F}$. Then, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*

2. Consider events A and B such that $A \subseteq B$ and $A, B \in \mathcal{F}$. Then, $\mathbb{P}(A) \leq \mathbb{P}(B)$

3. *If $A_1, A_2, \ldots, A_n$ are a finite number of disjoint events, then* $\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$
4. *For any* $A, B \in \mathcal{F}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *In general, for a family of events* $\{A_i\}_{i=1}^n \in \mathcal{F}$*, we get that* $\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} \mathbb{P}(\cap_{i=1}^n A_i)$
5. *If* $\{A_i, i \geq 1\}$ *are events, then* $\mathbb{P}(\cup_{i=1}^\infty A_i) = \lim_{m \to \infty} \mathbb{P}(\cup_{i=1}^m A_i)$ *(this is called the continuity of probability measures)*
6. *If* $\{A_i, i \geq 1\}$ *is a sequence of decreasing nested events* $A_{i+1} \subseteq A_i \forall i \geq 1$*, then* $\mathbb{P}(\cap_{i=1}^\infty A_i) = \lim_{m \to \infty} \mathbb{P}(A_m)$
7. *If* $\{A_i, i \geq 1\}$ *is a sequence of decreasing nested events i.e.* $A_{i+1} \subseteq A_i \forall i \geq 1$*, then* $\mathbb{P}(\cap_{i=1}^\infty A_i) = \lim_{m \to \infty} \{(A_m)$
8. *Suppose* $\{A_i, i \geq 1\}$ *are events, then* $\mathbb{P}(\cup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mathbb{P}(A_i)$

Proof.

1. Given some $A \in \mathcal{P}(\Omega)$, A and A^c partition the sample space. Thus, $A^c \cup A = \Omega$ and $A^c \cap A = \emptyset$. Because we have countable additivity we get that

$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$$

This implies that

$$\mathbb{P}(A) + \mathbb{P}(A^c) = 1\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

2. We know that $B = A \cup (A^c \cap B)$ and thus

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B) = \mathbb{P}(B)$$

Which shows that $\mathbb{P}(A) \leq \mathbb{P}(B)$ as $\mathbb{P}(A^c \cap B) \geq 0$.

3. This follows more or less directly from σ-additivity (verify this for yourself).
4. We get that

$$A \cup B = A \cup (A^c \cap B)$$

Because the events are disjoint,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$$

We can now partition B to be

$$B = (A \cap B) \cup (A^c \cap B)$$

Thus,

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$

Then we get that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

The general result follows by induction.

5. For this proof, we'll use two claims to prove it. Let

$$B_n = A_n \setminus \cup_{i=1}^n A_i$$

(a) $B_i \cap B_j = \emptyset, \forall i \neq j$

(b) $\cup_{i=1}^\infty A_i = \cup_{i=1}^\infty B_i$

Since the B_i's are a disjoint sequence of events, we can use the previous lemmas to get that

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i)$$

Thus

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} \mathbb{P}(B_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \end{aligned}$$

It is easy to deduce the result from here.

6. This is a corollary of the last property.

7. This is also a corollary of the last proven property.

8. Using the same two claims that we established in property 5, we will define the B_i's used in this proof in a similar manner. Using the previous claims we can also get that

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i)$$

Since

$$B_i \subseteq A_i \forall i \geq 1, \mathbb{P}(B_i) \leq \mathbb{P}(A_i) \forall i \geq 1$$

Hence,

$$\sum_{i=1}^{n} \mathbb{P}(B_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

As we go to infinity, we get that

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Eventually, we get

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

□

5 Returning to the Dart

With our newfound understanding of probability in terms of the Lebesgue measure, we can finally return to our central question: What is the chance a dart thrown randomly in $[0, 1]$ lands on a rational number? We now know another way to formulate this question.

Question 5.1. What is the Lebesgue measure of the subset $\mathbb{Q} \cap [0, 1] \in \mathbb{R}$?

We have the answer. Because $\mathbb{Q}$ is a countable set, its Lebesgue measure is zero. We say the dart will *almost surely* not hit a rational number, because it could conceivably occur, but only with probability zero. What if we are more lenient? What if we allow the dart to land close to a rational number?

Here's a surprise! It turns out we have control over that chance, depending on how we define the word "close." For example, enumerate the rationals in some way to get a sequence $\{q_i\}_{k=1}^{\infty}$. Assign to q_k an open interval A_k of length $\epsilon/2^k$, where $0 < \epsilon < 1$. Thus, each rational is covered. What is the Lebesgue measure of the union of the intervals $A = \bigcup_{k=1}^{\infty} A_k$? Because nearby fractions may overlap (take 1/2 and 11/23, for instance), calculating an exact answer for this setup is challenging, but it is likely that $\mu(A)$ will be close to ϵ depending on how we enumerate $\mathbb{Q}$. We can bound the measure:

$$0 < \mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k) = \epsilon \sum_{k=1}^{\infty} \frac{1}{2^k} = \epsilon.$$

If $\epsilon = 0.5$, then these intervals do not contain every real number despite containing every rational number in some interval. How is this possible? Aren't the rational numbers dense in the interval? It turns out that, although rational numbers converge to every real number, they do not converge fast enough. A real number that is difficult to cleanly approximate with rational numbers will not be included in the measure for low enough ϵ. Thus, we can fiddle with the chances that the randomly thrown dart lands near a rational.

Let's take a moment to appreciate how far we've come. Previously, our notion of size was constricted to a small subset of all cases. Now, with the help of the Lebesgue measure and σ-algebras of measurable sets, we can formally speak of many otherwise bizarre lengths, areas, and volumes. Countable sets pale in size to the real continuum, as shown by them having measure zero. But we can also finesse the covering to yield an arbitrary measure. If we owned a casino, we would have the power to rig the game to offer a drunken dart thrower an arbitrary maximum chance of hitting near to a rational number. With measure theory, we have a precise grasp of the notions of size and density. These concepts form the foundation of many theories that form the foundation of our world.

References

[1] Alexander Grigoryan. *Measure Theory and Probability.* CreateSpace Independent Publishing Platform, 24 October 2014.

[2] D.H. Fremlin. *Measure Theory* Vol. 1. Torres Fremlin. 2000.
`wiki.math.ntnu.no/_media/tma4225/2011/fremlin-mt1.pdf`

[3] Jeffrey Miller. "(PP 1.7) Measure theory: More Properties of Probability Measures." *YouTube* channel mathematicalmonk. 20 Apr. 2011.
`www.youtube.com/watch?v=t4cwYCVyQLM&list=PL17567A1A3F5DB5E4`

[4] Krishna Jagannathan. "Lecture 5: Properties of Probability Measures." July-November 2015.
`www.ee.iitm.ac.in/~krishnaj/EE5110_files/notes/lecture5_properties%20of%20prob%20measures.pdf`