# Measure and Measurability

Daniel K. Rui - June 3, 2020

**Abstract**

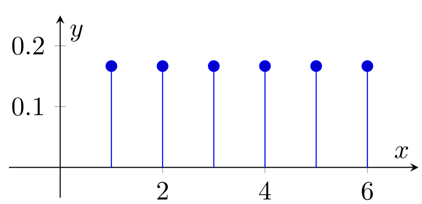
This paper seeks to introduce the concepts of measures and measurability, starting from the realm of probability theory. The goal is to motivate all the concepts, such as compactness, measurable sets, and $\sigma$-algebras, instead of just giving a definition and moving on as most textbooks do. Learning mathematics to me is a process of discovery, not blind acceptance of definitions.

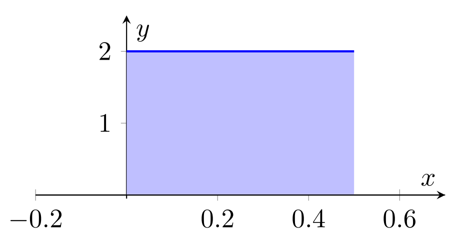
## 1 A Crisis in Probability

Suppose we are given a discrete random variable (henceforth r.v.), $X$, say for example the r.v. representing the outcome of a dice roll: $P(X = x) = \frac{1}{6}$ for $x \in \{1, \ldots, 6\}$. Then, the expected value of $X$ would be $\mathbb{E}[X] = \sum_x xP(X = x)$, which in our example would be $\frac{1}{6}(1+2+\ldots+6) = 3.5$. However, if we were to instead consider a continuous r.v., say for example the r.v. $Y$ that takes values uniformly in $[0, \frac{1}{2}]$. Intuitively, we would say that the expected value for this r.v. would be $\frac{1}{4}$. But our formula for expected value no longer works in the continuous case; instead we must use the analogous formula $\mathbb{E}[Y] = \int_{-\infty}^{\infty} xf_Y(x)\,dx$ where $f_Y(x)$ is called the "probability density function" of $Y$, which in our example is $f_Y(x) = 2 \cdot 1_{[0,\frac{1}{2}]}(x)$. Additionally, the probability that $X$ and $Y$ take values in some set $S$ is respectively $\sum_{x \in S} P(X = x)$ and $\int_S f(x)\,dx$.

Because of this disconnect between the discrete and continuous cases, the elementary probability student is burdened with learning two different "styles" of probability, in which the continuous case is treated as a more "abstract analogy" to the very intuitive discrete case, with no concrete meaning of its own: case in point, the function $f_Y(x)$ does NOT represent probability (as is obvious from the fact that it takes on the value 2), but rather a "likelihood". As such, it can seem at times that this framework is not elegant or intuitive, or not capturing the "true spirit" of probability theory.

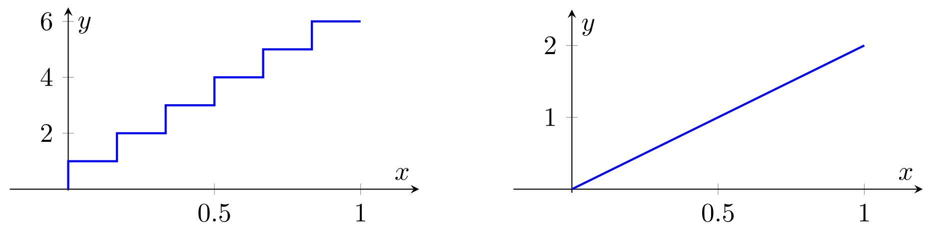
In order to develop a more elegant and intuitive description, let us first look at what functions we are, in a sense, integrating in the above examples:



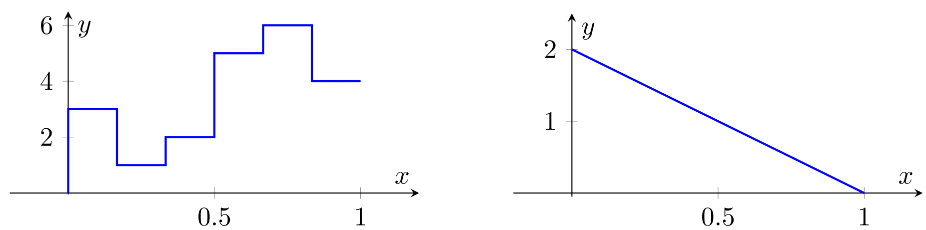




As you can see, the values that the r.v.'s take are on the $x$-axis, and the corresponding probability/likelihood is on the $y$-axis. But why would we put the values (i.e. the outcomes) of the r.v.'s on the $x$-axis?

Wouldn't it be more intuitive to put them on the $y$-axis, and instead use the $x$-axis as some sort of way to tell how much or how often the r.v. is taking on that value? For example, if we were to use this framework on our two examples, it would look like this:




In this framework, the probability that $X$ and $Y$ take values in some set $S$ is the "size" of the inverse images $X^{-1}(S)$ and $Y^{-1}(S)$. Notice that both functions are zero outside of $[0, 1]$. This is intentional, as we want the underlying space, $[0, 1]$ in this example, to have a "size" of 1, so that the "sizes" of inverse images can represent the notion of probability well. We usually denote these underlying spaces $\Omega$; think if it as the setting in which the action takes place. The expected value would be the integrals $\int_\Omega X\,dx$ and $\int_\Omega Y\,dx$. This framework also gives a lot of flexibility in what $X$ and $Y$ look like exactly, even with the restriction that they be representing a dice roll or a uniformly random number in $[0, 2]$; for example, we could also represent $X$ and $Y$ as




You can check for yourself that the probabilities and expected values are all the same. However, the astute reader may have noticed that I used the word "size" earlier when I talked about the probabilities being the "size" of the inverse image. So what exactly is the "size" of a set? This question is a particularly interesting one — it is the foundational question behind the field of measure theory. We will begin answering this question on the underlying space $\Omega = \mathbb{R}$.

# 2 A Crisis in Measures

As stated previously, we would like to have some function $\mu$ (called a "measure") to assign a number form $[0, \infty]$ to subsets of $\mathbb{R}$. What properties would we want such a function to have? Well,

- We would like it to agree with our intuition on basic sets that we already "know" the size of, such as intervals of the form $(a, b)$. Define a function $\ell$ ($\ell$ for "length") such that $\ell((a, b)) = b - a$ (where $a$ and $b$ can take on $-\infty$ and $\infty$ resp., and where $\varnothing = (a, a)$ has length 0). We want $\mu(I) = \ell(I)$ for all $I = (a, b)$.
- We would like it to be translation invariant: $\mu(A) = \mu(A + t)$. It would be contrary to our intuition if we could shift an interval left or right and have it change "size".

- For any sequence of sets $A_1, A_2, \ldots$ that are all disjoint (i.e. $A_i \cap A_j \neq \varnothing \implies i = j$), we would like $\mu$ to satisfy
$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$
Intuitively, if we have disjoint sets, the "size" of all of them should just be the sum of all of their "sizes" individually. This property is called "countable additivity".

- And we would like $\mu$ to assign a value to every subset of $\mathbb{R}$. (There is not even a function that satisfies finite additivity along with the other three properties)

First a couple lemmas about what a function $\mu$ satisfying these properties would be like:

*Lemma 2.1: monotonicity*

If $A \subseteq B$, $\mu(A) \leq \mu(B)$

*Proof:* from countable additivity on $\{A, B \backslash A, \varnothing, \ldots\}$, we have $\mu(B) = \mu(A \cup B \backslash A) = \mu(A) + \mu(B \backslash A) \geq \mu(A)$ (because $\mu \geq 0$).■

*Lemma 2.2: measure of closed intervals*

$\mu([a, b]) = b - a$

*Proof:* (consider the case first where there are no infinities and $a < b$). By the first property, $\mu((a - \epsilon, b + \epsilon)) = b - a + 2\epsilon$ and $\mu((a + \epsilon, b - \epsilon)) = b - a - 2\epsilon$. From the previous lemma, and the fact that $\mu$ must assign a number to every set in $2^{\mathbb{R}}$, it must be that $\mu([a, b]) = b - a$. For the null and infinite cases, similarly straightforward arguments show the claim. ■

*Lemma 2.3: countable subadditivity*

For any $A_1, A_2, \ldots \in 2^{\mathbb{R}}$, $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mu(A_k)$

*Proof:* consider $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus (A_1 \cup A_2)$, and so on (by the first lemma, $\mu(B_i) \leq \mu(A_i)$). These $B_i$ are clearly disjoint, and have union equal to the union of the $A_i$, and so the claim follows:
$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mu(B_k) \leq \sum_{k=1}^{\infty} \mu(A_k) \quad ■$$

Now comes the bad news:

*Theorem 2.1: no perfect measures*

There does not exist any function $\mu$ with the four properties listed above.

*Proof:* [1] (by Vitali set). For some $a \in [-1, 1]$, define the corresponding $\tilde{a}$ to be the set $\tilde{a} = \{a' \in [-1, 1] : a - a' \in \mathbb{Q}\}$; that is, $\tilde{a}$ is the equivalence class of $a$ (intersected with $[-1, 1]$) under the equivalence relation $x \sim y \iff x - y \in \mathbb{Q}$. Note that for any $a' \in \tilde{a}$. $|a - a'| \leq 2$ and so $a - a' \in \mathbb{Q} \cap [-2, 2]$.

It is clear that different equivalence classes are disjoint, i.e. $\tilde{a}_1 \cap \tilde{a}_2 \neq \varnothing \implies \tilde{a}_1 = \tilde{a}_2$. Define $S := \{\tilde{a} : a \in [-1,1]\}$. Now (using the Axiom of Choice) define $V$ to be a set that contains exactly one element from every equivalence class in $S$ (and so for every $s \in S$, $V \cap s$ has one element).

From the countability of the rationals, denote $\{r_1, r_2, \ldots\}$ to be an enumeration of all the rationals in $[-2,2]$. For every $a \in [-1,1]$, there is one $v \in V \cap \tilde{a}$, which satisfies $a - v \in \mathbb{Q} \cap [-2,2]$, there is some $k$ such that $a - v = r_k \iff a = v + r_k \iff a \in (r_k + V)$. Thus,

$$[-1,1] \subseteq \bigcup_{k=1}^{\infty} (r_k + V) \implies 2 = \mu([-1,1]) \leq \sum_{k=1}^{\infty} \mu(r_k + V) = \sum_{k=1}^{\infty} \mu(V) \implies \mu(V) > 0$$

But notice also that every $(r_k + V) \subseteq [-3,3]$ because $r_k \in [-2,2]$ and $V \subseteq [-1,1]$, so for any $n \in \mathbb{N}$,

$$\bigcup_{k=1}^{n} (r_k + V) \subseteq [-3,3] \implies \mu\left(\bigcup_{k=1}^{n} (r_k + V)\right) \leq \mu([-3,3]) = 6$$

But because $\mu(V) > 0$, there is $N$ large enough that $N\mu(V) > 6$, and so $\sum_{k=1}^{N} \mu(r_k + V) = N\mu(V) > 6$, while the measure of the union is $\leq 6$. The $(r_k + V)$ are all disjoint (because $r_i + v_1 = r_j + v_2 \implies v_1 - v_2 = r_j - r_i \in \mathbb{Q} \implies v_1 = v_2 \implies r_i = r_j \implies i = j$), and so this contradicts countable (and finite) additivity. ■

Well, what to do? The first three properties are non-negotiable, and it may seem like the fourth is as well. But, do we really need to have a size for *every* subset of $\mathbb{R}$? Surely there must be sets that are so pathological and rarely encountered that they won't be missed if we set them aside. So if we do set aside some sets, which sets should we consider? To answer this question, let's take a step back and try *constructing* a measure by hand, and see where problems arise.

# 3 Outer Measure

Let us take inspiration from Riemann's construction of integration via upper and lower sums, and first create some notion of "outer measure" and "inner measure" (applicable to all subsets of $\mathbb{R}$), and calling the sets on which they agree "measurable".

*Definition 3.1: Lebesgue outer measure*

Define $\mu^* : 2^{\mathbb{R}} \to [0,\infty]$ to be $\mu^*(A) = \inf\left\{\sum_{k=1}^{\infty} \ell(I_k) : A \subseteq \bigcup_{k=1}^{\infty} I_k, I_k \text{ an open interval}\right\}$

To clear up some notation, any sequence of open intervals $\{I_k\}$ where $A \subseteq \bigcup_{k=1}^{\infty} I_k$ is called a (open) covering of $A$ and $\sum_{k=1}^{\infty} \ell(I_k)$ is called the total length of the cover. (This definition should feel very intuitive; we know how to assign "length" to intervals, so the two most natural generalizations would be smallest container or largest thing that will fit inside – the first is of course the outer measure we're talking about now, and the second is called the inner measure which we will talk about later). Now, this function $\mu^*$ obeys several properties:

*Lemma 3.1: monotonicity*

$A \subseteq B \implies \mu^*(A) \leq \mu^*(B)$

*Proof:* any covering $\{I_k\}$ of $B$ also covers $A$. ■

*Lemma 3.2: countable sets have zero measure*

$\mu^*(\{x_1, x_2, \ldots\}) = 0$

*Proof:* $\{x_1, x_2, \ldots\} \subseteq \left(x_1 - \frac{\epsilon}{2^2}, x_1 + \frac{\epsilon}{2^2}\right) \cup \left(x_2 - \frac{\epsilon}{2^3}, x_2 + \frac{\epsilon}{2^3}\right) \cup \ldots$, which has a total length of $\epsilon$. Thus, the infimum must be 0.■

*Lemma 3.3: countable subadditivity*

$\mu^*\left(\bigcup_{k=1}^{\infty} A_i\right) \leq \sum_{k=1}^{\infty} \mu^*(A_i)$

*Proof:* let $I_{k1}, I_{k2}, \ldots$ cover $A_k$ s.t.

$$\mu^*\left(\bigcup_{i=1}^{\infty} I_{ki}\right) < \mu^*(A_k) + \frac{\epsilon}{2^k}$$

Such a covering exists because $\mu^*(A_k)$ is the infimum. If we do this for all $A_k$, then we get

$$\mu^*\left(\bigcup_{k=1}^{\infty}\bigcup_{i=1}^{\infty} I_{ki}\right) \leq \sum_{k=1}^{\infty} \mu^*\left(\bigcup_{i=1}^{\infty} I_{ki}\right) < \sum_{k=1}^{\infty} \mu^*(A_i) + \epsilon$$

We know that $\bigcup_{k=1}^{\infty} A_i \subseteq \bigcup_{k=1}^{\infty}\bigcup_{i=1}^{\infty} I_{ki}$ and that $\{I_{ki}\}$ is countable (b/c $\mathbb{N} \times \mathbb{N}$ is countable), so

$$\mu^*\left(\bigcup_{k=1}^{\infty} A_i\right) \leq \mu^*\left(\bigcup_{k=1}^{\infty}\bigcup_{i=1}^{\infty} I_{ki}\right) < \sum_{k=1}^{\infty} \mu^*(A_i) + \epsilon$$

And because $\epsilon$ is arbitrary, we conclude the proposition.■

*Lemma 3.4: approximating from the outside by open sets*

$\inf\left\{\mu^*(U) : A \subseteq U, U \text{ open}\right\} = \mu^*(A)$

*Proof:* we know from Lemma 3.1 that for any $U \supseteq A$, $\mu^*(U) \geq \mu^*(A)$, and so the infimum must also be $\geq \mu^*(A)$. Furthermore, for any $\epsilon > 0$, we know that there is a cover $\{I_k\}$ of $A$ which has total length $L < \mu^*(A) + \epsilon$. $U := \bigcup_{k=1}^{\infty} I_k$ is obviously covered by $\{I_k\}$, and so $\mu^*(U)$ must be $\leq L < \mu^*(A) + \epsilon$. $U$ is also an open set that contains $A$, and thus, the infimum value over all open sets that contain $A$ must be also $< \mu^*(A) + \epsilon$. $\epsilon$ is arbitrary, so we see that the infimum value must be exactly $\mu^*(A)$. ■

Note that in our definition of Lebesgue outer measure, we can restrict all our intervals to have length less than any fixed $\alpha$; i.e. we can instead define

$$\mu^*(A) = \inf\left\{\sum_{k=1}^{\infty} \ell(I_k) : \ I_k \text{ are open intervals s.t. } \ell(I_k) \leq \alpha \text{ and } A \subseteq \bigcup_{k=1}^{\infty} I_k\right\}$$

This is because given any $I_k$ from the original covering, we can replace it with a finite or countable collections of intervals of length $\leq \alpha$ without (much) changing the total length of the covering: in the case that $I_k$ is not bounded, w.l.o.g. take $I_k = (a, \infty)$ (the cases $(-\infty, b)$ and $(-\infty, \infty)$ are analogous), then we can replace $I_k$ with

$$I_k \subseteq (a, a+\alpha) \cup (a+\alpha-\epsilon, a+2\alpha-\epsilon) \cup (a+2\alpha-2\epsilon, a+3\alpha-2\epsilon) \cup \dots$$

We see that total length of the covering is $\infty$ in both cases (i.e. with or without the $\ell \leq \alpha$ condition). If e.g. $I_k = (a, b)$ is bounded, then the following covering with small intervals suffices:

$$\begin{aligned} I_k = (a,b) \subseteq (a, a+\alpha) &\cup \left(a+\alpha - \frac{1}{2^2}\frac{\epsilon}{2^k}, a+\alpha+\frac{1}{2^2}\frac{\epsilon}{2^k}\right) \\ &\cup (a+\alpha, a+2\alpha) \cup \left(a+2\alpha - \frac{1}{2^3}\frac{\epsilon}{2^k}, a+2\alpha+\frac{1}{2^3}\frac{\epsilon}{2^k}\right) \\ &\cup (a+(n-1)\alpha, a+n\alpha) \cup \left(a+n\alpha - \frac{1}{2^{n+1}}\frac{\epsilon}{2^k}, a+n\alpha+\frac{1}{2^{n+1}}\frac{\epsilon}{2^k}\right) \cup \dots \\ &\cup (a+n\alpha, b) \end{aligned}$$

for some appropriate $n$. The sum of the lengths of the $n$ intervals above is

$$= (b-a) + \left(\frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^n}\right)\frac{\epsilon}{2^k} = \ell(I_k) + \left(\frac{2^n - 1}{2^n}\right)\frac{\epsilon}{2^k} < \ell(I_k) + \frac{\epsilon}{2^k}$$

and so by replacing each $I_k$ with at most countably many small intervals, we get a covering of $A$, by countably many (b/c $\mathbb{N} \times \mathbb{N}$ is countable) small intervals, with a total length of $< \sum_{k=1}^{\infty} \ell(I_k) + \epsilon$.

In English, this all tells us that we can go from an arbitrary covering to one with only small intervals, with the caveat of needing an arbitrarily small amount of additional length.

*More explicitly:* to prove that the infimum total length for coverings with only small intervals is exactly $\mu^*(A)$, we need to show that for any $\epsilon > 0$, we can find a cover $\{I_k\}$ using only small intervals s.t. the total length of $\{I_k\}$ is $< \mu^*(A) + \epsilon$. We do this by finding a cover $\{J_k\}$ (using arbitrary intervals) s.t. the total length of $\{J_k\}$ is $< \mu^*(A) + \frac{\epsilon}{2}$, and then cutting $\{J_k\}$ into a cover involving only small intervals, which we know we can do while still keeping the total length $< (\mu^*(A) + \frac{\epsilon}{2}) + \frac{\epsilon}{2}$. QED! Let us continue listing properties of Lebesgue outer measure:

*Lemma 3.5: (finite) additivity for distant sets*

$$d(A,B) := \inf\{|x-y| : x \in A, y \in B\} > 0 \implies \mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$$

*Proof:* let $d(A, B) = \alpha > 0$; then from above we can just force all our intervals to have length $\leq \frac{\alpha}{2}$. This would imply that any covering of $A \cup B$ could be split into two coverings of $A$ and $B$ respectively, e.g. by putting each $I_k$ into the covering of $A$ if $I_k \cap A \neq \varnothing$ and into the covering of $B$ otherwise. The reason why we have to first restrict ourselves to intervals of length $\leq \frac{\alpha}{2}$ is because this way we can guarantee that none of the $I_k$ have non-empty intersections with both $A$ and $B$; if they did then we would not be guaranteed that we could split the original cover cleanly into two subcovers of $A$ and $B$.

More clearly the argument is that for any $\epsilon > 0$, we can find a covering $\{I_k\}$ with only small intervals (covering $A \cup B$) s.t. the total length is $< \mu^*(A \cup B) + \epsilon$. We can then split $\{I_k\}$ into two covers of $A$ and $B$ respectively. The sum of the total lengths of these two subcovers is clearly the same as the total length of the original cover, which means that $\mu^*(A) + \mu^*(B) < \mu^*(A \cup B) + \epsilon$. Because $\epsilon$ is arbitrary, this means that $\mu^*(A) + \mu^*(B) \leq \mu^*(A \cup B)$. Use countable subadditivity (3.3) for $\geq$. ■

*Lemma 3.6: inequality regarding outer measure of bounded intervals*

For any interval $I$ with endpoints $a$ and $b$, $\mu^*(I) \leq b - a$

*Proof:* $I \subset (a - \frac{\epsilon}{2}, b + \frac{\epsilon}{2}) \implies \mu^*(I) \leq \ell((a - \frac{\epsilon}{2}, b + \frac{\epsilon}{2})) = b - a + \epsilon$. $\epsilon$ is arbitrary, so $\mu^*(I) \leq b - a$. ■

The other direction of Lemma 3.6 turns out to be much trickier to prove, because we need to prove that for *any* cover $\{I_k\}$ of $I$, $\sum_{k=1}^{\infty} \ell(I_k) \geq b - a$. The difficulty here lies in the infinite sum, so one way forward would be to somehow make it so that we only deal with finite sums (where we can use tools like induction). Well, wouldn't it be nice if we had a certain type of interval where EVERY open cover could be reduced to a finite open cover? Indeed it would:

*Lemma 3.7: outer measure of compact sets*

For any interval $I$ with endpoints $a$ and $b$ with the property that for any open cover $\{I_k\}$, there exists $N$ s.t. $I$ is covered completely by just $I_1, \ldots, I_N$, $\sum_{k=1}^{N} \ell(I_k) \geq b - a$

*Proof:* as I hinted before, we proceed by induction. Base case: if $N = 1$, then $I \subseteq I_1$ implies that $I_1 = (c, d)$ where $c \leq a$ and $d \geq b$ and so $\ell(I_1) \geq b - a$. ✓
Now assuming that the statement holds for all $N$ up to $N = n - 1$, then $I \subseteq I_1 \cup \ldots \cup I_{n-1}$. $b$ must be in one of these intervals (or it is the right endpoint of one of them), i.e. there must be an $I_k = (c, d)$ (via relabeling let it be $I_n$) from the open cover s.t. $c < b \leq d$. If $c \leq a$, $\ell(I_n)$ is obviously $\geq b - a$. If not, then we have that $a < c < b \leq d$. Now we know that the interval $I \cap [a, c]$ is covered by $I_1 \cup \ldots \cup I_{n-1}$. If furthermore this interval has the special finite open cover property, we can use the induction hypothesis to get that $\sum_{k=1}^{n-1} \ell(I_k)$ must be $\geq c - a$, which means that $\sum_{k=1}^{n} \ell(I_k) = \sum_{k=1}^{n-1} \ell(I_k) + \ell(I_n) \geq (c - a) + (d - c) = d - a \geq b - a$. ■

The proof is a bit overburdened with caveats because we have not characterized the intervals with the special finite open cover property (we don't even know if such intervals exist!), but at least it demonstrates that once we do characterize such intervals, we can use the core ideas from this proof to prove results about the outer measure of those intervals. Now for the characterization:

**Theorem 3.1: Heine-Borel theorem**

The sets in $\mathbb{R}$ for which every open cover can be reduced to a finite open cover are exactly the sets which are closed and bounded (in fact, the "finite open cover property" is so important that we'll give it a name: compactness. I think this is the best motivation for developing the concept).

*Proof:* ( $\Longrightarrow$ ) Let $S$ be compact. First, $S$ must be bounded: consider the open cover $\{(-k,k)\}_{k=1}^{\infty}$, which necessarily covers $S$ because $\bigcup_{k=1}^{\infty}(-k,k)=\mathbb{R}$. By compactness, $S$ is covered by $(-1,1)\cup\ldots\cup(-N,N)$, and so $S\subseteq(-N,N)$ and so it's bounded.
$S$ must also be closed: suppose not. Then there is $p\notin S$ that is on the boundary. Consider the open cover $\{\mathbb{R}\setminus\overline{B(p,1/k)}\}_{k=1}^{\infty}$ where $B(p,1/k)$ is the open ball centered at $p$ with radius $1/k$ (and the overline is "closure of"), which must necessarily cover $S$ because

$$\bigcup_{k=1}^{\infty} I_k=\bigcup_{k=1}^{\infty}\mathbb{R}\setminus\overline{B(p,1/k)}=\mathbb{R}\setminus\{p\}\supseteq S$$

By compactness, $S$ can be covered by $I_1\cup\ldots\cup I_M$, and we know $(\mathbb{R}\setminus\overline{B(p,1/M)})$ and $B(p,1/M)$ are disjoint, so we know $S$ and $B(p,1/M)$ must be disjoint. This means that $p$ is actually not on the boundary of $S$ (it is in the exterior); contradiction.

( $\Longleftarrow$ ) Suppose $F$ is a closed bounded set in $\mathbb{R}$ and $\mathscr{C}$ an open cover of $F$. First consider the case where $F=[a,b]$. Define $D$ to be $D=\{d\in[a,b]:[a,d]$ has a finite subcover from $\mathscr{C}\}$. $D$ is not empty because $D$ must contain at least $a$, because $a$ must be contained in at least one open set in $\mathscr{C}$.

We want to prove that $s=\sup D$ is $b$. Well, $[a,b]$ is closed so the supremum $s$ is also in $[a,b]$. Suppose that $s<b$. Then, there must be some open set $G$ from $\mathscr{C}$ that covers $(s-\delta,s+\delta)$ for some $\delta>0$. But we know by definition of $s$ that $[a,s-\frac{\delta}{2}]$ is covered by finitely many open sets from $\mathscr{C}$, say $G_1\cup\ldots\cup G_N$. But then $[a,s+\frac{\delta}{2}]$ is also covered by finitely many open sets from $\mathscr{C}$, i.e. $G_1\cup\ldots G_N\cup G$ and so $s$ is not the supremum; contradiction; thus $s\geq b$, so $[a,b]$ can be covered by finitely many sets from $\mathscr{C}$.

Now let $F$ be any arbitrary closed and bounded set in $\mathbb{R}$ with open cover $\mathscr{C}$. By boundedness, $F\subset[a,b]$ for some $a,b\in\mathbb{R}$. Then $\mathscr{C}\cup\{\mathbb{R}\setminus F\}$ is an open cover covering $\mathbb{R}$ and hence $[a,b]$. From the first part, we know that for some $M$, $F\subset[a,b]\subset G_1\cup\ldots\cup G_M\cup(\mathbb{R}\setminus F)$. $F$ obviously is not in $\mathbb{R}\setminus F$, so $F\subset G_1\cup\ldots\cup G_M$. ■

Read through the proof of property 3.7 again, replacing $I$ with $[a,b]$ — it should make more sense this time around. To summarize, 3.6 and 3.7 give us that $\mu^*([a,b])=b-a$ as one would expect. Finally, we finish off with two last properties regarding outer measure:

**Lemma 3.8: null sets do not impact measure**

If $N$ is a null set, i.e. $\mu^*(N)=0$, then $\mu^*(A\cup N)=\mu^*(A)$.

*Proof:* $\mu^*(A) \leq \mu^*(A \cup N) = \mu^*(A) + \mu^*(N) = \mu^*(A)$. ■

*Lemma 3.9: outer measure of intervals*

Any interval $I$ with endpoints $a$ and $b$ (e.g. $(a,b)$ or $(a,b]$, etc.) have outer measure $b-a$.

*Proof:* practically trivial by (8) and the fact that $\mu^*([a,b]) = b-a$. Finally, any intervals with $\pm\infty$ at their endpoints must have outer measure $\infty$ because of monotonicity (3.1) applied to larger and larger bounded intervals contained in those unbounded intervals. ■

# 4 Inner Measure and Bounded Measurability

As promised, we now define Lebesgue inner measure (inspired by Lemma 3.4):

*Definition 4.1: Lebesgue inner measure*

$\mu_* : 2^{\mathbb{R}} \to [0,\infty]$ is defined as $\mu_*(A) = \sup\{\mu^*(K) : K \subseteq A, K \text{ compact}\}$.

A set with finite outer measure is Lebesgue measurable (sometimes just "measurable") if its inner and outer measure agree (the infinite case will be dealt with in a bit, because $\infty$ in a sense is not precise; think hypothetically of a scenario where we have a set $A$ in (0,1) with $\mu^*(A) = \frac{2}{3}$ but $\mu_*(A) = \frac{1}{3}$, and we copy and paste the set to (1,2) and (2,3) and so on ad infinitum. Then the outer and inner measure are both $\infty$, but such a set should clearly not be measurable). However, we will start with a more restrictive set of sets; those that not only have finite outer measure, but are bounded:

*Definition 4.2: bounded Lebesgue measurable*

If $A$ is bounded, and if $\mu^*(A) = \mu_*(A)$, then $A$ is bounded Lebesgue measurable. Denote the Lebesgue measure of a Lebesgue measurable set to be $\mu(A) = \mu^*(A) = \mu_*(A)$.

The reason why we don't define inner measure in terms of the length function $\ell$ on compact sets (i.e. $\sup\{\sum_{k=1}^{\infty} \ell(I_k) : A \supseteq \bigcup_{k=1}^{\infty} I_k, I_k \text{ a compact interval}\}$) is because then we would assign inner measure 0 to the irrational numbers in $[0,1]$ (call it $\mathbb{I}_{[0,1]}$), when the outer measure of that same set is 1 (because by 3.8, $\mu^*(\mathbb{I}_{[0,1]}) = \mu^*(\mathbb{I}_{[0,1]} \cup \mathbb{Q}_{[0,1]}) = \mu^*([0,1]) = 1$). It would be a pathetic definition if even something as simple as the irrational numbers in $[0,1]$ were non-measurable! Let us now dig into the above definitions; the following three facts should be fairly immediate:

*Lemma 4.1: inner measure $\leq$ outer measure*

For any set $A$, $\mu_*(A) \leq \mu^*(A)$.

*Proof:* $K \subseteq A \implies \mu^*(K) \leq \mu^*(A)$ and so the sup over all compact $K \subseteq A$ is still $\leq \mu^*(A)$. ■

*Lemma 4.2: (bounded) intervals are Lebesgue measurable*

If $I$ is an interval with endpoints $a$ and $b$, then $\mu(I) = \mu^*(I) = \mu_*(I) = b-a$.

*Proof:* if $a = b$, then the inner measure is obviously 0. Otherwise if $a < b$, then for all $\epsilon > 0$ small enough, $a + \epsilon < b - \epsilon$, and so we $[a + \epsilon, b - \epsilon]$ is a compact set contained in $I$, which means that $\mu_*(I) \geq \mu^*([a + \epsilon, b - \epsilon]) = b - a + 2\epsilon$. $\epsilon$ was arbitrary, so $\mu_*(I) \geq b - a$. The other direction follows from the above lemma (4.1). ■

*Lemma 4.3: compact sets are Lebesgue measurable*

If $K$ is compact, then $\mu^*(K) = \mu_*(K)$.

*Proof:* well, $K \subseteq K$, so $\mu^*(K) \leq \mu_*(K)$ (because $\mu_*(K)$ takes the sup over the outer measure of all subsets of $K$), but the lemma above tells us $\geq$, and so the quantities must be exactly equal. ■

This next theorem is a huge affirmation in our progress — we originally wanted a measure to be countably additive for all subsets of $\mathbb{R}$ (a wish that turned out sadly to be unfullilable), but now we'll see that if we stick to Lebesgue measurable sets, everything will work out the way we want it to:

*Theorem 4.1: measurable sets are countably additive*

For disjoint bounded measurable $A_1, A_2, \ldots$, and $A = \bigcup\limits_{n=1}^{\infty} A_n$, $\mu^*(A) = \mu_*(A) = \sum\limits_{n=1}^{\infty} \mu(A_n)$.

*Proof:* [4](note that the above "bounded" applies to $A$ as well). From countable additivity, we have that

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

Now fix $\epsilon < 0$, and choose a sequence of $K_n \subseteq A_n$ such that for each $n \in \mathbb{N}$, $\mu^*(K_n) \geq \mu_*(A_n) - \frac{\epsilon}{2^n}$ (there must be such $K_n$ by the definition of inner measure), where the RHS is equal to $\mu(A_n) - \frac{\epsilon}{2^n}$ by measurability. The $K_n$ are disjoint (being subsets of disjoint $A_n$) and compact, and so the distances between them are positive (see below lemma, 4.4), which means that we can use Lemma 3.5 to get that for every $N \in \mathbb{N}$,

$$\mu^*(K_1 \cup \ldots \cup K_N) = \mu^*(K_1) + \ldots + \mu^*(K_N) \geq \sum_{n=1}^{N} \left( \mu(A_n) - \frac{\epsilon}{2^n} \right) \geq \left( \sum_{n=1}^{N} \mu(A_n) \right) - \epsilon$$

But for all $N \in \mathbb{N}$, $A \supseteq (K_1 \cup \ldots \cup K_N)$ (where $(K_1 \cup \ldots \cup K_N)$ is compact), and so it must be that for all $N \in \mathbb{N}$,

$$\mu_*(A) \geq \mu^*(K_1 \cup \ldots \cup K_N) \geq \left( \sum_{n=1}^{N} \mu(A_n) \right) - \epsilon$$

and so the inequality holds as we take the limit $N \to \infty$. Lastly, $\epsilon$ was arbitrary, and so

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

which means that $\mu_*(A) = \mu^*(A)$, (where the other direction is from 4.1), and $\mu(A) = \sum\limits_{n=1}^{\infty} \mu(A_n)$. ■

In the above proof, we used that if $K_1$ and $K_2$ are compact (actually only one needs to be compact; the other can just be closed), then $d(K_1, K_2) > 0$. We prove this for the sake of being self-contained:

*Lemma 4.4: disjoint compact sets are distant*

If $K_1$ is compact and $K_2$ is closed, then $d(K_1, K_2) > 0$.

*Proof:* denote $S_n := \{x \in \mathbb{R} : d(x, K_2) > \frac{1}{n}\}$. Now for all $x \in S_n$, $d(x, K_2) = \alpha > \frac{1}{n}$, which means that for all $k_2 \in K_2, |x - k_2| \geq \alpha$, and so for all $y \in B(x, \frac{1}{2}(\alpha - \frac{1}{n}))$, we have that

$$\alpha \leq |x - k_2| = |x - y + y - k_2| \leq |x - y| + |y - k_2| < \frac{1}{2}(\alpha - \frac{1}{n}) + |y - k_2|$$

and so $\frac{1}{2}(\alpha + \frac{1}{n}) < |y - k_2|$. Because $k_2$ was arbitrarily chosen from $K_2$, the infimum of all such distances must still be $\geq \frac{1}{2}(\alpha + \frac{1}{n})$, i.e. $d(y, K_2) \geq \frac{1}{2}(\alpha + \frac{1}{n}) > \frac{1}{n}$. Thus we've found a neighborhood of $x$ that is also inside $S_n$, and so all the $S_n$ are open. Furthermore, $\bigcup_{n=1}^{\infty} S_n = (\mathbb{R} \setminus K_2)$, because $(\mathbb{R} \setminus K_2)$ is open (the complement of a closed set), and so for every $x \in (\mathbb{R} \setminus K_2)$, there is some $r > 0$ such that $B(x, r) \subseteq (\mathbb{R} \setminus K_2)$, which means that $|x - k_2| > \frac{r}{2}$ for all $k_2 \in K_2$, implying of course that $d(x, K_2) \geq \frac{r}{2} > 0$ (and so $x$ is all the $S_n$ for $n > \frac{2}{r}$). $\{S_n\}$ is hence an open cover of $K_1$, and so by compactness, $K_1$ is covered by finitely many $S_n$, say $S_1, \ldots, S_N$, and so $d(K_1, K_2) \geq \frac{1}{N} > 0$. ■

With Theorem 4.1, we can now prove that open sets are measurable (with the help of a lemma):

*Lemma 4.5: open sets are disjoint countable unions of open intervals*

If $U$ is an open set, then there are disjoint open intervals $I_1, I_2, \ldots$ such that $U = \bigcup_{n=1}^{\infty} I_n$.

*Proof:* (recall that an interval $(a, b) := \{x \in \mathbb{R} : a < x < b\}$, and that $(a, b) = \bigcup_{a<\alpha<\beta<b}(\alpha, \beta)$). For all rational $x \in U$, define

$$I_x = \bigcup_{x \in (a,b) \subseteq U} (a, b)$$

which is itself an open interval (because all the $(a, b)$ are open and share the point $x$ — more explicitly, for all $\alpha, \beta \in I_x$, $\alpha < \beta$, there must be intervals such that $\alpha \in (a, b)$ and $\beta \in (c, d)$; but both intervals contain $x$, and so $a < x < b$ and $c < x < d$, implying that $c < d$ and so $(\alpha, \beta) \subseteq (a, d) = (a, b) \cup (c, d) \subseteq I_x$, where again $\alpha$ and $\beta$ were arbitrary in $I_x$), and a subset of $U$.
For all irrational $x \in U$, by openness, there is some $r_x$ such that $B(x, r_x) \subseteq U$, and so picking any rational $y$ in $B(x, r_x)$, we see that $B(x, r_x) \subseteq I_y$ (by definition of $I_y$). In other words, for all $x \in U$, there is $q \in U \cap \mathbb{Q}$ such that $x \in I_q$, so

$$U \subseteq \bigcup_{q \in U \cap \mathbb{Q}} I_q.$$

For the other direction, note that $I_q \subseteq U$ for all $q \in U \cap \mathbb{Q}$. Lastly, for all $p, q \in U \cap \mathbb{Q}$, either $I_p = I_q$ or $I_p \cap I_q = \varnothing$ because case one, $x \in I_p \cap I_q \implies$ that $I_p \cup I_q$ is an open interval (because $I_p$ and $I_q$ are both open intervals sharing a point $x$), and so by the definition of $I_p$ and $I_q$, $I_p \cup I_q \subseteq I_p, I_q \implies I_p = I_q$. Case two is $\nexists x \in I_p \cap I_q$, i.e. $I_p \cap I_q = \varnothing$. If we want to be more rigorous, we can replace $\{I_k\}$ with $\{J_k\}$, where we discard all duplicate $I_k$'s. ■ (See more proofs here)

*Lemma 4.6: open sets are Lebesgue measurable*

If $U$ is any open set, then $\mu(U) = \mu^*(U) = \mu_*(U)$.

*Proof:* from the lemma above, $U$ can be written as $\bigcup_{n=1}^{\infty} I_n$ for disjoint $I_n$, and so by Lemma 4.2 and Theorem 4.1, $U$ is Lebesgue measurable. ■

We are now almost at the point where we can introduce our most general and useful definition of measurability; we first have to establish a connection between the inner measure of a set and the outer measure of its "complement":

*Theorem 4.2: inner measure vs. outer measure of complement*

For a bounded set $A$ contained in a bounded open (hence measurable) set $X$, we have that $\mu_*(A) = \mu(X) - \mu^*(X \setminus A)$.

*Proof:* fix any $\epsilon > 0$. Then, there is some compact $K$ such that $\mu(K) = \mu^*(K) \geq \mu_*(A) - \epsilon$. $X \setminus K$ is an open set (hence measurable), and so by Theorem 4.1, $\mu(K) + \mu(X \setminus K) = \mu(X)$. But $(X \setminus K) \supseteq (X \setminus A) \implies \mu(X \setminus K) = \mu^*(X \setminus K) \geq \mu^*(X \setminus A)$, and so putting everything together we have

$$\mu_*(A) - \epsilon \leq \mu(K) = \mu(X) - \mu(X \setminus K) \leq \mu(X) - \mu^*(X \setminus A)$$

On the other hand, countable additivity gives that $\mu(X) = \mu^*(X) \leq \mu^*(X \setminus A) + \mu^*(A) \implies \mu(X) - \mu^*(X \setminus A) \leq \mu^*(A)$, and so

$$\mu_*(A) - \epsilon \leq \mu(X) - \mu^*(X \setminus A) \leq \mu^*(A)$$

Since $\epsilon > 0$ was arbitrary, we have the desired equality. ■

We now establish a couple equivalent conditions for bounded Lebesgue measurability, the last of which we will then take as our new definition of measurable:

*Lemma 4.7: alternative definition of bounded Lebesgue measurable*

A bounded set $A$ (contained in some bounded open set $X$) is Lebesgue measurable if and only if $\mu^*(A) = \mu(X) - \mu^*(X \setminus A) \iff \mu(A) + \mu^*(X \setminus A) = \mu(X)$.

*Proof:* using Theorem 4.2, $\mu_*(A) = \mu^*(A) \iff \mu(X) - \mu^*(X \setminus A) = \mu^*(A)$. ■

*Lemma 4.8: complements (taken in an open set) are measurable*

For a bounded Lebesgue measurable set $A$ contained in a bounded open (hence measurable) set $X$, $(X \setminus A)$ is also Lebesgue measurable.

*Proof:* $(X \setminus A)$ is a bounded set in $X$, and furthermore $\mu^*(X \setminus A) + \mu^*(X \setminus (X \setminus A)) = \mu^*(X \setminus A) + \mu^*(A) = \mu(X)$ and so the above lemma tells us that $(X \setminus A)$ is in fact Lebesgue measurable. ■

We would like to now prove that the intersection of two Lebesgue measurable sets $A$ and $B$, is still Lebesgue measurable. Unfortunately, the alternate definition doesn't give us much help, so let's see what we can do just from the original definition. Well, from the measurability of $A$ and $B$, we can find outer open sets $U_A$ and $U_B$ and inner compact sets $K_A$ and $K_B$ with outer measures arbitrarily close to $\mu(A)$ and $\mu(B)$ respectively. Now is it the case that $U_A \cap U_B$ and $K_A \cap K_B$ are outer open and inner compact approximations to $A \cap B$? From a cursory glance, we have that $\mu(U_A \cap U_B) - \mu(K_A \cap K_B) = \mu((U_A \cap U_B) \setminus (K_A \cap K_B)) \leq \mu((U_A \setminus K_A) \cup (U_B \setminus K_B)) \leq \mu(U_A \setminus K_A) + \mu(U_B \setminus K_B)$ is arbitrarily small, and so it looks like the intersection also is Lebesgue measurable. We formalize things below:

*Lemma 4.9: another alternative definition of bounded Lebesgue measurable*

A bounded set $A$ is Lebesgue measurable if and only if for every $\epsilon > 0$, there is an open set $U$ and compact $K$ such that $K \subseteq A \subseteq U$ and $\mu^*(U \setminus K) = \mu(U \setminus K) < \epsilon$ (where we removed the asterisk by 4.8). $A$ satisfying this condition is said to be "bounded open-closed measurable".

*Proof:* ( $\implies$ ) fix any $\epsilon > 0$. From Lemma 3.4 and the definition of inner measurability, we can find open $U$ and compact $K$ such that $K \subseteq A \subseteq U$ and $\mu(U) < \mu^*(A) + \frac{\epsilon}{2}$ and $\mu(K) > \mu^*(A) - \frac{\epsilon}{2}$. But $K$ is a bounded Lebesgue measurable set in an open set $U$, and so by Lemma 4.7, we have that $\mu(K) + \mu(U \setminus K) = \mu(U) \implies \mu(U \setminus K) = \mu(U) - \mu(K) < (\mu^*(A) + \frac{\epsilon}{2}) - (\mu^*(A) - \frac{\epsilon}{2}) = \epsilon$.

( $\impliedby$ ) we have (from 4.7 and the assumption) that $\mu(U) = \mu(U \setminus K) + \mu(K) < \epsilon + \mu(K)$, and so $\mu^*(A) \leq \mu(U) < \epsilon + \mu(K) \leq \epsilon + \mu_*(A)$. $\epsilon > 0$ is arbitrary, and the other direction follows from 4.1. ■

*Lemma 4.10: intersections are measurable*

For bounded Lebesgue measurable sets $A, B$, $A \cap B$ is also Lebesgue measurable.

*Proof:* fix any $\epsilon > 0$. From the above lemma (4.9), we can find $U_A, K_A, U_B$, and $K_B$ such that $K_A \subseteq A \subseteq U_A$, and $\mu(U_A \setminus K_A) < \frac{\epsilon}{2}$ (likewise for the $B$ sets). Then, because

$$(U_A \cap U_B) \setminus (K_A \cap K_B) = \Big((U_A \cap U_B) \setminus K_A\Big) \cup \Big((U_A \cap U_B) \setminus K_A\Big) \subseteq (U_A \setminus K_A) \cup (U_B \setminus K_B),$$

we get that $\mu((U_A \cap U_B) \setminus (K_A \cap K_B)) \leq \mu((U_A \setminus K_A) \cup (U_B \setminus K_B)) \leq \mu(U_A \setminus K_A) + \mu(U_B \setminus K_B) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Obviously, $(K_A \cap K_B) \subseteq A \cap B \subseteq (U_A \cap U_B)$, so by 4.9, $A \cap B$ must be Lebesgue measurable. ■

*Lemma 4.11: countable intersections are measurable*

For bounded Lebesgue measurable sets $A_1, A_2, \ldots$, $A = \bigcap_{n=1}^{\infty} A_n$ is also Lebesgue measurable.

*Proof:* fix any $\epsilon > 0$. From 4.9, we have that for all $n \in \mathbb{N}$, $K_n \subseteq A_n \subseteq U_n$ and $\mu(U_n \setminus K_n) < \frac{\epsilon}{2^n}$. Then, because

$$\left(\bigcap_{n=1}^{\infty} U_n\right) \setminus \left(\bigcap_{n=1}^{\infty} K_n\right) = \bigcup_{n=1}^{\infty} \left(\left(\bigcap_{n=1}^{\infty} U_n\right) \setminus K_n\right) \subseteq \bigcup_{n=1}^{\infty} \Big(U_n \setminus K_n\Big),$$

we get that

$$\mu\left(\left(\bigcap_{n=1}^{\infty} U_n\right) \setminus \left(\bigcap_{n=1}^{\infty} K_n\right)\right) \leq \mu\left(\bigcup_{n=1}^{\infty} \left(U_n \setminus K_n\right)\right) \leq \sum_{n=1}^{\infty} \mu(U_n \setminus K_n) < \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon.$$

Obviously, $\bigcap_{n=1}^{\infty} K_n \subseteq \bigcap_{n=1}^{\infty} A_n \subseteq \bigcap_{n=1}^{\infty} U_n$, so by 4.9, $\bigcap_{n=1}^{\infty} A_n$ must be Lebesgue measurable. ■

*Theorem 4.3: strengthening of alternative definition*

A bounded set $A$ is Lebesgue measurable if and only if for ALL subsets $X$ of $\mathbb{R}$, $\mu^*(A \cap X) + \mu^*(X \setminus A) = \mu^*(A \cap X) + \mu^*(X \cap A^{\complement}) = \mu^*(X)$. A bounded set $A$ satisfying this condition is said to be "bounded Carathéodory measurable".

*Proof:* ( $\Longleftarrow$ ) simply choose open $X \supseteq A$ and use Lemma 4.7.
( $\Longrightarrow$ ) (inspired by this) for the arbitrary set $X$, define a sequence of open sets $X_n$ such that $X_n \supseteq X$ and $\mu(X_n) < \mu^*(X) + \frac{1}{n}$ (we can do this by definition of outer measure). Then, by Lemma 4.7 (because $A$ is assumed to be Lebesgue measurable, so 4.10 gives that $(X_n \cap A)$ is Lebesgue measurable in an open set $X_n$), we have that $\mu(X_n) = \mu(X_n \cap A) + \mu(X_n \setminus (X_n \cap A)) = \mu(X_n \cap A) + \mu(X_n \setminus A)$, and so

$$\mu^*(X) > \mu(X_n) - \tfrac{1}{n} = \mu(X_n \cap A) + \mu(X_n \setminus A) - \tfrac{1}{n} \geq \mu(X \cap A) + \mu(X \setminus A) - \tfrac{1}{n}$$

Because $n$ can be arbitrarily large, we get that $\mu^*(X) \geq \mu(X \cap A) + \mu(X \setminus A)$. The other direction follows from countable additivity. ■

As a recap of this section, we have proved that:

*Theorem 4.4: three equivalent definitions of bounded measurable*

For a bounded set $A$, the three notions of bounded measurability: Lebesgue, open-closed, and Carathéodory, are equivalent; i.e. any bounded set that satisfies one satisfies the other two.

# 5 Unbounded Measurability and Sigma Algebras

We are now ready to extend our ideas about measurability to unbounded sets — we will propose two definitions inspired by the bounded measurability definitions from the previous section, and eventually see that they are equivalent:

*Definition 5.1: Carathéodory measurable*

For some arbitrary set $A$, if for every $X \subseteq \mathbb{R}$, $\mu^*(X \cap A) + \mu^*(X \setminus A) = \mu^*(A \cap X) + \mu^*(X \cap A^{\complement}) = \mu^*(X)$, then $A$ is Carathéodory measurable.

*Definition 5.2: open-closed measurable*

For some arbitrary set $A$, if for every $\epsilon > 0$, there is an open set $U$ and closed $F$ such that $F \subseteq A \subseteq U$ and $\mu(U \setminus F) < \epsilon$, then $A$ is open-closed measurable.

I've already told you what the goal of this section is (proving the equivalence of the two definitions), but along the way, we'll prove some lemmas and develop some important structures:

*Lemma 5.1: countable intersections are open-closed measurable*

For arbitrary open-closed measurable sets $A_1, A_2, \ldots$, $A = \bigcap_{n=1}^{\infty} A_n$ is also open-closed measurable.

*Proof:* the same proof as Lemma 4.11 with references to Lemma 4.9 replaced with references to Definition 5.2, "Lebesgue" replaced with "open-closed", and "$K$" replaced with "$F$". ■

*Lemma 5.2: intersections are Carathéodory measurable*

For arbitrary Carathéodory measurable sets $A, B$, $A \cap B$ is also Carathéodory measurable.

*Proof:* for any $X \subseteq \mathbb{R}$, and by the Carathéodory measurability of $A$ and $B$, we have that

$$\begin{aligned}
\mu^*(X) &= \mu^*(X \cap A) + \mu^*(X \cap A^\complement) \\
&= \mu^*((X \cap A) \cap B) + \mu^*((X \cap A) \cap B^\complement) + \mu^*((X \cap A^\complement) \cap B) + \mu^*((X \cap A^\complement) \cap B^\complement) \\
&\geq \mu^*(X \cap A \cap B) + \mu^*\Big(\Big(X \cap A \cap B^\complement\Big) \cup \Big(X \cap A^\complement \cap B\Big) \cup \Big(X \cap A^\complement \cap B^\complement\Big)\Big) \\
&= \mu^*(X \cap (A \cap B)) + \mu^*(X \cap (A \cap B)^\complement)
\end{aligned}$$

where the inequality is due to subadditivity. ■

We keep seeing these type of results pop up; so far we've proved that:

- For general open-closed measurable sets, we established in Lemma 5.1 that countable intersections preserve general open-closed measurability.
- Carathéodory measurable sets remain Carathéodory measurable after:
  - finite intersections (apply induction to the two case in Lemma 5.2)
  - complements (obvious from the definition; no caveats about boundedness like in Lemma 4.8)
- Early in section 4, we had Theorem 4.1, which said a countable union of disjoint bounded Lebesgue measurable sets was also bounded Lebesgue measurable.
- We ultimately showed that bounded (Lebesgue/open-closed/Carathéodory) measurable sets remained bounded measurable after:
  - complements within an open set $X$ (Lemma 4.8)
  - countable intersections (Lemma 4.11)

With all these facts in mind, we freestyle a couple definitions (the first condition in these definitions will always be to make sure we're talking about non-empty collections):

*Definition 5.3: $\pi$-system*

A collection $\mathscr{P}$ of sets (subsets of some overarching set $\Omega$) is called a $\pi$-system if

- $\mathscr{P}$ is not empty
- $A, B \in \mathscr{P} \implies A \cap B \in \mathscr{P}$

*Definition 5.4: algebra (or field)*

A collection $\mathscr{A}$ of sets (subsets of some overarching set $\Omega$) is called an algebra (or a field) if

- $\varnothing \in \mathscr{A}$
- $A \in \mathscr{A} \implies A^{\complement} = \Omega \setminus A \in \mathscr{A}$
- $A, B \in \mathscr{A} \implies A \cap B \in \mathscr{A}$ (we could replace $\cap$ with $\cup$ by DeMorgan's laws)

*Definition 5.5: $\lambda$-system*

A collection $\mathscr{L}$ of sets (subsets of some overarching set $\Omega$) is called a $\lambda$-system if

- $\varnothing \in \mathscr{L}$
- $A \in \mathscr{L} \implies A^{\complement} = \Omega \setminus A \in \mathscr{L}$
- disjoint $A_1, A_2, \ldots \in \mathscr{L} \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{L}$

*Definition 5.6: $\sigma$-algebra (or $\sigma$-field)*

A collection $\mathscr{A}$ of sets (subsets of some overarching set $\Omega$) is called a $\sigma$-algebra (or $\sigma$-field) if

- $\varnothing \in \mathscr{A}$
- $A \in \mathscr{A} \implies A^{\complement} = \Omega \setminus A \in \mathscr{A}$
- $A_1, A_2, \ldots \in \mathscr{A} \implies \bigcap_{n=1}^{\infty} A_n \in \mathscr{A}$ (we could replace $\cap$ with $\cup$ by DeMorgan's laws)

Just to familiarize you with the language, we rephrase our results in the above bulleted list using the new terminology (notice that $\varnothing$ is measurable, which is obvious from the bounded/general Carathéodory definition):

*Example 5.1: fitting previous theorems into new framework*

- general open-closed measurable sets form a $\pi$-system (where $\Omega = \mathbb{R}$)
- general Carathéodory measurable sets form an algebra (where $\Omega = \mathbb{R}$)
- bounded (Lebesgue/open-closed/Carathéodory) measurable sets form a $\lambda$-system (where $\Omega = X$ for an open bounded set $X$)
- bounded (Lebesgue/open-closed/Carathéodory) measurable sets form a $\sigma$-algebra (where $\Omega = X$ for an open bounded set $X$)

We are at a good place to finish off the theorem we wanted to prove at the start of the section:

**Theorem 5.1: Carathéodory measurable equivalent to open-closed measurable**

For a set $A$, every $X \subseteq \mathbb{R}$ satisfies $\mu^*(X) = \mu^*(X \cap A) + \mu^*(X \setminus A)$ if and only if for every $\epsilon > 0$, there is an open set $U$ and closed $F$ such that $F \subseteq A \subseteq U$ and $\mu(U \setminus F) < \epsilon$.

*Proof:* the case where $A$ is bounded is done above, so let $A$ be unbounded.
( $\Longleftarrow$ ) if $\mu^*(X) = \infty$, the RHS is also $\infty$ by the countable additivity inequality, and so the equality holds. Now we can assume that $\mu^*(X) < \infty$. Then, simply use the proof of ( $\Longrightarrow$ ) in Theorem 4.3 (with the reference to Lemma 4.10 replaced with a reference to Lemma 5.1).

( $\Longrightarrow$ ) fix any $\epsilon > 0$ and denote $S_1 = (-1,1), S_2 = (-2,-1] \cup [1,2), \ldots, S_n = (-n,-n+1] \cup [n-1,n), \ldots$ (which are all bounded Lebesgue measurable, as disjoint unions of intervals), so that for all $n \in \mathbb{N}$, $A \cap S_n$ is bounded and Carathéodory measurable (by 5.2). Then, we can apply Theorem 4.4 to get $K_n, U_n$ such that $K_n \subseteq A \cap S_n \subseteq U_n$ and $\mu(U_n \setminus K_n) < \frac{\epsilon}{2^n}$. But then we have that $\bigcup_{n=1}^{\infty} K_n \subseteq \bigcup_{n=1}^{\infty} (A \cap S_n) = A \subseteq \bigcup_{n=1}^{\infty} U_n$, and that

$$\left(\bigcup_{n=1}^{\infty} U_n\right) \setminus \left(\bigcup_{n=1}^{\infty} K_n\right) = \bigcup_{n=1}^{\infty} \Big(U_n \setminus \left(\bigcup_{n=1}^{\infty} K_n\right)\Big) \subseteq \bigcup_{n=1}^{\infty} \Big(U_n \setminus K_n\Big),$$

so

$$\mu\left(\left(\bigcup_{n=1}^{\infty} U_n\right) \setminus \left(\bigcup_{n=1}^{\infty} K_n\right)\right) \leq \mu\left(\bigcup_{n=1}^{\infty} \Big(U_n \setminus K_n\Big)\right) \leq \sum_{n=1}^{\infty} \mu(U_n \setminus K_n) < \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon.$$

The union of open sets is open, and the union of a locally finite collection of closed sets is closed. ■

*Remark:* one can take the definition of Lebesgue measurability for an unbounded set $A$ to be that $A \cap [-n,n]$ is Lebesgue measurable for all $n \in \mathbb{N}$; I personally did not find this to the most elegant way to introduce the concept of measurability for unbounded sets (I mean, the choice of $[-n,n]$ is kind of arbitrary, and one kind of has to know that $[-n,n]$ is measurable before making such a definition), but the reader can decide whether or not this definition is worth thinking about. An exercise could be to prove that unbounded Lebesgue measurability $\iff$ unbounded Carathéodory measurability.

# 6 Lebesuge Measure and Extension

We are now fast approaching the finish line; the theorems of this section will all be of great importance, being after all the fruits of the hard work we've done in the last dozen or so pages.

**Theorem 6.1: Carathéodory measurable sets form a $\sigma$-algebra**

The collection $\mathscr{C}$ of all sets $A$ such that $\mu^*(X) = \mu^*(X \cap A) + \mu^*(X \setminus A)$ for all $X \subseteq \mathbb{R}$ is a $\sigma$-algebra.

*Proof:* [3] as we wrote out in Example 5.1, general Carathéodory measurable sets form an algebra on $\mathbb{R}$, so we just need to prove either countable unions or countable intersections (we'll go with unions).

Consider measurable $A_1, A_2, \ldots \in \mathscr{C}$ and $B_n = \bigcup_{k=1}^n A_k$, where all the $B_n$ are also in $\mathscr{C}$ because we know $\mathscr{C}$ is an algebra. By the definition of Carathéodory measurability, we have that (remembering that $B_n^{\complement} \supseteq A^{\complement}$ and monotonicity of outer measure):

$$\begin{aligned}
\mu^*(X) &= \mu^*(X \cap B_n) + \mu^*(X \cap B_n^{\complement}) \geq \mu^*(X \cap B_n) + \mu^*(X \cap A^{\complement}) \\
&= \mu^*((X \cap B_n) \cap A_1) + \mu^*((X \cap B_n) \cap A_1^{\complement}) + \mu^*(X \cap A^{\complement}) \\
&= \mu^*(X \cap A_1) + \mu^*\left(X \cap \bigcup_{k=2}^n A_k\right) + \mu^*(X \cap A^{\complement}) \\
&= \mu^*(X \cap A_1) + \mu^*\left(\left(X \cap \bigcup_{k=2}^n A_k\right) \cap A_2\right) + \mu^*\left(\left(X \cap \bigcup_{k=2}^n A_k\right) \cap A_2^{\complement}\right) + \mu^*(X \cap A^{\complement}) \\
&= \mu^*(X \cap A_1) + \mu^*(X \cap A_2) + \mu^*\left(X \cap \bigcup_{k=3}^n A_k\right) + \mu^*(X \cap A^{\complement}) \\
&= \ldots = \sum_{k=1}^n \mu^*(X \cap A_k) + \mu^*(X \cap A^{\complement})
\end{aligned}$$

Because this inequality is true for every $n \in \mathbb{N}$, it holds for the limit as $n \to \infty$ as well, so

$$\mu^*(X) \geq \sum_{k=1}^\infty \mu^*(X \cap A_k) + \mu^*(X \cap A^{\complement}) \geq \mu^*(X \cap A) + \mu^*(X \cap A^{\complement})$$

where the second inequality follows by countable subbadditivity of outer measure. The other direction ($\leq$) also follows by subbadditivity of outer measure, and so we have equality and the claim is proven. ■

Because we know that all open intervals are measurable, the above theorem tells us that EVERY set in the $\sigma$-algebra generated by open intervals in $\mathbb{R}$ is measurable (this $\sigma$-algebra has a name actually: the Borel sets/subsets of $\mathbb{R}$, denoted $\mathscr{B}$). Here's a list to give you an idea of what sets $\mathscr{B}$ contains:

*Example 6.1: Borel sets*

- Any open interval is in $\mathscr{B}$ because $\mathscr{B}$ by definition contains all the open intervals of $\mathbb{R}$.
- By closure of countable unions, any (countable) collection of open intervals is in $\mathscr{B}$; this means that any open set in $\mathbb{R}$ is in $\mathscr{B}$.
- Any closed set is in $\mathscr{B}$ because closed sets are (by definition) complements of open sets.
- Any interval of the form $(a, b]$ is in $\mathscr{B}$ because $(a, b] = (a, \frac{a+b}{2}) \cup [\frac{a+b}{2}, b]$.
- Any singleton $\{x_0\}$ is in $\mathscr{B}$ because its closed.
- Any countable list of points $\{x_0, x_1, x_2, x_3, \ldots\}$ is in $\mathscr{B}$ because it's just a countable union of singletons, so e.g. $\mathbb{Q} \in \mathscr{B}$ (and hence the set of irrationals is in $\mathscr{B}$ too, by complements)
- The Cantor set $= \bigcap_{i=1}^\infty \bigcap_{j=1}^{3^{i-1}-1} \left[0, \frac{3j+1}{3^i}\right] \cup \left[\frac{3j+2}{3^i}, 1\right]$ is in $\mathscr{B}$ because it's a countable intersection of unions of closed intervals, which are in $\mathscr{B}$.

We now have an update to Theorem 4.1 to encompass general measurable sets as well:

**Theorem 6.2: Outer measure is countably additive for general Carathéodory measurable sets**

For general disjoint measurable $A_1, A_2, \ldots$, and $A = \bigcup_{n=1}^{\infty} A_n$, $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$.

*Proof:* take the equality $\mu^*(X) = \sum_{k=1}^{\infty} \mu^*(X \cap A_k) + \mu^*(X \cap A^{\complement})$ from the proof of Theorem **??**, and set $X = A$ to yield $\mu^*(A) = \sum_{k=1}^{\infty} \mu^*(A_k)$. We can replace $\mu^*$ with $\mu$ because all sets $A_i$ and $A$ are measurable. ■

To recap, we have found a collection $\mathscr{B} \subset 2^{\mathbb{R}}$ such that $\mu^*$, a function taking a set $A$ to the infimum of the total lengths (according to the baby seedling measure $\ell$) over all coverings of $A$ consisting of sets from $\mathscr{O}$ (the set of open intervals), is countably additive for disjoint sets in $\mathscr{B}$. Making the definition:

**Definition 6.1: measure on a collection $\mathscr{C}$**

A function $\mu : \mathscr{C} \to [0, \infty]$ is called a measure on $\mathscr{C}$ if for disjoint $C_1, C_2, \ldots \in \mathscr{C}$ such that $C = \bigcup_{k=1}^{\infty} C_k$ is also in $\mathscr{C}$, $\mu(C) = \sum_{k=1}^{\infty} \mu(C_k)$.

we see that our $\mu^*$ (Lebesgue outer measure) is in fact a measure on $\mathscr{B}$ — the Lebesgue measure. However, the story is not over yet; you see, the machinery we've so painstakingly developed? It's not just good for $\mathbb{R}$. Let's say we have some overarching set $\Omega$ and some collection $\mathscr{C}$ of subsets of $\Omega$ (requiring that $\varnothing \in \mathscr{C}$ so the above countable union and sum can be reduced to a finite union/sum, and that $\Omega \in \mathscr{C}$ to guarantee that a covering exists), and some baby seedling measure $m$ on $\mathscr{C}$. Then one could define the outer measure to be

**Definition 6.2: general outer measure**

Define $\mu^* : 2^{\Omega} \to [0, \infty]$ to be $\mu^*(A) = \inf\left\{ \sum_{k=1}^{\infty} m(C_k) : A \subseteq \bigcup_{k=1}^{\infty} C_k, C_k \in \mathscr{C} \right\}$

which still satisfies monotonicity and countable subadditivity (the proofs are word-for-word the proofs of Lemmas 3.1 and 3.3). Measurability can still be defined using the general Carathéodory definition (which as opposed to the open-closed definition does not require a notion of open or closed):

**Definition 6.3: measurable sets**

For some arbitrary set $A$, if for every $X \subseteq \Omega$, $\mu^*(X \cap A) + \mu^*(X \setminus A) = \mu^*(A \cap X) + \mu^*(X \cap A^{\complement}) = \mu^*(X)$, then $A$ is Carathéodory measurable (or just measurable).

which means that the proof of Theorem 6.1 (and Lemma 5.2 before it), and therefore the proof of Theorem 6.2) carry over word-for-word as well. Finally, in order to get a true extension from the measure $m$ on $\mathscr{C}$ to $\mu = \mu^*$ on the $\sigma$-algebra generated by $\mathscr{C}$ (denoted $\sigma[\mathscr{C}]$), we want that $\mu^*(C) = m(C)$ for all $C \in \mathscr{C}$ (so that $\mu^*$ is in fact an extension of $m$ and not some unrelated measure), and that all $C \in \mathscr{C}$ are measurable (in order for Theorems 6.1 and 6.2 to kick in).

Unfortunately, these claims are not true for arbitrary collections $\mathscr{C}$; take for example $\Omega = [0,2]$ and $\mathscr{C} = \{\varnothing, [0,1], (1,2], (0.5,1.5), \Omega\}$ with $m$ values $\{0,1,1,10,2\}$ respectively (check that despite its weird definition, $m$ is in fact a measure on $\mathscr{C}$ as in the definition above; the issue is that Lemma 2.1 need not hold in arbitrary collections $\mathscr{C}$ because those collections may not be closed under set differences) which has that $\mu^*((0.5,1.5))$ is in fact 2, not $m((0.5,1.5)) = 10$. Moreover, the property of being closed under set differences and unions/intersections is also quite important in the Carathéodory definition, and so it seems reasonable to consider $\mathscr{C}$ that do have those properties. We'll even give such collections a name:

*Definition 6.4: weak semi-ring*

A collection $\mathscr{W}$ of sets (subsets of some overarching set $\Omega$) is called a weak semi-ring if

- $A, B \in \mathscr{W} \implies A \setminus B \in \mathscr{W}$
- $A, B \in \mathscr{W} \implies A \cap B \in \mathscr{W}$ (actually unnecessary because $A \setminus (A \setminus B) = A \cap B$)

If our collection $\mathscr{C}$ is a weak semi-ring, then we'll be able to get the prove the two remaining claims. First, some lemmas:

*Lemma 6.1: monotonicity for weak semi-rings*

If $A, B \in \mathscr{C}$ and $A \subseteq B$, $m(A) \leq m(B)$

*Proof:* $A$ is in $\mathscr{C}$, and from the set difference property of weak semi-rings $B \setminus A$ is also $\in \mathscr{C}$. Furthermore, $A$ and $B \setminus A$ are disjoint and have union $B$, which is in $\mathscr{C}$, so by the definition of the measure $m$ on $\mathscr{C}$, $m(B) = m(A) + m(B \setminus A) \geq m(A)$ (because $m \geq 0$). ■

*Lemma 6.2: countable subadditivity for weak semi-rings*

For any $A_1, A_2, \ldots \in \mathscr{C}$ such that $A = \bigcup_{k=1}^{\infty} A_k$ is also in $\mathscr{C}$, $m(A) \leq \sum_{k=1}^{\infty} m(A_k)$

*Proof:* consider $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus (A_1 \cup A_2) = (A_3 \setminus A_1) \cap (A_3 \setminus A_2)$, and so on; i.e. $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) = (A_n \setminus A_1) \cap \ldots \cap (A_n \setminus A_{n-1})$. By the set difference property and intersection property of weak semi-rings all the $B_n$ are also $\in \mathscr{C}$. From monotonicity (6.1), $\mu(B_i) \leq \mu(A_i)$. These $B_i$ are clearly disjoint, and have union equal to the union of the $A_i$ (i.e. $A \in \mathscr{C}$), and so the claim follows by the definition of the measure $m$ on $\mathscr{C}$ and monotonicity:

$$m(A) = m\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} m(B_k) \leq \sum_{k=1}^{\infty} m(A_k) \quad ■$$

Now the claims:

*Lemma 6.3: $\mu^*$ and $m$ agree for sets in $\mathscr{C}$*

For $C \in \mathscr{C}$, $\mu^*(C) = m(C)$

*Proof:* trivially, $\mu^*(C) \le m(C)$ because $\{C, \varnothing, \dots\}$ is a covering of $C$. For the other direction, consider any covering $\{C_1, C_2, \dots\}$ of $C$. Now $\bigcup_{k=1}^{\infty}(C_k \cap C) = C \in \mathscr{C}$, and $(C_k \cap C) \in \mathscr{C}$ by the intersection property of weak semi-rings, so by countable subadditivity and monotonicity,

$$m(C) = m\left(\bigcup_{k=1}^{\infty}(C_k \cap C)\right) \le \sum_{k=1}^{\infty} m(C_k \cap C) \le \sum_{k=1}^{\infty} m(C_k)$$

Because this inequality is true for every covering, it is true for the infimum, so $m(C) \le \mu^*(C)$. ■

*Lemma 6.4: all sets in $\mathscr{C}$ are measurable*

For all $C \in \mathscr{C}$, $C$ is (Carathéodory) measurable; i.e. for all $X \subseteq \Omega$, $\mu^*(X) = \mu^*(X \cap C) + \mu^*(X \setminus C)$

*Proof:* fix any $\epsilon > 0$. By the definition of outer measure, there is a covering $\{C_1, C_2, \dots\}$ of $X$ such that $\sum_{k=1}^{\infty} m(C_k) \le \mu^*(X) + \epsilon$. But because all the $C_k$ and $C$ are in $\mathscr{C}$, $(C_k \cap C)$ and $(C_k \setminus C)$ are also in $\mathscr{C}$ (by the definition of weak semi-ring), and are disjoint with union $C_k$, meaning that $m(C_k) = m(C_k \cap C) + m(C_k \setminus C)$. Furthermore, $\{(C_1 \cap C), (C_2 \cap C), \dots\}$ and $\{(C_1 \setminus C), (C_2 \setminus C), \dots\}$ are coverings of $(X \cap C)$ and $(X \setminus C)$ respectively, and so putting everything together, we have that

$$\mu^*(X) + \epsilon \ge \sum_{k=1}^{\infty} m(C_k) = \sum_{k=1}^{\infty} m(C_k \cap C) + \sum_{k=1}^{\infty} m(C_k \setminus C) \ge \mu^*(X \cap C) + \mu^*(X \setminus C)$$

To finish things off, $\epsilon > 0$ is arbitrary, and the other direction follows by countable subadditivity of outer measure. ■

*Example 6.2: weak semi-rings*

The collection we started off our study on, $\mathscr{O}$, the set of open intervals is not a weak semi-ring (an open interval differenced with an open interval is not necessarily open). However, the collection of half open intervals of the form $(a, b]$ IS a weak semi-ring. Defining $m((a, b]) = b - a$ (and infinite/null cases appropriately), we recover Lebesgue measure.

Now what if we go into $\mathbb{R}^2$? Will the collection of the analogues of half-open intervals, or more precisely the half-open rectangles $(a_1, b_1] \times (a_2, b_2]$ form a weak semi-ring? Well, not quite. If we take $A = (1, 3] \times (1, 3]$ and $B = (0, 2] \times (0, 2]$, then $A \setminus B$ will look like an L-shaped tromino. However, we can still break it up into three pieces of the aforementioned form, namely $(1, 2] \times (2, 3]$, $(2, 3] \times (2, 3]$, and $(2, 3] \times (1, 2]$. Thus, we define

*Definition 6.5: semi-ring*

A collection $\mathscr{S}$ of sets (subsets of some overarching set $\Omega$) is called a semi-ring if

- $A, B \in \mathscr{S} \implies A \setminus B$ is a disjoint finite union of $S_1, \dots, S_n \in \mathscr{S}$.
- $A, B \in \mathscr{S} \implies A \cap B \in \mathscr{S}$

and redo the four above lemmas where now $\mathscr{C}$ is a semi-ring [2], now as theorems (but with the help of some more lemmas):

*Lemma 6.5: difference of finite union can be written as finite disjoint union*

For any $A, A_1, \ldots, A_n \in \mathscr{C}$ (a semi-ring), there is $M < \infty$ and disjoint sets $C_1, \ldots, C_M \in \mathscr{C}$ such that

$$A \setminus \bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{M} C_k$$

*Proof:* by the definition of semi-ring, $A \setminus A_1 = \bigcup_{k=1}^{m} C_k$ for some $m \in \mathbb{N}$ and disjoint $C_k \in \mathscr{C}$. Then, $A \setminus (A_1 \cup A_2) = (A \setminus A_1) \setminus A_2 = (\bigcup_{k=1}^{m} C_k) \setminus A_2 = \bigcup_{k=1}^{m}(C_k \setminus A_2) = \bigcup_{k=1}^{m} \bigcup_{i=1}^{m_k} D_i$ where for all $k \in \{1, \ldots, m\}$, $D_1, \ldots, D_{m_k}$ are disjoint and have union $C_k \setminus A_2$. Because the $C_k$ are all disjoint, all the $D_i$ are disjoint; thus, we can reindex the $D_i$ to go from $i \in \{1, \ldots, M\}$ where $M = m_1 + \ldots m_k$ and get that $A \setminus (A_1 \cup A_2) = \bigcup_{i=1}^{M} D_i$. The claim for any finite $n$ follows by induction. ■

*Theorem 6.3: monotonicity for semi-rings*

If $A, A_1, \ldots, A_n \in \mathscr{C}$ and $\bigcup_{k=1}^{n} A_k \subseteq A$, $\sum\limits_{k=1}^{n} m(A_k) \leq m(A)$

*Proof:* we know that $A$ and the $A_i$ are in $\mathscr{C}$, and by the above lemma (6.5), $A \setminus \bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{M} C_k$ for disjoint $C_k \in \mathscr{C}$. Furthermore, $A_1, \ldots, A_n, C_1, \ldots, C_M$ are all disjoint and have union $A$, which is in $\mathscr{C}$, so by the definition of the measure $m$ on $\mathscr{C}$, $m(A) = \sum_{k=1}^{n} m(A_k) + \sum_{k=1}^{M} m(C_k) \geq \sum_{k=1}^{n} m(A_k)$ (because $m \geq 0$). ■

*Theorem 6.4: countable subadditivity for semi-rings*

For any $A_1, A_2, \ldots \in \mathscr{C}$ such that $A = \bigcup\limits_{k=1}^{\infty} A_k$ is also in $\mathscr{C}$, $m(A) \leq \sum\limits_{k=1}^{\infty} m(A_k)$

*Proof:* consider $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus (A_1 \cup A_2)$, and so on; i.e. $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$. By the above lemma (6.5), each $B_n$ can be written as a disjoint union of $C_{n,1}, \ldots, C_{n,m_n} \in \mathscr{C}$. From Theorem **??** (monotonicity), $\sum_{k=1}^{m_n} m(C_{n,k}) \leq m(A_n)$. All of these $\{C_{1,1}, \ldots, C_{1,m_1}, \ldots, C_{n,1}, \ldots, C_{n,m_n}, \ldots\}$ are clearly disjoint, and have union equal to the union of the $A_n$ (i.e. $A \in \mathscr{C}$), and so the claim follows by the definition of the measure $m$ on $\mathscr{C}$ and monotonicity:

$$m(A) = m\left(\bigcup_{n=1}^{\infty} \bigcup_{k=1}^{m_n} C_{n,k}\right) = \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} m(C_{n,k}) \leq \sum_{k=1}^{\infty} m(A_k) \quad ■$$

Now the claims:

*Theorem 6.5: $\mu^*$ and $m$ agree for sets in $\mathscr{C}$*

For $C \in \mathscr{C}$, $\mu^*(C) = m(C)$

*Proof:* trivially, $\mu^*(C) \leq m(C)$ because $\{C, \varnothing, \ldots\}$ is a covering of $C$. For the other direction, consider any covering $\{C_1, C_2, \ldots\}$ of $C$. Now $\bigcup_{k=1}^{\infty}(C_k \cap C) = C \in \mathscr{C}$, and $(C_k \cap C) \in \mathscr{C}$ by the intersection property of semi-rings, so by countable subadditivity and monotonicity,

$$m(C) = m\left(\bigcup_{k=1}^{\infty}(C_k \cap C)\right) \leq \sum_{k=1}^{\infty} m(C_k \cap C) \leq \sum_{k=1}^{\infty} m(C_k)$$

Because this inequality is true for every covering, it is true for the infimum, so $m(C) \leq \mu^*(C)$. ■

*Theorem 6.6: all sets in $\mathscr{C}$ are measurable*

For all $C \in \mathscr{C}$, $C$ is (Carathéodory) measurable; i.e. for all $X \subseteq \Omega$, $\mu^*(X) = \mu^*(X \cap C) + \mu^*(X \setminus C)$

*Proof:* fix any $\epsilon > 0$. By the definition of outer measure, there is a covering $\{C_1, C_2, \ldots\}$ of $X$ such that $\sum_{k=1}^{\infty} m(C_k) \leq \mu^*(X) + \epsilon$. But because all the $C_k$ and $C$ are in $\mathscr{C}$, $(C_k \cap C) \in \mathscr{C}$ and $(C_k \setminus C) = \bigcup_{i=1}^{m_k} D_{k,i}$ for $D_{k,i} \in \mathscr{C}$ (by the definition of semi-ring), and where the sets $\{(C_k \cap C), D_{k,1}, \ldots, D_{k,m_k}\}$ are all disjoint with union $C_k$, meaning that $m(C_k) = m(C_k \cap C) + \sum_{i=1}^{m_k} m(D_{k,i})$. Furthermore, $\{(C_1 \cap C), (C_2 \cap C), \ldots\}$ and $\{D_{1,1}, \ldots, D_{1,m_1}; \ldots; D_{k,1}, \ldots, D_{k,m_k}; \ldots\}$ are coverings of $(X \cap C)$ and $(X \setminus C)$ respectively, and so putting everything together, we have that

$$\mu^*(X) + \epsilon \geq \sum_{k=1}^{\infty} m(C_k) = \sum_{k=1}^{\infty} m(C_k \cap C) + \sum_{k=1}^{\infty}\sum_{i=1}^{m_k} m(D_{k,i}) \geq \mu^*(X \cap C) + \mu^*(X \setminus C)$$

To finish things off, $\epsilon > 0$ is arbitrary, and the other direction follows by countable subadditivity of outer measure. ■

To summarize the main result of this section and the culmination of our twenty pages of work, we give the statement of the famed Carathéodory extension theorem:

*Theorem 6.7: Carathéodory extension theorem*

Let $\Omega$ be some overarching set, and let $\mathscr{C}$ be a semi-ring of subsets of $\Omega$. Let $m$ be a measure on $\mathscr{C}$. Then,

$$\mu^*(A) = \inf\left\{\sum_{k=1}^{\infty} m(C_k) : A \subseteq \bigcup_{k=1}^{\infty} C_k, C_k \in \mathscr{C}\right\}$$

is a measure on $\sigma[\mathscr{C}]$ satisfying $\mu^*(C) = m(C)$ for all $C \in \mathscr{C}$.

*Remark:* there are statements about the uniqueness of the extension as well; perhaps I will expand upon this at some later revision of the document.

# Concluding Remarks

The Carathéodory extension theorem opens many doors for us; in one fell swoop, we have Lebesgue measure on the Borel sets of $\mathbb{R}^n$, for ALL $n \in \mathbb{N}$ (and in far more general spaces). From here, one can define a very powerful theory of integration on very general spaces. However, to tie back into the opening section on probability, the Carathéodory extension theorem allows us to think of distribution functions as measures. As a high level overview, let $X : \Omega \to \mathbb{R}$ be some function on some $\Omega$ equipped with a measure $P$ such that $P(\Omega) = 1$. Then, define the cumulative distribution function $F_X(x)$ to be $P(X^{-1}(-\infty, x])$, otherwise denoted as $P([X \leq x])$. Then, it is clear that $P([a < X \leq b]) = F_X(b) - F_X(a)$. $F_X$ is clearly a function $\mathbb{R} \to [0, 1]$, but we can change our perspective a little bit. Define $F_X((a, b])$ to be $F_X(b) - F_X(a)$. Now, this is a measure on the semi-ring of sets of the form $(a, b] \subset \mathbb{R}$, which we can extend to the Borel sets of $\mathbb{R}$ by the extension theorem. In this way, we've transformed a question about distribution functions to measures. (In general, measures defined this way are called Lebesgue-Stieltjes measure).

Moreover, the extension theorem can be used to construct arbitrary product measures (not just on $\mathbb{R}^n$), and with additional work (the Kolmogorov extension theorem), we can get measures on infinite dimensional spaces, which can be used in probability theory to construct stochastic processes in continuous time. For those with less interest in probability, the "outer measure to bona-fide measure" part of the proof (i.e. ignoring the baby seedling measure and the semi-ring of starting sets), one can construct the Hausdorff measure, which is of great importance in pure analysis.

This is just a taste of measure-theoretic probability, and there is much to look forward to/be interested in. With measures, one gives probability a rigorous framework on which to build, and allows mathematicians to study questions about randomness and chance with the powerful tools of analysis. Without measure theory, the tools available were vague notions about chance, combinatorics, and distribution functions (which are notoriously difficult to calculate, usually involving messy integrals), but with measure theory, we can study the random variables themselves as functions, distribution functions as measures, expectations as integrals, information as the sets available to us in the $\sigma$-algebras, and limiting behavior as convergence of functions/measures/integrals/$\sigma$-algebras.

# References


[1] Sheldon Axler. *Measure, Integration, & Real Analysis*. Vol. 282. Graduate Texts in Mathematics. Springer International Publishing, 2020, pp. XVIII, 411. ISBN: 978-3-030-33143-6, 978-3-030-33142-9.

[2] Stanley Sawyer. "Measures on Semi-Rings in $\mathbb{R}^1$ and $\mathbb{R}^k$". Handout. Sep. 2009. URL: https://www.math.wustl.edu/~sawyer/handouts/MeasuresSemiRings.pdf.

[3] Galen Shorack. *Probability for Statisticians*. 2nd. Springer Texts in Statistics. Springer-Verlag New York, 2000, pp. XVIII, 586. ISBN: 978-0-387-22760-3, 978-0-387-98953-2.

[4] Shlomo Sternberg. "Lebesgue Measure". Math212a1411 Lecture. Oct. 2014. URL: http://people.math.harvard.edu/~shlomo/212a/11.pdf.