Just Daniel's Trivial Ramblings

Daniel Rui - January 5, 2020

Abstract

My sketches/notes on measure and probability theory. Meant to serve as a guide to Axler's *Measure, Integration, and Real Analysis*. After writing about 6 sections, I now see that the title is rather...inaccurate, but I like it, so stay, it shall.

1 Probability!

The definition of probability measure (a function that maps select events of a sample space Ω, i.e. select "subsets" of Ω, to real numbers) consists of three parts:

1. $P(\Omega) = 1$
2. For any event $\mathcal{E}$, $P(\mathcal{E}) \geq 0$
3. Countable Additivity: if a countably infinite family of pairwise disjoint sets $\{A_1, A_2, \ldots\}$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

This video gives some simple exercises that are within my intellectual capability (lol) so here goes:

1. $P(\varnothing) = 0$ because for the family of pairwise disjoint subsets $\{\Omega, \varnothing, \varnothing, \ldots\}$, which is pairwise disjoint because there does not exist any element in $\varnothing$ and Ω (because there exists no elements in $\varnothing$) so using part three of the definition we get that $P(\Omega \cup \varnothing \cup \ldots) = P(\Omega) + P(\varnothing) + \ldots$ but $\Omega \cup \varnothing \cup \ldots = \Omega$ (again lol because there exists no elements in $\varnothing$) so $P(\Omega) = P(\Omega) + P(\varnothing) + \ldots$ so $P(\varnothing)$ must not be a positive quantity else the sum would reach infinity, but contradiction because $P(\Omega) = 1$.
2. $P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$ is literally just part three of the definition using $\{A_1, \ldots, A_n, \varnothing, \ldots\}$
3. Using the just proved (2) take the disjoint subsets of Ω: $\{A, A^{\complement}\}$ to see that $P(A \cup A^{\complement}) = P(A) + P(A^{\complement})$. But $A \cup A^{\complement} = \Omega$ by definition of complement so $P(S) = 1 = P(A) + P(A^{\complement})$ so $P(A^{\complement}) = 1 - P(A)$.
4. If $A \subset B$, then $P(A) \leq P(B)$ because $A \cap B \setminus A = B$, so $P(B) = P(A) + P(B \setminus A) \geq P(A)$ because all probabilities are ≥ 0 by part two of the definition.
5. Finally, all sets satisfy $\varnothing \subseteq A \subseteq \Omega$ so $0 \leq P(A) \leq 1$ for all subsets of the sample space.

The video goes on to talk about the principle of inclusion-exclusion, which I won't talk about here. I guess go read about it elsewhere (like the Axler book). The thing I want to think about here is his definition of independence. He first defines conditional probability (denoted $P_B(A)$ in Axler's book)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

and then he says independence is defined as when $P(A|B) = P(A)$ because then

$$P(A|B^{\complement}) = \frac{P(A \cap B^{\complement})}{P(B^{\complement})}$$

but we know that $P(A) = P((A \cap B) \cup (A \cap B^{\complement})) = P(A \cap B) + P(A \cap B^{\complement})$ so we can write

$$P(A|B^{\complement}) = \frac{P(A \cap B^{\complement})}{P(B^{\complement})} = \frac{P(A) - P(A \cap B)}{1 - P(B)}$$

but by conditional probability $P(A \cap B) = P(A|B) \cdot P(B)$ which equals $P(A) \cdot P(B)$ because $P(A \cap B) = P(A)$ by independence so

$$P(A|B^{\complement}) = \frac{P(A \cap B^{\complement})}{P(B^{\complement})} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{P(A) - P(A) \cdot P(B)}{1 - P(B)} = P(A)$$

so regardless whether or not an event in B happens $P(A|B) = P(A|B^{\complement}) = P(A)$, which is our intuitive definition of independence because the probability of A given B and $B^{\complement}$ is the same.

1.1 Axler's definition

Axler's definition of independence is $P(A \cap B) = P(A) \cdot P(B)$. From our definition of conditional probability and independence: $P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)}$, we see that Axler's definition is equivalent.

2 Sigma Algebras

Before we go further into probability theory, we first have to figure out what exactly *is* probability, and how to study it mathematically. First, we must define a set Ω (a universal set) where all the action takes place (i.e. a "space" where we can build structures and functions). From above, you saw that we had some sort of probability "function" which we gave "events" (which were subsets of a sample space Ω) as inputs, and which returned a real number between 0 and 1 that we called the "probability".

We can think of what the probability function is doing as looking at the "size" of a subset of Ω — the "bigger" an event, the higher the probability and the "smaller" an event, the lower the probability. But how can we define a notion of "size" on events/sets, and what sets can we apply it to? We will actually answer the second question first: we define a structure on Ω called a σ-algebra, which will hopefully remind you of the probability axioms and exercises above.

An important starting point when talking about sigma algebras is the power set, denoted $\mathscr{P}(\Omega)$ or 2^Ω, meaning the set of all subsets of Ω. $\mathscr{P}(\Omega)$ is one example of a sigma algebra on Ω. Now for the definition of a σ-algebra: a set $\mathcal{S}$ is a σ-algebra (on Ω) if

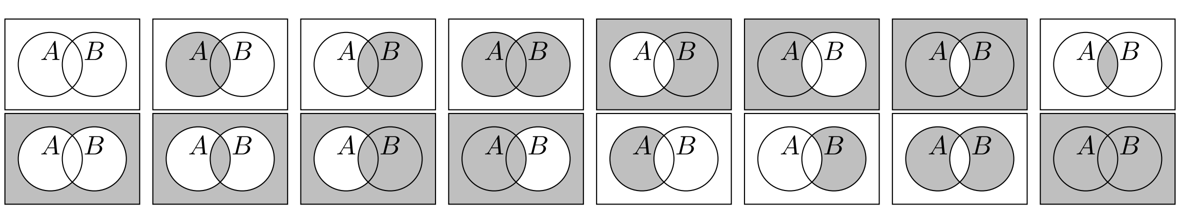
1. $\varnothing$ is an element of $\mathcal{S}$
2. for some subset $A \subseteq \Omega$, if $A \in \mathcal{S}$, then $\Omega \setminus A$ is too (closed under complements)
3. for some subsets $A_i \subseteq \Omega$ already elements of $\mathcal{S}$, $\bigcup_{n=1}^{\infty} A_n$ is too (closed under countable unions)

You may ask, "why not *finite* unions?" — well, intuitively, finite unions don't get enough structure to be particularly interesting; however, they are useful enough to warrant a name: "algebra". Uncountable unions on the other hand give too much structure — basically, this condition would allow some really really nasty sets inside the sigma algebra, but it's complicated enough topic that we'll have to come back at a later time. Just to hammer in this point, **all elements of a σ-algebra are subsets of Ω. So, to be clear with notation, $A \in \mathcal{S}$ but $A \subseteq \Omega$.** Some examples of sigma algebras are ones we've already encountered in Axler's book

1. $\{\varnothing, \Omega\}$
2. $\{\varnothing, A, (\Omega \setminus A), \Omega\}$
3. For two sets A, B it gets really hard really fast so I'll list them out vertically for better readability. The bullet points are read top to bottom then left to right.

- $\varnothing$
- A
- B
- $(A \cup B)$
- $(\Omega \setminus A)$
- $(\Omega \setminus B)$
- $((\Omega \setminus A) \cup (\Omega \setminus B))$
- $(\Omega \setminus ((\Omega \setminus A) \cup (\Omega \setminus B)))$
- $(\Omega \setminus (A \cup B))$
- $((\Omega \setminus ((\Omega \setminus A) \cup (\Omega \setminus B))) \cup (\Omega \setminus (A \cup B)))$
- $((\Omega \setminus A) \cup B)$
- $((\Omega \setminus B) \cup A)$
- $(\Omega \setminus ((\Omega \setminus A) \cup B))$
- $(\Omega \setminus ((\Omega \setminus B) \cup A))$
- $((\Omega \setminus ((\Omega \setminus A) \cup B)) \cup (\Omega \setminus ((\Omega \setminus B) \cup A)))$
- Ω

The last one I had to look at some diagrams (but we can be sure we got them all because we can that these 16 sets satisfy the three conditions in the definition of σ-algebra). In left to right then top to bottom order, all the sets (where white means empty and shaded means filled) are



Ouch. Also from above, we see that using deMorgan's laws we see that if A, B are in the sigma algebra, $A \cap B$ is too:

$$((\Omega \setminus A) \cup (\Omega \setminus B)) = \Omega \setminus (A \cap B) \implies (\Omega \setminus ((\Omega \setminus A) \cup (\Omega \setminus B))) = A \cap B$$

We can expand to countably infinite sets; but first we need the countably infinite version of deMorgan's law:

$$\Omega \setminus \left(\bigcap_{n=1}^{\infty} A_n \right) = \bigcup_{n=1}^{\infty} (\Omega \setminus A_n)$$

which we see to be true because in English the left side is "the x's not shared by all of the sets A_i" which is equivalent to the right side "the x's that lie outside of at least one of the A_i". Hence, we see that sigma algebras are also (closed under countable intersection):

$$A_1, A_2, A_3, \ldots \in \mathcal{S} \implies \left(\bigcap_{n=1}^{\infty} A_n \right) \in \mathcal{S}$$

3 Borel Sets

The most important sigma algebra we're going to encounter would be the Borel subsets of $\mathbb{R}$. I'm going to denote it $\mathcal{B}$. $\mathcal{B}$ is the smallest sigma algebra containing all the open intervals of $\mathbb{R}$. A subset $B \in \mathcal{B}$ is called a Borel set. Here are some examples of some Borel sets:

1. Any open interval is in $\mathcal{B}$ obviously because $\mathcal{B}$ by definition contains all the open intervals of $\mathbb{R}$
2. By closure of countable unions, any (countable) collection of open intervals is in $\mathcal{B}$
3. Any closed interval is in $\mathcal{B}$ because closed sets are (by definition) complements of open sets.
4. Any half open interval is in $\mathcal{B}$ because any half open interval can be written as the union of a closed interval and open interval.
5. Any individual point $\{x_0\}$ is in $\mathcal{B}$ because its complement, $(-\infty, x_0) \cup (x_0, \infty)$ is the union of two open intervals.
6. Any countable list of points $\{x_0, x_1, x_2, x_3, \ldots\}$ is in $\mathcal{B}$ because it's just a countable union of $\{x_0\}, \{x_1\}, \{x_2\}, \ldots$. E.g. $\mathbb{Q} \in \mathcal{B}$ (and hence the set of irrationals is in $\mathcal{B}$ too, by complements)

We are ready to go back to the question a while back of why we do not define σ-algebras to contain the *uncountable* union of sets; this is because if we were to allow that, ALL subsets of $\mathbb{R}$ would be in the σ-algebra because ANY set is the uncountable union of the singleton points in the set (and we know that every individual singleton is in $\mathcal{B}$).

Basically all the sets we'll normally encounter are Borel sets; those that aren't are very nasty and maybe I'll talk about them later. (The A in this video: https://www.youtube.com/watch?v=Ur3ofJ61bpk is not Borel. I think the video is quite good, much better than reading it alone from Axler's book, section 2A). BTW proving that it's not Borel requires more work, and maybe I'll talk about that later).

4 Measurable Functions

At this point in time, I'm not sure why we call them measurable functions, but I suppose that will come later. As for now, let's have the definition:

- For a set, Ω, and σ-algebra on Ω, $\mathcal{S}$, an $\mathcal{S}$-measurable function means that $f^{-1}(B)$ is one of the sets in the σ-algebra $\mathcal{S}$ for every Borel set $B \in \mathcal{B}$
- $f^{-1}(B)$ is the set $\{x \in \Omega : f(x) \in B\}$

It's not obvious at all what this definition is doing, so let's see some examples:

1. For $\mathcal{S} = \{\varnothing, \Omega\}$, a function is $\mathcal{S}$-measurable if $\{x \in \Omega : f(x) \in B\}$ equals the empty set or the entire set Ω for every Borel set $B \in \mathcal{B}$. This function can't possibly be a non-constant function, because then there would exist $a, b \in \Omega$ such that $f(a) \neq f(b)$, which would mean that for the Borel set (a singleton) $\{f(a)\}$, $\{x \in \Omega : f(x) = f(a)\}$ would include a (hence not $\varnothing$), but would not include b (hence not the entire set Ω), which is a contradiction. Hence only constant functions can possibly be $\{\varnothing, \Omega\}$-measurable.

   To prove that constant functions *are* actually $\{\varnothing, \Omega\}$-measurable, just note that there are exactly two cases: either $c \in B$ or $c \notin B$ (where c is the value the constant function takes for all $x \in \Omega$). In the first case, $f^{-1}(B) = \Omega$, and in the second, $f^{-1}(B) = \varnothing$. And so $\{\varnothing, \Omega\}$-measurable $\iff f(x) = c, \forall x \in \Omega$

2. For $\mathcal{S} = \{\varnothing, A, (\Omega \setminus A), \Omega\}$, the only $\mathcal{S}$-measurable functions are those that are constant over A and constant over $(\Omega \setminus A)$ (for pretty much the same reasoning as the first example).

3. Generally, if you partition the set Ω into N pairwise disjoint pieces (i.e. $A_i \cap A_j = \varnothing$ and $\bigcup_{i=1}^{N} A_i = \Omega$), then find the smallest σ-algebra that contains all the A_i, the only $\mathcal{S}$-measurable functions are those that are constant on all those A_i (again for the same reasonings as example one), i.e. those that are the sums of characteristic functions of A_i, denoted $1_{A_i}(x)$ or $\chi_{A_i}(x)$, defined as

   $$1_{A_i}(x) = \begin{cases} 1 & x \in A_i \\ 0 & \text{otherwise} \end{cases}$$

   Using these characteristic functions, the $\mathcal{S}$-measurable functions can be written as

   $$\sum_{i=1}^{N} c_i \cdot 1_{A_i}(x)$$

   As you can imagine, as we get finer and finer partitions, these sums of characteristic functions can get better and better at approximating normal continuous functions. (Bonus:) Just to be a little more thorough, let's prove that the smallest sigma algebra containing all the A_i (usually denoted $\sigma(\{A_1, \ldots, A_N\})$) is

   $$\left\{ \bigcup_{i \in I} A_i \text{ for all } I \subseteq \{1, \ldots, N\} \right\}$$

In English that's basically "union over every distinct subset of $\{A_1, \ldots, A_N\}$". To show that $\sigma(\{A_1, \ldots, A_N\} = \left\{\bigcup_{i \in I} A_i : I \subseteq \{1, \ldots, N\}\right\}$, we need to show that, one, the right hand side is actually a σ-algebra, and two, that $A_1, \ldots, A_N \in \mathcal{S} \implies \bigcup_{i \in I} A_i \in \mathcal{S}, \forall I \subseteq \{1, \ldots, N\}$.

Part one is as follows:

(a) $\varnothing \in \left\{\bigcup_{i \in I} A_i : I \subseteq \{1, \ldots, N\}\right\}$ obviously, choosing $I = \varnothing$.

(b) All the complements are also in the set (i.e. can be represented as unions of A_i); more explicitly,
$$(\Omega \setminus (A_i \cup A_j \cup A_k \cup \ldots)) = \bigcup_{i \notin I} A_i \text{ where } I = \{i, j, k, \ldots\}$$

(c) All possible unions are also in the set — we only have the A_i as building blocks, and $\Omega = A_1 \cup \ldots \cup A_N$ already, and we can't get bigger than Ω.

Part two is trivial: by (closure under countable unions) all the sets in $\left\{\bigcup_{i \in I} A_i : I \subseteq \{1, \ldots, N\}\right\}$ must be in the σ-algebra.

4. See this MSE question: `https://math.stackexchange.com/questions/3358230/any-nice-pedagogical-examples-of-s-measurable-functions` for more? It's my question but the answer seems a bit complicated and at present I don't really understand it.

4.1 A small simplification

To check whether a function is $\mathcal{S}$-measurable, you'll notice that we have to find the inverse image of *every* Borel set B and make sure it's in $\mathcal{S}$. For functions more compicated than constant functions however, checking inverse images for all Borel sets is all but infeasible. Fortunately, there's a way to cut down on the amount of work we have to do: if we find a collection of sets, $\mathcal{C}$, such that $\sigma(\mathcal{C}) \supseteq \mathcal{B}$ and $f^{-1}(C) \in \mathcal{S}$ for all $C \in \mathcal{C}$ — then f is $\mathcal{S}$-measurable.

Axler gives a nice proof (section 2B). In short, he proves that the set $\mathcal{A} = \{A \subset \mathbb{R} \text{ where } f^{-1}(A) \in \mathcal{S}\}$ is a sigma algebra (using properties of inverse images and the fact that $\mathcal{S}$ is a sigma algebra). We then know that all the sets in $\mathcal{C}$ are in $\mathcal{A}$ (by definition of $\mathcal{C}$), and any sigma algebra containing $\mathcal{C}$ must contain the smallest sigma algebra of $\mathcal{C}$, which is $\sigma(\mathcal{C}) \supseteq \mathcal{B}$, so $\mathcal{A} \supseteq \mathcal{B}$, and thus finally f must be $\mathcal{S}$-measurable.

- An example: take $\mathcal{C} = \{(a, \infty) \text{ for all } a \in \mathbb{R}\}$. $\sigma(\mathcal{C})$ must contain all $(-\infty, a] = \Omega \setminus (a, \infty)$, and all $(a, b] = (-\infty, b] \cap (a, \infty)$, and hence all finite open intervals $(a, b) = \bigcup_{k=1}^{\infty} \left(a, b - \frac{1}{Nk}\right]$ where N is chosen so that $b - \frac{1}{N} > a$. Lastly, it must also include all $(-\infty, b) = (-\infty, a] \cup (a, b)$. Thus $\sigma(\mathcal{B})$ includes all open intervals, so it must also contain $\mathcal{B}$.

  We can now check for $\mathcal{S}$-measurability by checking if $f^{-1}((a, \infty)) = \{x \in \Omega \text{ where } f(x) > a\} \in \mathcal{S}$ for all $a \in \mathbb{R}$

4.2 Limits of measurable functions

I end this section with the following result: if we have a sequence $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ of $\mathcal{S}$-measurable functions s.t. $\lim_{k\to\infty} f_k(x)$ exists for all $x \in \Omega$, then $f(x) := \lim_{k\to\infty} f_k(x)$ is also $\mathcal{S}$-measurable. This is because

$$\begin{aligned} f^{-1}((a,\infty)) &= \left\{x \in \mathbb{R} : \exists j \text{ s.t. } f(x) > a + \frac{1}{j}\right\} \\ &= \left\{x \in \mathbb{R} : \exists j \in \mathbb{N} \text{ s.t. } \exists m \in \mathbb{N} \text{ s.t. } \forall k \geq m,\ f_k(x) > a + \frac{1}{j}\right\} \\ &= \bigcup_{j=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} f_k^{-1}(\left(a + \frac{1}{j}, \infty\right)) \end{aligned}$$

which is a set in $\mathcal{S}$ because countable unions and intersections of sets in $\mathcal{S}$ are also in $\mathcal{S}$.

5 Borel Measurability

As you might have guessed, Borel measurable functions are those that are $\mathcal{B}$-measurable (where $\mathcal{B}$ denotes the Borel subsets of $\mathbb{R}$, i.e. the smallest sigma algebra containing all the open intervals of $\mathbb{R}$): functions whose inverse images of Borel sets are themselves Borel sets.

What kinds of functions are Borel measurable? Well for starters, all continuous functions are! First, let us denote $U_a := f^{-1}((a,\infty))$. Then for any $x \in U_a$, by continuity there exists a corresponding δ_x such that $|y - x| < \delta_x \implies |f(y) - f(x)| < \epsilon := |f(x) - a|$, which is just a fancy way of saying that $f(y) > a$ for all $y \in (x - \delta_x, x + \delta_x)$, or in other words that $x \in U_a \implies (x - \delta_x, x + \delta_x) \in U_a$. This tells us that

$$\bigcup_{x \in U_a} (x - \delta_x, x + \delta_x) \subseteq U_a := f^{-1}((a,\infty))$$

For inclusion in the other direction, note that

$$x \in f^{-1}((a,\infty)) \implies x \in (x - \delta_x, x + \delta_x) \implies x \in \left(\bigcup_{x \in U_a} (x - \delta_x, x + \delta_x)\right).$$

We now have that $f^{-1}((a,\infty))$ is a union of open intervals, but we can't immediately conclude that this it is a Borel set, because the union involves an *uncountable* number of intervals. However, the following result shows that any open set can always be written as a countable union of open intervals:

Recall that the definition of an open subset of $\mathbb{R}$ is that for every point x in an arbitrary open subset U, there exists some δ_x such that the neighborhood around the point, $(x - \delta_x, x + \delta_x)$, lies fully within U. That means U can be written as

$$U = \bigcup_{x \in U} (x - \delta_x, x + \delta_x)$$

which you can verify yourself by proving both $\subseteq$ and $\supseteq$. This does not guarantee a *countable* union, however, so we aren't done yet. Let's make the following tweak: find $q_x, r_x \in \mathbb{Q}$ where $x - \delta_x \leq q_x < r_x \leq x + \delta_x$ for every $x \in U$. We see that $(q_x, r_x) \subseteq (x - \delta_x, x + \delta_x) \subseteq U$, so

$$U = \bigcup_{x \in U} (q_x, r_x)$$

(again by $\subseteq$ and $\supseteq$). Furthermore, because $\mathbb{Q}$ is countable, $\mathbb{Q} \times \mathbb{Q}$ is as well: any countable set can be given natural number indices $c_1, c_2, c_3, \ldots$, and so all pairs of elements from that countable set can be written out in a grid like so

$$\begin{array}{cccc} (c_1, c_1) & (c_1, c_2) & (c_1, c_3) & \cdots \\ (c_2, c_1) & (c_2, c_2) & (c_2, c_3) & \cdots \\ (c_3, c_1) & (c_3, c_2) & (c_3, c_3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

This grid of pairs can be given natural number indices using the same diagonal weaving pattern as in the proof that $\mathbb{Q}$ is countable. Thus, all open subsets of the real line can be represented as a countable union of (distinct) open intervals — i.e., *all* uncountable unions of open intervals can be reduced to countable unions.

6 Outer Measure

After all that set up, we now set out to answer the fundamental question of measure theory: how do we assign size to subsets of the real line? First, we define this notion of length for intervals:

$$\ell((a, b)) = b - a, \text{ where } a, b \in \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}, \text{ and } a \leq b$$

and then make the following preliminary intuitive "guess" as to what a measure could be:

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

To clear up some notation, any sequence $\{I_k\}$ where $A \subseteq \bigcup\limits_{k=1}^{\infty} I_k$ is called a covering of A and $\sum\limits_{k=1}^{\infty} \ell(I_k)$ is called the total length of the cover. We call this the Lebesgue outer measure of A. Now, this function obeys several properties:

1. Monotonocity: $A \subseteq B \implies \mu^*(A) \leq \mu^*(B)$ — any covering $\{I_k\}$ that covers B also covers A.

2. Countable sets have zero measure: $\mu^*(\{x_1, x_2, \ldots\}) = 0$ —

   $$\{x_1, x_2, \ldots\} \subseteq \left(x_1 - \frac{\epsilon}{2^2}, x_1 + \frac{\epsilon}{2^2}\right) \cup \left(x_2 - \frac{\epsilon}{2^3}, x_2 + \frac{\epsilon}{2^3}\right) \cup \ldots$$

   which has a total length of ϵ. Thus, the infimum must be 0.

3. Countable subadditivity: $\mu^*\left(\bigcup\limits_{k=1}^{\infty} A_i\right) \leq \sum\limits_{k=1}^{\infty} \mu^*(A_i)$ — let the sequence of intervals $I_{k1}, I_{k2}, \ldots$

cover A_k s.t.

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

Such a covering exists because $\mu^*(A)$ is the infimum. If we do this for all $k \in \mathbb{N}$, then we get

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

But we know that $\bigcup\limits_{k=1}^{\infty} A_i \subseteq \bigcup\limits_{k=1}^{\infty}\bigcup\limits_{i=1}^{\infty} I_{ki}$ 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 ϵ is arbitrary, we conclude the proposition.

4. Approximating from the outside by open sets: $\mu^*(A) = \inf\left\{\mu^*(U) : A \subseteq U, U \text{ open}\right\}$ — we know from the property 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 $\{I_k\}$ which has total length $< \mu^*(A) + \epsilon$. But $U_k := \bigcup_{k=1}^{\infty} I_k$ is an open set that contains A, and that is covered by $\{I_k\}$, which implies that $\mu^*(U_k) < \mu^*(A) + \epsilon$! And thus, the infimum value over all U must be $\leq \mu^*(A) + \epsilon$, and because ϵ is arbitrary, 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 α; 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: e.g., if I_k is not bounded, w.l.o.g. $I_k = (a, \infty)$, then we can replace I_k with

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

We see that total length of the covering is ∞ 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:

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

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

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

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 $< \sum_{k=1}^{\infty} \ell(I_k) + \epsilon$. And because ϵ is arbitrary, the infimum total length of covers with arbitrary open intervals is ALSO the infimum total length of covers with only small open intervals.

Explained more clearly, this is because 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 $\{I_k\}$ 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! With this in hand, we can get two more properties:

5. Additivity for distant sets: $d(A, B) > 0 \implies \mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ — 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 ϵ is arbitrary, this means that $\mu^*(A) + \mu^*(B) \leq \mu^*(A \cup B)$. The other direction is obvious by countable subadditivity.

7 Measures

Onward to the titular protagonist of measure theory: a measure on a *measurable space* $(\Omega, \mathcal{S})$ is function $\mu : \mathcal{S} \to [0, \infty]$ such that

1. $\mu(\varnothing) = 0$
2. Countable Additivity: if a countably infinite family of pairwise disjoint sets $\{E_1, E_2, \ldots\}$, then

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

(If you think this looks an awful lot like probability measures, yes). I think I mentioned this before but (countable additivity) implies (finite additivity). Any function satisfying the above definition is a measure, and we can define a (measure space) as the triple $(\Omega, \mathcal{S}, \mu)$ for some Ω, σ-algebra, and measure. Check out the examples from Axler's book (section 2C) — there's not much I can add.

7.1 Lebesgue measure

And the moment we've all been waiting for: the famous Lebesgue measure. Define the length of an open interval, $\ell((a,b))$, to be $b-a$ (for a or b equal to ∞ we get $\ell = \infty$, and if $a = b$ then we have $\varnothing$ whose length is of course 0). All super trivial stuff. Now define the *outer measure* of any set to be

$$|A| = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) \text{ where the } I_k \text{ are open intervals such that } A \subset \bigcup_{k=1}^{\infty} I_k \right\}$$

This makes a lot of intuitive sense — find the "smallest" covering of your set using open intervals and find the length of that covering. We will later see that this *is* the Lebesgue measure, the only difference being that outer measure applies to all subsets of $\mathbb{R}$ whereas Lebesgue measure only applies to Lebesgue measurable sets (a concept I'll get into more later).