Math/CS 120: Intro. to Math — Professor: Padraic Bartlett

Lecture 5: Measure, Integrals and Limits

Week 6 — *UCSB 2014*

> O! it is excellent
> To have a giant's strength; but it is tyrannous
> To use it like a giant.
>
> Isabella, scene II, Measure for Measure

1 Measure

1.1 Definitions.

Loosely speaking, a "measure" on some space X is simply a way to assign a notion of "length" to some of the subsets of that space. For example, you already know how to assign a notion of length to the interval subsets of $\mathbb{R}$: you just set the length of $[a, b]$ to be $b - a$, for any interval $[a, b]$.

But how can we make this formal? In other words: how can we talk about lengths for **lots** of things, rather than just the "nice" ones that we know? Can we create a notion of length that is "smart" enough to know that the length of $[0, 1/2] \cup [1/2, 3/4] \cup [3/4, 7/8] \cup \ldots$ should be 1?

The answer is yes! In this set of notes, we set about doing just this.

Formally, we say that a measure is the following sort of object:

Definition. A **measure** on $\mathbb{R}$ is a function μ from some collection of subsets of $\mathbb{R}$ to $\mathbb{R} \cup \{\infty\}$, such that μ satisfies the following properties:

1. $\mu(\emptyset) = 0$.
2. For any measurable set A, $\mu(A) \geq 0$.
3. If X, Y are a pair of measurable sets, and $X \subseteq Y$, we have $\mu(X) \leq \mu(Y)$.
4. For any countable collection of disjoint measurable sets $\{A_i\}_{i=1}^{\infty}$, the set $\bigcup_{i=1}^{\infty} A_i$ is also measurable. Moreover, we have the following equality:

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

This definition, however, is missing some of the aspects we really like about measures on $\mathbb{R}$. For example, we could take the "zero" measure, that assigns every set measure 0 — this satisfies our axioms, but isn't really a property that want to think length satisfies! So we have two more properties that we'd like measures on the real line $\mathbb{R}$ to satisfy:

Definition. A measure μ on the real line is called a **Lebesgue measure** if it satisfies the following additional two properties:

5. $\mu([0,1]) = 1$.

6. μ is **scale and translation invariant**: in other words, given any measurable set A and real number t, the set $t + A = \{t + x \mid x \in A\}$ is measurable and has the same measure as A, while the set $tA = \{t \cdot x \mid x \in A\}$ is also measurable and has measure $|t| \cdot \mu(A)$.

Throughout this talk, we will assume that "measure" means Lebesgue measure from here on out, and never refer to measures that are not Lebesgue unless explicitly stated as such.

1.2 Properties; basic cases.

Let's assume we have some such measure μ. What can we say about it?

Well: one plausible goal might be to verify some of the things that we know length "ought" to satisfy. For example, we could ask about the length of an arbitrary interval $[a, b]$:

Proposition. For any $a < b$, $a, b \in \mathbb{R}$, we have $\mu([a, b]) = b - a$.

Proof. First, take the interval $[0, 1]$. By scaling, the measure of $[0, b - a]$ is just $b - a$ times the measure of $[0, 1]$; therefore we know that $\mu([0, b - a]) = (b - a) \cdot 1$.

Now, translate the interval $[0, b-a]$ by a. This doesn't change the length of our interval by our properties; consequently, we have that $\mu([a, b-a+a]) = \mu([a, b]) = b-a$, as claimed. □

We could also ask about the length of a single point:

Proposition. For any $a \in \mathbb{R}$, we have $\mu(\{a\}) = 0$.

Proof. We present three proofs, to give a feel for how measures work. On one hand, the "fast" proof is simply to scale $[0, 1]$ by 0; this tells us that the measure of the set $[0, 0] = \{0\}$ is 0, and therefore by scaling that the measure of any single-point set is 0.

On the other hand, we could also prove this fact by contradiction: assume not, that there is some single point $a \in \mathbb{R}$ such that $\mu(\{a\}) \neq 0$. Because measures only output positive real numbers or possible ∞, we can actually say that there is some $\epsilon > 0$ with $\mu(\{a\}) = \epsilon$. (In practice, this strategy of "this thing is nonzero: let's give a name to that nonzero quantity" is something you should do on autopilot in your proofs! Always give objects names!)

But, by translation, this would mean that the measure of **any** single-point set is ϵ! In particular, this would force the measure of any set containing infinitely many points — say, $[0, 1]$ — to be infinity, as we would have

$$\mu([0,1]) \geq \mu\left(\bigcup_{n=1}^{\infty} \left\{\frac{1}{n}\right\}\right) = \sum_{n=1}^{\infty} \mu\left(\left\{\frac{1}{n}\right\}\right) = \sum_{n=1}^{\infty} \epsilon = \infty.$$

But this is impossible, as $\mu([0,1]) = 1$! So we have a contradiction.

Finally, we can present a third proof, by "covering." Suppose that we want to show that some set S has measure zero. To do this, we can simply show that it is "covered" by arbitrarily small sets! In other words: suppose that for any $\epsilon > 0$, we could find a set T with measure less than ϵ, such that $S \subseteq T$. Then, because this implies that $\mu(S) \leq \mu(T)$, we would have shown that $\mu(S) \leq \epsilon$, for any $\epsilon > 0$. But the only nonnegative number less than all positive numbers is 0; therefore $\mu(S) = 0$!

To do this here: take any a, and notice that for any $\epsilon > 0$, the interval $[a-\epsilon/2, a+\epsilon/2]$ is a set of measure ϵ (by our first proposition) that contains $\{a\}$. Therefore, for any $\epsilon > 0$ we have that $\mu(\{a\}) \leq \mu([a-\epsilon/2, a+\epsilon/2]) = \epsilon$; consequently $\mu(\{a\})$ must equal 0, as claimed. □

From here, you can make observations like the following:

Proposition. For any $a < b$, $a, b \in \mathbb{R}$, we have $\mu((a,b)) = b - a$.

Proof. Simply notice that

$$b - a = \mu([a,b]) = \mu(\{a\} \cup (a,b) \cup \{b\}) = \mu(\{a\}) + \mu((a,b)) + \mu(\{b\}) = 0 + \mu((a,b)) + 0,$$

and therefore that our claim is true. □

From here, we can start to study stranger sets. Recall the Cantor set, from past problem sets:

Definition. The **Cantor set** $\mathcal{C}_\infty$ is defined as follows:

- Start with the interval $[0,1]$. Call this set C_0.
- Remove the middle-third of this set, so that you have $[0,1/3]$ and $[2/3,1]$ left over. Call this set C_1.
- Remove the middle-third of those two sets, so that you have $[0,1/9], [2/9,1/3], [2/3,7/9], [8/9,1]$ left over. Call this set C_2.
- Repeat this process! In other words, to form C_n, take all of the closed intervals that make up C_{n-1}, and remove the middle-third from each of these intervals. The resulting collection of closed intervals is C_n.

Define $\mathcal{C}_\infty$, the Cantor set, as the set made by taking all of the elements x such that x is in C_i, for every i. In other words, set

$$\mathcal{C}_\infty = \bigcap_{n=1}^{\infty} C_n.$$

Proposition. The measure of the Cantor set $\mathcal{C}_\infty$ is 0.

Proof. Consider the construction of the Cantor set above. In particular, let's define the "leftover sets" L_n from the Cantor construction as follows:

- $L_1 = (1/3, 2/3)$: in other words, L_1 consists of the middle-third that we removed from C_0 to make C_1.
- $L_2 = (1/9, 2/9) \cup (7/9, 8/9)$. In other words, L_2 is the collection of all of the open intervals we deleted from C_1 to make C_2.
- In general, $L_n = C_{n-1} \setminus C_n$.

Now, make the following observations about how the sets C_n, L_n grow and interact:

- C_n is a union of 2^n distinct closed intervals. To see this, simply apply induction. Our base case, C_0 consists of $2^0 = 1$ interval. Finally, if we inductively assume that C_n is made out of 2^n intervals, when we delete the middle third out of each of those intervals to make C_{n+1}, we will have at the end 2^{n+1} many intervals (as we split each of our 2^n into two pieces.)
- Consequently, we have that L_n consists of a union of 2^{n-1} intervals, as there are as many pieces in L_n as there were middle-thirds deleted from C_{n-1}.
- As well, the length of each interval in C_n is $\frac{1}{3^n}$. This is again justifiable by induction; our base case has length $0 = 1/3^0$, while if C_n's intervals all have length $1/3^n$, chopping out their middle-thirds will leave us with intervals of length $(1/3)(1/3^n)$ for C_{n+1}, as claimed.
- Consequently, L_n's intervals all have length $1/3^n$, as they correspond to precisely the deleted middle-thirds that transform C_{n-1} into C_n.
- Finally, all of the L_n's are disjoint from each other and from $\mathcal{C}_\infty$, as the L_n's correspond to deleted middle-thirds from the Cantor set (check this if you don't see why it is true!)
- As a result, we know the measure of L_n for each n: because L_n is made of 2^{n-1} disjoint open intervals of length $1/3^n$, we can conclude that $\mu(L_n) = \frac{2^{n-1}}{3^n}$ by our earlier results and properties.

What do these observations tell us? Well: if we write

$$[0,1] = \mathcal{C}_\infty \cup \left(\bigcup_{n=1}^{\infty} L_n \right),$$

and use the fact that C_n and all of the L_n's are disjoint, we have

$$\begin{aligned} 1 = \mu([0,1]) = \mu\left(\mathcal{C}_\infty \cup \left(\bigcup_{n=1}^{\infty} L_n \right) \right) &= \mu(\mathcal{C}_\infty) + \sum_{n=1}^{\infty} \mu(L_n). \\ &= \mu(\mathcal{C}_\infty) + \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n}. \\ &= \mu(\mathcal{C}_\infty) + \frac{1/3}{1-(2/3)} = \mu(\mathcal{C}_\infty) + 1. \end{aligned}$$

Consequently, $\mu(\mathcal{C}_\infty) = 0$, as claimed. ☐

With somewhat less work, we can study a few stranger sets:

Proposition. For any countable set A, $\mu(A) = 0$.

Proof. Because A is countable, there is some bijection between A and the natural numbers! Therefore, we can use this bijection to "label" the elements of A: that is, if $\phi : \mathbb{N} \to A$ is a bijection, simply label each element of a with the element of $\mathbb{N}$ that maps to it!

Doing this lets us write $A = \{a_n\}_{n=0}^{\infty}$, as we have "labeled" all of our elements in A. This is a tremendously useful thing to do to A, and is often the first/only thing we will do with a set when we know that it is countable! In later proofs, we will skip this discussion, and simply say "Write $A = \{a_n\}_{n=1}^{\infty} \ldots$" to start our proofs; but it's a useful trick!

With this done: simply notice that

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} \{a_n\}\right) = \sum_{n=1}^{\infty} \mu(\{a_n\}) = \sum_{n=1}^{\infty} 0 = 0.$$

Therefore, $\mu(A) = 0$! □

This has some strange corollaries:

Corollary. $\mu(\mathbb{Q}) = 0$.

Corollary. $\mu([0,1] \setminus \mathbb{Q}) = 1$.

Think about these two results for a while; try to persuade yourself why they are true, and also try to think about why they are surprising!

From here, you might wonder how much further you can keep going: i.e. how many sets can we define our measure on? Can you make **every** set measurable?

Surprisingly enough, the answer is no! Within the axioms that we work with in mathematics, we can construct sets that cannot have any notion of length in any Lebesgue measure. We give a proof of this here:

Theorem. There is a nonmeasurable set.

Proof. Consider the following construction. Take the interval $[0, 1]$. Define an equivalence relation on $[0, 1]$ as follows: set $x \sim y$ iff $x - y \in \mathbb{Q}$. (Check that this is an equivalence relation!)

For example, the collection of all elements equivalent to 0 here is just the set

$$\{x \in \mathbb{R} \mid x - 0 \in \mathbb{Q}\};$$

in other words, the equivalence class for 0 is just $\mathbb{Q}$! Similarly, the equivalence class for π is just

$$\{x \in \mathbb{R} \mid x - \pi \in \mathbb{Q}\} = \{\pi + q \mid q \in \mathbb{Q}\} = \mathbb{Q} + \pi,$$

and our other equivalence classes look similar.

Using this equivalence relation, we can partition $[0, 1]$ into its equivalence classes under this relation! Denote these equivalence relations $\{E_i\}_{i \in I}$, for some indexing set I. (This trick, of labeling things with some index set I without saying what I is, is an incredibly useful one! For example, right now, we don't know what I is: it would be a mistake to assume that I is countable, or perhaps the size of $\mathcal{P}(\mathbb{R})$, without knowing more. But we can certainly assert that there is some way to collect all of our equivalence relations together; once we do this, we can simply let I be any way to label all of these sets!)

Now, take each of these sets E_i, and pick out[1] some element $e_i \in E_i$ from each of them. Furthermore, when we do this, pick out e_i so that it lives in the interval $[0, 1]$. This is not hard to do: if we have any element $e_i \in E_i$, then any translation $e_i + q$ is also in E_i. Therefore we can just shift our element into $[0, 1]$ if it wasn't already there!

Let $A+q$ denote the translation of A by some rational q. Notice that for any $q, r \in \mathbb{Q}$, if $A+q$ and $A+r$ have an element in common, we would have some $e_i + q = e_j + r$. But this means that $e_i \sim e_j$; consequently, because there is only one member from any equivalence class in A, this tells us that $e_i = e_j$, and therefore that $q = r$! In other words, if $q \neq r$, $A+q \neq A+r$.

Conversely; for any $x \in \mathbb{R}$, there is some $q \in \mathbb{Q}$ such that $x \in A+q$. To see this, simply find the equivalence class E_i of $\mathbb{R}$ that contains x, and look up the representative e_i we chose for E_i. We know by definition that $x \sim e_i$; consequently, there is some q such that $x = e_i + q$, and therefore that $x \in A + q$.

Therefore, the collection of sets $\{A+q \mid q \in \mathbb{Q}\}$ partition the real numbers! That's... strange.

How strange? Well: consider the union $\mathcal{A}$ of these sets given by all translations of A by a rational number in $[-1, 1]$,

$$\mathcal{A} = \bigcup_{q \in [-1,1] \cap \mathbb{Q}} (A+q)$$

On one hand: for any $x \in [0, 1]$, if x is in the equivalence class E_i, we can see that

$$x \sim e_i \Rightarrow x = e_i + q, \text{ for some } q \Rightarrow q \in [-1, 1],$$

---

[1] The particularly astute among you might notice that this step — unique among the steps we've done throughout this class since we started axiomatizing things in week 2 to build the natural numbers! — is, strictly speaking, not something that our axioms let us do. We can make power sets and unions and formulas and all sorts of stuff, but there is no axiom that lets us simply pick out elements from other sets without defining a rule! So, let's fix that here:

The Axiom of Choice : For every family Φ of nonempty sets, there is a choice function

$$f : \Phi \to \bigcup_{S \in \Phi} S,$$

such that $f(S) \in S$ for every $S \in \Phi$.

When this was first proposed as an axiom, mathematicians were opposed to it on several grounds:

- Constructivist and intutionist mathematicians opposed it, on the grounds that it posits the existence of functions without any clue whatsoever as to how to find them!
- Many other working mathematicians just thought it was a true statement; i.e. that AC was a trivial consequence of any logical framework of mathematics.

Surprisingly enough, however, Paul Cohen and Kurt Gödel proved that the axiom of choice is independent of the Zermelo-Fraenkel axioms of set theory, the current framework within which we do mathematics: i.e. that it is its own proper axiom! Pretty much all of modern mathematics accepts the Axiom of Choice; it's a pretty phenomenally useful axiom, and most fields of mathematics like to be able to call on it when pursuing nonconstructive proofs. We'll talk lots more about it later.

because both $x, e_i \in [0,1]$, and therefore their difference is at least -1 and at most 1.

On the other hand: no element in $\mathcal{A}$ is greater than 2 (as elements in A are all ≤ 1) nor smaller than -1 (as elements in A are all nonnegative.)

Therefore, we have

$$(-1,2) \supseteq \bigcup_{q \in [-1,1] \cap \mathbb{Q}} (A+q) \supseteq [0,1],$$

and therefore that

$$3 = \mu((-1,2)) \geq \mu\left(\bigcup_{q \in [-1,1] \cap \mathbb{Q}} (A+q) \right) \geq \mu([0,1]) = 1$$

But all of the $A+q$'s are just translations of each other. Therefore they all have the same measure, $\mu(A)$: so we have

$$3 \geq \sum_{q \in [-1,1] \cap \mathbb{Q}} \mu(A) \geq 1.$$

So: what is $\mu(A)$? We can clearly see that it is nonzero, because otherwise we would not satisfy our rightmost inequality, that $\sum_{q \in [-1,1] \cap \mathbb{Q}} \mu(A) \geq 1$. But we cannot have that it is equal to any $\epsilon > 0$, because this would break our leftmost inequality $3 \geq \sum_{q \in [-1,1] \cap \mathbb{Q}} \mu(A)$. But any nonnegative number is either 0 or greater than 0; consequently, we have proven that there is **no** value we can assign to $\mu(A)$ without creating a contradiction!

In other words: for any measure μ that we make, there are sets (like A!) that **cannot** be measured by that μ! In other words; there are nonmeasurable sets. □

Strange!

Despite this oddity, though, measures are still pretty ridiculously useful. One of the best applications of the concept of measure is to **integration,** which we discuss in the next section:

2 Integration

2.1 The old-school approach.

Way back in calculus, you probably saw the integral of a function defined as "the area under the curve." As a rigorous definition, this seems lacking — what is this area, and how are we calculating it? — but it is not a bad place to start. For example, this definition works just fine at telling us what the integral of certain very simple functions should be:

Definition. Take any interval $[a,b)$, for $a < b \in \mathbb{R}$. We define the **characteristic function** $\chi_{[a,b)}(x)$ of this interval as follows:

$$\chi_{[a,b)}(x) = \begin{cases} 1, & x \in [a,b), \\ 0, & x \notin [a,b). \end{cases}$$

Definition. For any interval $[a, b)$ and corresponding characteristic function $\chi_{[a,b)}$, we define the **integral** of $\chi_{[a,b)}(x)$ over $\mathbb{R}$ as $b - a$, and write

$$\int_{\mathbb{R}} \chi_{[a,b)}(x)dx = b - a.$$

Notice that this is the area between our function and the x-axis!

We can extend this definition in a few logical ways:

Definition. For any interval $[a, b)$ and corresponding characteristic function $\chi_{[a,b)}$, and any constant c, we set the **integral** of $c \cdot \chi_{[a,b)}(x)$ over $\mathbb{R}$ as $c \cdot (b - a)$, and write

$$\int_{\mathbb{R}} c\chi_{[a,b)}(x)dx = c(b - a).$$

This, too, is the area between our function and the x-axis, if we interpret area "beneath" the x-axis as "negative."

Similarly, for any disjoint intervals $[a_1, b_1), [a_2, b_2), \ldots [a_n, b_n)$, corresponding characteristic functions $\chi_{[a_1,b_1)}(x), \ldots \chi_{[a_n,b_n)}(x)$, and any constants $c_1, \ldots c_n$, we set the **integral** of $\sum_{k=1}^{n} c_k \cdot \chi_{[a_k,b_k)}(x)$ over $\mathbb{R}$ to be the sum $\sum_{k=1}^{n} c_k(b_k - a_k)$: that is,

$$\int_{\mathbb{R}} \left(\sum_{k=1}^{n} c_k \cdot \chi_{[a_k,b_k)}(x) \right) dx = \sum_{k=1}^{n} c_k(b_k - a_k).$$

This, as well, is still the area between our function and the x-axis, because the sum of our individual areas is just the area of our sum!

The sorts of functions that we could integrate above are particularly useful, and merit a definition of their own:

Definition. Take any collection of intervals $[a_1, b_1), [a_2, b_2), \ldots [a_n, b_n)$, corresponding characteristic functions $\chi_{[a_1,b_1)}(x), \ldots \chi_{[a_n,b_n)}(x)$, and any constants $c_1, \ldots c_n$.

The **step function** associated to these objects is the map

$$\sum_{k=1}^{n} c_k \cdot \chi_{[a_k,b_k)}(x).$$

Notice that if any two scaled characteristic functions $\chi_{[a_1,b_1)}(x), \chi_{[a_2,b_2)}(x)$ overlap, then we can actually write their sum as the sum of three scaled non-overlapping characteristic functions! For example, we can write

$$3\chi_{[1,4)}(x) + 5\chi_{[2,7)}(x) = 3\chi_{[1,2)}(x) + 8\chi_{[2,4)}(x) + 5\chi_{[4,7)}(x).$$

So, when we work with the concepts above, we can actually omit the "disjoint" condition above, as it doesn't change our results; this will simplify[2] our work.

So: how can we integrate other functions? Well: the central idea that you probably saw back in calculus was that if you wanted to find the area beneath **any** curve, you could just **approximate** that curve with rectangles, and take the area beneath those rectangles! Somewhat more formally, this is the following definition:

Definition. Take any function $f : \mathbb{R} \to \mathbb{R}$, and any interval $[a, b]$. We say that f is **integrable** on the interval $[a, b]$ if the following equation holds:

$$\inf_{\substack{\varphi(x) \geq f(x) \text{ on } [a,b], \\ \varphi(x) \text{ a step function,} \\ \varphi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \varphi(x)dx = \sup_{\substack{\psi(x) \leq f(x) \text{ on } [a,b], \\ \psi(x) \text{ a step function,} \\ \psi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \psi(x)dx.$$

If this does happen, then we say that the integral $\int_a^b f(x)dx$ is equal to both the left- and right-hand-sides above.

There are a lot of undefined concepts here! Let's back up and define some of them:

- Given any set $A \subseteq \mathbb{R}$, we define the **infimum** of that set as the largest element $I \in \mathbb{R}$ such that $I \leq a, \forall a \in A$. This, in a sense, is describing the "best possible" lower bound for A. We denote this by "inf" for shorthand. For example, in the above definition, the left-hand-side is the infimum of the set

  $$\left\{ \int_{\mathbb{R}} \varphi(x)dx \;\middle|\; \varphi(x) \geq f(x) \text{ on } [a, b], \varphi(x) = 0 \text{ off of } [a, b], \varphi(x) \text{ a step function.} \right\}$$

  To offer another example: the infimum of the set $(0, 1)$ is 0, because

  - $0 \leq x$, for all $x \in (0, 1)$; that is, 0 is a lower bound!
  - Moreover, 0 is the largest possible lower bound. To see this, take any any other potential lower bound $\epsilon > 0$. We can trivially observe that $\epsilon < 1$, as otherwise ϵ would fail to be a lower bound for **any** of the elements in our set!

    But then we have that $\epsilon \not\leq \epsilon/2$, while $\epsilon/2 \in (0, 1)$ (because $\epsilon < 1$.) So no value greater than 0 can be a lower bound!

[2] Relatedly, this is why we're working with intervals of the form $[a, b)$ instead of the more standard closed $[a, b]$ or open (a, b) intervals. Tntervals of this form have the nice property that when you stack/intersect/union them, the resulting intervals all still have the same form — this simplifies some of our arguments! You could of course do everything here with other sorts of intervals, and it would all work; you'd just have weird things where equations like the above would not hold at single points of overlap or stuff like that.

Checking these two properties is how we show that something is an infimum.

- Similarly, given any set $A \subseteq \mathbb{R}$, we define the **supremum** of that set as the smallest element $S \in \mathbb{R}$ such that $S \geq a, \forall a \in A$. This, in a sense, is describing the "best possible" upper bound for A. We denote this by "sup" for shorthand. Again, for example, the right-hand-side above is the supremum of the set

$$\left\{ \int_{\mathbb{R}} \psi(x)dx \;\middle|\; \psi(x) \leq f(x) \text{ on } [a,b], \psi(x) = 0 \text{ off of } [a,b], \psi(x) \text{ a step function.} \right\}$$

To offer another example: the supremum of the set $[3,7]$ is 7, because

  - $7 \geq x$, for all $x \in [3,7]$; that is, 7 is a upper bound!
  - Moreover, 7 is the smallest possible upper bound. Take any other possible upprt bound that is less than 7; we can describe this possible bound as $7 - \epsilon$, for some positive ϵ. Notice that $\epsilon \leq 4$, as otherwise our upper bound would not be less than 3 (and thus in particular not less than any element in our set!

    But then we have that $7 - \epsilon \not\leq 7 - (\epsilon/2)$, while $7 - \epsilon/2$ is an element in $[3,7]$ (because $\epsilon < 4$.) So no value less than 7 can be a upper bound!

Checking these two properties is how we show that something is an infimum.

Ok. With that done, let's check back in with our definition and make sure that it's actually describing what we think an integral should be!

On one hand, we can interpret the left-hand-side of

$$\inf_{\substack{\varphi(x) \geq f(x) \text{ on } [a,b], \\ \varphi(x) \text{ a step function}, \\ \varphi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \varphi(x)dx = \sup_{\substack{\psi(x) \leq f(x) \text{ on } [a,b], \\ \psi(x) \text{ a step function}, \\ \psi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \psi(x)dx.$$

as the "smallest set of rectangles that are greater than $f(x)$ on $[a,b]$." On the other, we can interpret the right-hand-side as the "largest set of rectangles that are smaller than $f(x)$ on $[a,b]$."

If our function has a well-defined area beneath its curve, then these two methods of measuring area — creating the best upper bound, versus making the best lower bound — should agree! Therefore, we are simply saying that a function has well-defined area beneath its curve if and only if these two methods of measuring agree with each other.

To illustrate how this definition works, consider the following example:

Theorem. The integral of x^2 on $[0,1]$ is $\frac{1}{3}$.

Proof. To do this, a natural first step is to try to find some likely-looking step functions that seem to be going to the same values. How can we do this?

Well: one trick that will usually work is to simply divide the interval $[0,1]$ into n equal parts – i.e. to consider the division $[0,1/n),[1/n,2/n),\ldots[(n-1)/n,1]$. Under this division, the best upper-bound function $\varphi(x)$ we can make is

$$\varphi(x)=\begin{cases} 1/n^2, & x\in[0,1/n),\\ 4/n^2, & x\in[1/n,2/n),\\ 9/n^2, & x\in[2/n,3/n),\\ & \vdots \\ 1, & x\in[(n-1/n,1]. \end{cases}$$

Similarly, the best lower-bound function $\psi(x)$ we can have is

$$\varphi(x)=\begin{cases} 0, & x\in[0,1/n),\\ 1/n^2, & x\in[1/n,2/n),\\ 4/n^2, & x\in[2/n,3/n),\\ & \vdots \\ (n-1)^2/n^2, & x\in[(n-1/n,1]. \end{cases}$$

We can integrate these two functions, and get

$$\begin{aligned}\int\varphi(x)dx &= \sum_{k=1}^{n}\frac{k^2}{n^2}\cdot\text{length}([(k-1)/n,(k/n))\\ &= \sum_{k=1}^{n}\frac{k^2}{n^3}\\ &= \frac{\sum_{k=1}^{n}k^2}{n^3}\\ &= \frac{\frac{n(n+1)(2n+1)}{6}}{n^3}\\ &= \frac{2n^3+3n^2+1}{6n^3}\\ &= \frac{1}{3}+\frac{1}{2n}+\frac{1}{6n^2},\end{aligned}$$

while

$$\begin{aligned}
\int \psi(x)dx &= \sum_{k=1}^{n} \frac{(k-1)^2}{n^2} \cdot \text{length}([(k-1)/n, (k/n)) \\
&= \sum_{k=1}^{n} \frac{(k-1)^2}{n^3} \\
&= \frac{\sum_{k=0}^{n-1} k^2}{n^3} \\
&= \frac{\frac{(n-1)n(2n-1)}{6}}{n^3} \\
&= \frac{2n^3 - 3n^2 + 1}{6n^3} \\
&= \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}.
\end{aligned}$$

(In both cases, we used the identity $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, which is easily proved by induction if you haven't seen it!)

What does this mean? Well: we know that on one hand, for any n we have

$$\inf_{\substack{\varphi(x) \geq f(x) \text{ on } [a,b], \\ \varphi(x) \text{ a step function}, \\ \varphi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \varphi(x)dx \leq \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2},$$

because the smallest value must be at least as small as these values for φ we've just found. Similarly, we have that

$$\sup_{\substack{\psi(x) \leq f(x) \text{ on } [a,b], \\ \psi(x) \text{ a step function}, \\ \psi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \psi(x)dx \geq \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2},$$

because the largest value must be at least as large as these values for ψ we've just found.

But we also know that

$$\inf_{\substack{\varphi(x) \geq f(x) \text{ on } [a,b], \\ \varphi(x) \text{ a step function}, \\ \varphi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \varphi(x)dx \geq \sup_{\substack{\psi(x) \leq f(x) \text{ on } [a,b], \\ \psi(x) \text{ a step function}, \\ \psi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \psi(x)dx,$$

because for any $\varphi(x) \geq f(x)$ and $\psi(x) \leq f(x)$ on $[a, b]$, we have $\varphi(x) \geq \psi(x)$ for all x, and therefore that $\int \varphi(x) \geq \int \psi(x)$ must hold!

But, if we combine all of our observations, we have that for all n,

$$\frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \geq \inf_{\substack{\varphi(x) \geq f(x) \text{ on } [a,b], \\ \varphi(x) \text{ a step function}, \\ \varphi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \varphi(x)dx \geq \sup_{\substack{\psi(x) \leq f(x) \text{ on } [a,b], \\ \psi(x) \text{ a step function}, \\ \psi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \psi(x)dx \geq \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}.$$

For large values of n, the far-left-hand and far-right-hand sides go to $1/3$; therefore the middle two terms must be equal to $1/3$! Therefore our function x^2 is integrable, and its integral is $1/3$, as claimed. $\square$

So we can integrate things! However, we cannot integrate all things with our current definitions:

Theorem. The function

$$f(x) = \begin{cases} 1, & x \in [0,1] \cap \mathbb{Q}, \\ 0, & \text{otherwise} \end{cases}$$

is not integrable on $[0,1]$.

Proof. To see why, consider any step function $\varphi(x)$ such that $\varphi(x) \geq f(x)$ on $(0,1)$. Notice that it is impossible for $\varphi(x)$ to be anything less than 1 on this interval: if $\varphi(x) < 1$ at some $x \in (0,1)$, then, because φ is a sum of a finite number of disjoint scaled copies of characteristic functions of intervals, there is some entire interval on which φ is less than 1. But any interval contains rational points! Therefore, $\varphi(x)$ is not greater than $f(x)$ on this interval, which is a contradiction.

Similarly, notice that if $\psi(x)$ is any step function such that $\psi(x) \leq f(x)$ on $(0,1)$, then $\psi(x)$ must be no greater than 0 on this interval (by the same logic!)

But this means that

$$\inf_{\substack{\varphi(x) \geq f(x) \text{ on } [a,b], \\ \varphi(x) \text{ a step function}, \\ \varphi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \varphi(x)dx \geq 1, \text{ while}$$

$$\sup_{\substack{\psi(x) \leq f(x) \text{ on } [a,b], \\ \psi(x) \text{ a step function}, \\ \psi(x)=0 \text{ off of [a,b]}}} \int_{\mathbb{R}} \psi(x)dx \leq 0.$$

Consequently, these two quantities cannot be equal! Therefore our function is not integrable. □

This is ... bothersome. Luckily, we can fix this with our concept of **measure!**

2.2 Measures and integration.

Consider the following alternate definition for integrals:

Definition. Take any subset S of $\mathbb{R}$. We define the **characteristic function** $\chi_S(x)$ of this set as follows:

$$\chi_s(x) = \begin{cases} 1, & x \in S, \\ 0, & x \notin S \end{cases}$$

Definition. For any measurable subset $S \subseteq \mathbb{R}$ and corresponding characteristic function χ_s, we define the **Lebesgue integral** of $\chi_s(x)$ over $\mathbb{R}$ as $\mu(S)$, and write

$$\int_{\mathbb{R}} \chi_s(x)dx = \mu(S).$$

Notice that this is the "area" between our function and the x-axis, if we interpret area as "length" times height, and stick with our intuitive idea of the measure of a set being its length!

We extend this to combinations of such characteristic functions in the same way as before:

Definition. Take any collection of sets $S_1, \ldots S_n$ and corresponding characteristic functions $\chi_{s_1}(x), \ldots \chi_{s_n}(x)$, and any constants $c_1, \ldots c_n$.

The **simple function** associated to these objects is the map

$$\sum_{k=1}^{n} c_k \cdot \chi_{s_n}(x).$$

We define the Lebesgue integral of any such simple function in the same way as before:

$$\int_{\mathbb{R}} \left(\sum_{k=1}^{n} c_k \cdot \chi_{s_n}(x) \right) dx = \sum_{k=1}^{n} c_k \cdot \mu(S_k).$$

Definition. Take any function $f : \mathbb{R} \to \mathbb{R}$, and any measurable set S. We say that f is **Lebesgue-integrable** on the set S if the following equation holds:

$$\inf_{\substack{\varphi(x) \geq f(x) \text{ on } S \\ \varphi(x) \text{ a simple function,} \\ \varphi(x)=0 \text{ off of S}}} \int_{\mathbb{R}} \varphi(x)dx = \sup_{\substack{\psi(x) \leq f(x) \text{ on } S, \\ \psi(x) \text{ a simple function,} \\ \psi(x)=0 \text{ off of S}}} \int_{\mathbb{R}} \psi(x)dx.$$

If this does happen, then we say that the integral $\int_S f(x)dx$ is equal to both the left- and right-hand-sides above.

In other words: this is just our definition from before, except with measures used to calculate lengths instead of simply using intervals! As promised, this will let us integrate new functions:

Theorem. The function

$$f(x) = \begin{cases} 1, & x \in [0,1] \cap \mathbb{Q}, \\ 0, & \text{otherwise} \end{cases}$$

is Lebesgue-integrable on $[0, 1]$, and has integral 0.

Proof. To see this, simply notice that $f(x)$ is just the characteristic function of the set $\mathbb{Q} \cap [0, 1]$. We could be done right here, as we know how to integrate characteristic functions; but let's pretend we forgot this for a second.

Then, we can pick $\varphi(x) = f(x) = \psi(x)$ as valid simple functions such that $\varphi(x) \geq f(x) \geq \psi(x)$, and notice that because

$$\int_{\mathbb{R}} \varphi(x)dx = \mu(\mathbb{Q} \cap [0,1]) = 0 = \int_{\mathbb{R}} \psi(x)dx,$$

we have

$$0 \geq \inf_{\substack{\varphi(x) \geq f(x) \text{ on } S \\ \varphi(x) \text{ a simple function,} \\ \varphi(x)=0 \text{ off of S}}} \int_{\mathbb{R}} \varphi(x)dx \geq \sup_{\substack{\psi(x) \leq f(x) \text{ on } S, \\ \psi(x) \text{ a simple function,} \\ \psi(x)=0 \text{ off of S}}} \int_{\mathbb{R}} \psi(x)dx \geq 0$$

and therefore that both sides are equal, and in particular equal to 0. □

Instead of delving further into this concept, we take a sidetrack to one of the more interesting concepts we hit on with our discussion of infimum and supremum earlier: the idea of limits!

3 Limits of Sequences

3.1 Sequences: Definitions

We list the basic definitions here.

Definition. A **sequence** of real numbers is a collection of real numbers $\{a_n\}_{n=1}^{\infty}$ indexed by the natural numbers.

Definition. A sequence $\{a_n\}_{n=1}^{\infty}$ is called **bounded** if there is some value $B \in \mathbb{R}$ such that $|a_n| < B$, for every $n \in \mathbb{N}$. Similarly, we say that a sequence is **bounded above** if there is some value U such that $a_n \leq U, \forall n$, and say that a sequence is **bounded below** if there is some value L such that $a_n \geq L, \forall n$.

Definition. A sequence $\{a_n\}_{n=1}^{\infty}$ is said to be **monotonically increasing** if $a_n \leq a_{n+1}$, for every $n \in \mathbb{N}$; conversely, a sequence is called **monotonically decreasing** if $a_n \geq a_{n+1}$, for every $n \in \mathbb{N}$.

Definition. A sequence $\{a_n\}_{n=1}^{\infty}$ converges to some value λ if the a_n's "go to λ" at infinity. To put it more formally, $\lim_{n\to\infty} a_n = \lambda$ iff for any distance ϵ, there is some cutoff point N such that for any n greater than this cutoff point, a_n must be within ϵ of our limit λ.

In symbols:

$$\lim_{n\to\infty} a_n = \lambda \text{ iff } (\forall\epsilon)(\exists N)(\forall n > N)\ |a_n - \lambda| < \epsilon.$$

Convergence is one of the most useful properties of sequences! If you know that a sequence converges to some value λ, you know, in a sense, where the sequence is "going," and furthermore know where almost all of its values are going to be (specifically, close to λ.)

Because convergence is so useful, we've developed a number of tools for determining where a sequence is converging to. We omit the proofs of these results here; you'll see them when you take a proper analysis class!

3.2 Sequences: Convergence Tools

1. **The definition of convergence**: The simplest way to show that a sequence converges is sometimes just to use the definition of convergence. In other words, you want to show that for any distance ϵ, you can eventually force the a_n's to be within ϵ of our limit, for n sufficiently large.

   How can we do this? One method I'm fond of is the following approach:

   - First, examine the quantity $|a_n - L|$, and try to come up with a very simple upper bound that depends on n and goes to zero. Example bounds we'd love to run into: $1/n, 1/n^2, 1/\log(\log(n))$.

- Using this simple upper bound, given $\epsilon > 0$, determine a value of N such that whenever $n > N$, our simple bound is less than ϵ. This is usually pretty easy: because these simple bounds go to 0 as n gets large, there's always some value of N such that for any $n > N$, these simple bounds are as small as we want.
- Combine the two above results to show that for any ϵ, you can find a cutoff point N such that for any $n > N$, $|a_n - L| < \epsilon$.

2. **The definition of nonconvergence**: Sometimes, you will want to show that a given sequence $\{a_n\}_{n=1}^{\infty}$ **fails** to converge! How can you do this? Well: if we're working with our definition, we would want to prove that

$$\neg \left(\exists L \; \forall \epsilon \; \exists N \; \forall n > N, \; |a_n - \lambda| < \epsilon\right).$$

If you recall from week 1 how to work with negation and quantifiers, you can express this statement equivalently as the following claim:

$$\forall L \; \exists \epsilon \; \forall N \; \exists n > N, \; |a_n - \lambda| \geq \epsilon.$$

In other words, for any limit L we want to find an ϵ such that some of our terms in our sequence are at least ϵ-away from L, no matter how far away we are restricted to look for such terms!

How can we do this? Well: we can actually just mimic how we use the definition!

- Start by taking any L, as this must be our first step.
- First, examine the quantity $|a_n - L|$, and try to come up with a very simple **lower** bound that depends on n but does not go to 0, like say 1. You don't need this bound to hold for all n: you just want it to hold **infinitely often**, that is for any N you want to be able to find a $n > N$ for which this bound will hold.
- Using this simple lower bound, pick ϵ that is smaller than this lower bound!
- Then, we now have that for any L, we have found ϵ such that for all N, we can find a $n > N$ with $|a_n - L| \geq \epsilon$! So we've proven our claim.

3. **Arithmetic and sequences**: These tools let you combine previously-studied results to get new ones. Specifically, we have the following results:

- *Additivity of sequences*: if $\lim_{n\to\infty} a_n, \lim_{n\to\infty} b_n$ both exist, then $\lim_{n\to\infty} a_n + b_n = (\lim_{n\to\infty} a_n) + (\lim_{n\to\infty} b_n)$.
- *Multiplicativity of sequences*: if $\lim_{n\to\infty} a_n, \lim_{n\to\infty} b_n$ both exist, then $\lim_{n\to\infty} a_n b_n = (\lim_{n\to\infty} a_n) \cdot (\lim_{n\to\infty} b_n)$.

4. **Monotone and bounded sequences**: if the sequence $\{a_n\}_{n=1}^{\infty}$ is bounded above and nondecreasing, then it converges; similarly, if it is bounded below and nonincreasing, it also converges. If a sequence is monotone, this is usually the easiest way to prove that your sequence converges, as both monotone and bounded are "easy" properties to work with. One interesting facet of this property is that it can tell you that a sequence converges without necessarily telling you what it converges to! So, it's often of particular use in situations where you just want to show something converges, but don't actually know where it converges *to*.

5. **Squeeze theorem for sequences**: if $\lim_{n\to\infty} a_n, \lim_{n\to\infty} b_n$ both exist and are equal to some value l, and the sequence $\{c_n\}_{n=1}^{\infty}$ is such that $a_n \leq c_n \leq b_n$, for all n, then the limit $\lim_{n\to\infty} c_n$ exists and is also equal to l. This is particularly useful for sequences with things like sin(horrible things) in them, as it allows you to "ignore" bounded bits that aren't changing where the sequence goes.

6. **Cauchy sequences**: We say that a sequence is **Cauchy** if and only if for every $\epsilon > 0$ there is a natural number N such that for every $m > n \geq N$, we have

$$|a_m - a_n| < \epsilon.$$

You can think of this condition as saying that Cauchy sequences "settle down" in the limit – i.e. that if you look at points far along enough on a Cauchy sequence, they all get fairly close to each other.

The Cauchy theorem, in this situation, is the following: a sequence is Cauchy if and only if it converges.

3.3 Sequences: Applications of Convergence Tools

In this section, we work an example for each of these tools. We start by illustrating how to prove a sequence converges using just the definition:

Claim. (Definition of convergence example:)

$$\lim_{n\to\infty} \sqrt{n+1} - \sqrt{n} = 0.$$

Proof. When we discussed the definition as a convergence tool, we talked about a "blueprint" for how to go about proving convergence from the definition: (1) start with $|a_n - L|$, (2) try to find a simple upper bound on this quantity depending on n, and (3) use this simple bound to find for any ϵ a value of N such that whenever $n > N$, we have

$$|a_n - L| < (\text{simple upper bound}) < \epsilon.$$

Let's try this! Specifically, examine the quantity $|\sqrt{n+1} - \sqrt{n} - 0|$:

$$\begin{aligned}
|\sqrt{n+1} - \sqrt{n} - 0| &= \sqrt{n+1} - \sqrt{n} \\
&= \frac{(\sqrt{n+1} - \sqrt{n})(\sqrt{n+1} + \sqrt{n})}{\sqrt{n+1} + \sqrt{n}} \\
&= \frac{n+1-n}{\sqrt{n+1} + \sqrt{n}} \\
&= \frac{1}{\sqrt{n+1} + \sqrt{n}} \\
&< \frac{1}{\sqrt{n}}.
\end{aligned}$$

All we did here was hit our $|a_n - L|$ quantity with a ton of random algebra, and kept trying things until we got something simple. The specifics aren't as important as the idea here: just start with the $|a_n - L|$ bit, and try everything until it's bounded by something simple and small!

In our specific case, we've acquired the upper bound $\frac{1}{\sqrt{n}}$, which looks rather simple: so let's see if we can use it to find a value of N.

Take any $\epsilon < 0$. If we want to make our simple bound $\frac{1}{\sqrt{n}} < \epsilon$, this is equivalent to making $\frac{1}{\epsilon} < \sqrt{n}$, i.e $\frac{1}{\epsilon^2} < n$. So, if we pick $N > \frac{1}{\epsilon^2}$, we know that whenever $n > N$, we have $n > \frac{1}{\epsilon^2}$, and therefore that our simple bound is $< \epsilon$. But this is exactly what we wanted!

In specific, for any $\epsilon > 0$, we've found a N such that for any $n > N$, we have

$$|\sqrt{n+1} - \sqrt{n} - 0| < \frac{1}{\sqrt{n}} < \frac{1}{\sqrt{N}} < \epsilon,$$

which is the definition of convergence. So we've proven that $\lim_{n\to\infty} \sqrt{n+1} - \sqrt{n} = 0$. □

Claim. (Arithmetic and Sequences example:) Does the sequence

$$a_1 = 1,$$
$$a_{n+1} = \sqrt{1 + a_n^2}$$

converge?

Proof. On the homework! □

Claim. (Cauchy sequence example:) The sequence

$$a_n = \sum_{k=1}^{n} \frac{1}{k^2}$$

converges.

Proof. To show that this sequence converges, we will use the Cauchy convergence tool, which tells us that sequences converge if and only if they are Cauchy.

How do we prove that a sequence is Cauchy? As it turns out, we can use a similar blueprint to the methods we used to show that a sequence converges:

- First, examine the quantity $|a_m - a_n|$, and try to come up with a very simple upper bound that depends on m and n and goes to zero. Example bounds we'd love to run into: $\frac{1}{n}, \frac{1}{mn}, \frac{1}{n}, \frac{1}{m^4 \log(n)}$. Things that won't work: $\frac{n}{m}$ (if n is really big compared to m, we're doomed!), $\frac{m}{n^{34}}$ (same!), 4.
- Using this upper bound, given $\epsilon > 0$, determine a value of N such that whenever $m > n > N$, our simple bound is less than ϵ.
- Combine the two above results to show that for any ϵ, you can find a cutoff point N such that for any $m > n > N$, $|a_m - a_n| < \epsilon$.

Let's apply the above blueprint, and study $|a_m - a_n|$. Remember that we're assuming that $m > n$ here:

$$\begin{aligned} |a_m - a_n| &= \left| \sum_{k=1}^{m} \frac{1}{k^2} - \sum_{k=1}^{n} \frac{1}{k^2} \right| \\ &= \sum_{k=n+1}^{m} \frac{1}{k^2} \end{aligned}$$

The following step may seem quite weird: it's motivated by partial fractions (because we want a way to simplify our $\frac{1}{k^2}$'s into simpler things), but it's mostly just an algebraic trick. The important thing is not to remember these tricks, but to just try tons of things until eventually *one* of them sticks:

$$\begin{aligned} \sum_{k=n+1}^{m} \frac{1}{k^2} &< \sum_{k=n+1}^{m} \frac{1}{k(k-1)} \\ &= \sum_{k=n+1}^{m} \left(\frac{1}{k-1} - \frac{1}{k} \right) \\ &= \sum_{k=n+1}^{m} \frac{1}{k-1} - \sum_{k=n+1}^{m} \frac{1}{k} \\ &= \sum_{k=n}^{m-1} \frac{1}{k} - \sum_{k=n+1}^{m} \frac{1}{k} \\ &= \frac{1}{n} - \frac{1}{m} \\ &< \frac{1}{n}. \end{aligned}$$

This looks fairly simple!

Moving onto the second step: given $\epsilon > 0$, we want to force this quantity $\frac{1}{n} < \epsilon$. How can we do this? Well: if $m > n > N$, we have that $\frac{1}{n} < \frac{1}{N}$; so it suffices to pick N such that $\frac{1}{N} < \epsilon$.

Thus,we've shown that for any $\epsilon > 0$ we can find a N such that for any $m, n > N$,

$$|a_m - a_n| < \frac{1}{n} < \frac{1}{N} < \epsilon.$$

But this just means that our sequence is Cauchy! So, because all Cauchy sequences converge, we've proven that our sequence converges. □

The example above is interesting for a number of reasons: not only was it a nice way to illustrate the Cauchy condition, it was the first example of a **series**! We define series here:

Definition. A sequence is called **summable** if the sequence $\{s_n\}_{n=1}^\infty$ of partial sums

$$s_n := a_1 + \ldots a_n = \sum_{k=1}^{n} a_k$$

converges.

If it does, we then call the limit of this sequence the **series** associated to $\{a_n\}_{n=1}^\infty$, and denote this quantity by writing

$$\sum_{n=1}^{\infty} a_n.$$

We say that a series$\sum_{n=1}^\infty a_n$ **converges** or **diverges** if the sequence $\{\sum_{k=1}^n a_k\}_{n=1}^\infty$ of partial sums converges or diverges, respectively.

Claim. (Nonconvergence example:) The **harmonic series**

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges.

Proof. Notice the following inequalities:

$$\begin{aligned}
\left(\frac{1}{2}\right) &\geq \frac{1}{2} = \frac{1}{2}, \\
\left(\frac{1}{3} + \frac{1}{4}\right) &\geq 2 \cdot \frac{1}{4} = \frac{1}{2}, \\
\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) &\geq 4 \cdot \frac{1}{8} = \frac{1}{2}, \\
\left(\frac{1}{9} + \quad \ldots \quad + \frac{1}{16}\right) &\geq 8 \cdot \frac{1}{16} = \frac{1}{2}, \\
&\vdots
\end{aligned}$$

In particular, notice that by applying these inequalities, we can show that

$$\begin{aligned}
\sum_{n=1}^{2^n} \frac{1}{n} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{2^n} \\
&\geq 1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \ldots + \left(\frac{1}{2^{n-1}} + \ldots + \frac{1}{2^n}\right) \\
&\geq 1 + \frac{1}{2} + \ldots + \frac{1}{2} \\
&= 1 + \frac{n}{2},
\end{aligned}$$

for **any** value of n.

Therefore, for any fixed value of L, we have that for $n > 2L$,

$$\left|\sum_{n=1}^{2^n} \frac{1}{n} - L\right| \geq \left|1 + \frac{n}{2} - L\right| = 1 + \frac{n}{2} - L$$

So: take any L, and set $\epsilon = 1$. Then, for any N, if we pick n such that $2^n > N$ and $n > 2L$, we have

$$|a_{2^n} - L| = \left|\sum_{n=1}^{2^n} \frac{1}{n} - L\right| \geq \left|1 + \frac{n}{2} - L\right| = 1 + \frac{n}{2} - L > 1 = \epsilon.$$

This is the definition of nonconvergence! So we have proven our series does not have a limit, as claimed.

$\square$