Lebesgue integral

We will define a very powerful integral, far better than Riemann in the sense that it will allow us to integrate pretty much every reasonable function and we will also obtain strong convergence results. That is if we take a limit of integrable functions we will get an integrable function and the limit of the integrals will be the integral of the limit under very mild conditions. We will focus only on the real line, although the theory easily extends to more abstract contexts.

In Riemann integral the basic block was a rectangle. If we wanted to integrate a function that was identically 1 on an interval $[a, b]$, then the integral was simply the area of that rectangle, so $1 \times (b-a) = b-a$. For Lebesgue integral what we want to do is to replace the interval with a more general subset of the real line. That is, if we have a set $S \subset \mathbb{R}$ and we take the *indicator function* or *characteristic function* χ_S defined by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{else.} \end{cases}$$

Then the integral of χ_S should really be equal to the area under the graph, which should be equal to the "size" of S.

Example: Suppose that S is the set of rational numbers between 0 and 1. Let us argue that its size is 0, and so the integral of χ_S should be 0. Let $\{x_1, x_2, \ldots\} = S$ be an enumeration of the points of S. Now for any $\epsilon > 0$ take the sets

$$I_j = (x_j - \epsilon 2^{-j-1}, x_j + \epsilon 2^{-j-1}),$$

then

$$S \subset \bigcup_{j=1}^{\infty} I_j.$$

The "size" of any I_j should be $\epsilon 2^{-j}$, so it seems reasonable to say that the "size" of S is less than the sum of the sizes of the I_j's. At worst we are grossly overestimating; every I_j contains infinitely many other points of S, so there is a lot of overlap. So

$$\text{"size of } S\text{"} \leq \sum_{j=1}^{\infty} \text{"size of } I_j\text{"} = \sum_{j=1}^{\infty} \epsilon 2^{-j} = \epsilon.$$

So the "size of S" (whatever that concept should be) seems like it ought to be 0. And hence the integral of χ_S should be 0.

So to begin, we want to have a way to "measure" sets. We focus only on the real numbers and so suppose we wish to measure subsets of the real numbers. We would like (our Christmas wish) to have a function

$$m \colon \mathcal{P}(\mathbb{R}) \to [0, \infty]$$

that is a function that takes subsets of the real numbers and gives nonnegative extended real numbers, such that

$$m(\emptyset) = 0$$

and if $\{S_j\}$ is a countable collection of pairwise disjoint sets then

$$\sum_{j=1}^{\infty} m(S_j) = m\left(\bigcup_{j=1}^{\infty} S_j\right)$$

It should also replicate what we normally think of size of intervals, that is $m((a,b)) = m([a,b)) = m([a,b]) = m((a,b]) = b-a$.

Unfortunately, such a function is impossible. At least there is no such function on all of $\mathcal{P}(\mathbb{R})$ (the power set of the reals). We do have such a function on a subset of the powerset. That is, we will define a smaller set of subsets called measurable sets and on these sets we will be able to define such a function.

So let's talk about certain collections of sets. The collections we will want are so called σ-algebras (Rudin talks about σ-rings, the idea is very similar, I'll note what the difference is).

Definition: Let X be a set. A collection of sets $\mathcal{M} \subset \mathcal{P}(X)$ is a *σ-algebra* if

(i) $\mathcal{M}$ is nonempty,
(ii) $\mathcal{M}$ is closed under complements, that is, if $A \in \mathcal{M}$ then $A^c = X \setminus A \in \mathcal{M}$,
(iii) $\mathcal{M}$ is closed under countable unions, that is if $\{A_j\}$ is a countable collection of sets in $\mathcal{M}$ then

$$\bigcup_{j=1}^{\infty} A_j \in \mathcal{M}.$$

If $\mathcal{M}$ is closed only under finite unions, then we say that $\mathcal{M}$ is an *algebra*.

Most of the time below we will assume that $X = \mathbb{R}$, so you might as well think of subsets of the real line.

Definition of σ-ring and ring is similar but only needs closure under relative complements. A σ-algebra is always a σ-ring, and a σ-ring is a σ-algebra if it contains the whole set X as an element.

The sets in $\mathcal{M}$ are usually called *measurable sets*. We will define a certain function on the powerset and define a certain σ-algebra on which it has the desired properties. Our σ-algebra will be so large that we will essentially be able to integrate anything we want. It will be very hard to come up with sets that are not in our σ-algebra.

We will work with the *extended real numbers* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. We have previously used really only its order properties such as $-\infty < x < \infty$ for all $x \in \mathbb{R}$. Now we will also often use arithmetic on $\overline{\mathbb{R}}$. We have to be careful as $\overline{\mathbb{R}}$ will not be a field like $\mathbb{R}$. In fact, some operations are not even defined. Let us define

$$\begin{aligned}
&x \cdot \infty = \infty \qquad \text{for all } x > 0 \\
&x \cdot \infty = -\infty \qquad \text{for all } x < 0 \\
&x + \infty = \infty \quad \text{and} \quad x - \infty = -\infty \qquad \text{for all } x \in \mathbb{R} \\
&\frac{x}{\pm\infty} = 0 \qquad \text{for all } x \in \mathbb{R}
\end{aligned}$$

and so on. Everything that is not an indefinite form $\infty - \infty$, $\frac{\pm\infty}{\pm\infty}$, or $0 \cdot \infty$ has an obvious definition. It will be convenient for measure theory to define

$$0 \cdot \infty = 0.$$

We will have to avoid $\infty - \infty$ and $\frac{\pm\infty}{\pm\infty}$.

Definition: Let $\mathcal{M}$ be a σ-algebra. Let

$$\mu \colon \mathcal{M} \to \overline{\mathbb{R}}.$$

We say μ is *additive* if given $A, B \in \mathcal{M}$, disjoint ($A \cap B = \emptyset$) then

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

We say μ is *countably additive* if given $\{A_j\}$ a collection of sets in $\mathcal{M}$ such that $A_j \cap A_k = \emptyset$ for all $j \neq k$, then

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

Of course the sums have to make sense, so usually we will assume that μ does not achieve both $-\infty$ and ∞.

We will say that μ is nonnegative or monotonic if $\mu(A) \geq 0$ for all $A \in \mathcal{M}$.

We also say that μ is *countably subadditive* if for every collection $\{A_j\}$ we have

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j).$$

It is not too hard to show that if μ is additive then $\mu(\emptyset) = 0$. We also have additivity for arbitrary finite unions by induction.

If $B \subset A$ and $\mu(B)$ is finite, then writing $A = B \cup (A \setminus B)$ we obtain that

$$\mu(A \setminus B) = \mu(A) - \mu(B).$$

Another useful property for additive functions is

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B).$$

This follows by looking at the disjoint unions $B = (B \setminus A) \cup (A \cap B)$ and noting that $A \cup B = A \cup (B \setminus A)$. So for example a nonnegative additive function is also (finitely) subadditive:

$$\mu(A \cup B) \leq \mu(A) + \mu(B).$$

Countably additive functions are additive of course. Also, countably additive μ play nicely with limits.

Theorem 11.3: Suppose that μ is a countably additive function on a σ-algebra $\mathcal{M}$ and $A_1 \subset A_2 \subset \cdots$ are sets in $\mathcal{M}$ and $A = \cup_j A_j$, then

$$\lim_{n\to\infty} \mu(A_n) = \mu(A),$$

where the limit has the obvious interpretation for ∞ (or $-\infty$).

Proof. Write $B_1 = A_1$ and $B_j = A_j \setminus A_{j-1}$. Then the B_j's are pairwise disjoint and $A = \cup_j B_j$, so

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

As $A_n = B_1 \cup B_2 \cup \cdots \cup B_n$ then

$$\mu(A_n) = \sum_{j=1}^{n} \mu(B_j),$$

and the result follows. $\square$

Definition: If we have a σ-algebra $\mathcal{M}$ of measurable sets, then we call a function

$$\mu \colon \mathcal{M} \to \overline{\mathbb{R}}$$

a *measure* if it is a nonnegative and countably additive. Sometimes $\mu(\emptyset) = 0$ is also given as requirement, but that follows from additivity. Also some authors require μ to not be identically zero.

It turns out there are many different measures. The simplest measure can be defined as follows. Let $\mathcal{M}$ be all of $\mathcal{P}(X)$, and define $\mu(A) = |A|$, the cardinality of A. This μ is called the *counting measure*. Despite how trivial this example is, it does happen to be useful; we will see it later on.

Let us construct the Lebesgue measure. What we will actually construct is a subadditive nonnegative function on all of $\mathcal{P}(\mathbb{R})$, which will turn out to be a measure (so countably additive) on some large σ-algebra in $\mathcal{P}(\mathbb{R})$.

Let us define a bounded interval to be a set of the form

$$\{x : a < x < b\} \qquad \text{or} \qquad \{x : a \leq x < b\} \qquad \text{or} \qquad \{x : a < x \leq b\} \qquad \text{or} \qquad \{x : a \leq x \leq b\}$$

for real numbers $a \leq b$. We allow $a = b$, meaning we allow $\emptyset$ and the single point set $\{x\}$ to also be intervals. If I is a bounded interval, define

$$m(I) = b - a.$$

It is easy to see that given any bounded interval I and any $\epsilon > 0$, there are a closed interval F and an open interval G, with $F \subset A \subset G$ such that

$$m(G) - \epsilon \leq m(I) \leq m(F) + \epsilon.$$

Now the point is to show that we can extend m to a countably additive function on a σ-algebra that contains all the intervals.

Let $E \subset \mathbb{R}$ be any set. Let $\{I_j\}$ be a countable collection of bounded open intervals covering E, that is

$$E \subset \bigcup_{j=1}^{\infty} I_j.$$

Define the *outer measure* as

$$m^*(E) = \inf \sum_{j=1}^{\infty} m(I_j),$$

where the inf is taken over all coverings of E by countably many bounded open intervals.

It is immediate that m^* is nonnegative ($m^*(A) \geq 0$) and monotone (if $A \subset B$ then $m^*(A) \leq m^*(B)$).

Theorem 11.8: If I is a bounded interval, then $m(I) = m^*(I)$. Also m^* is countably subadditive.

That is m^* is a countably subadditive extension of m.

Proof. Suppose that I is a bounded interval, and let $\epsilon > 0$ be given. Then there exists an open bounded interval G, $I \subset G$, such that $m(G) \leq m(I) + \epsilon$. As G is a covering of I by bounded open intervals,

$$m^*(I) \leq m(G).$$

So $m^*(I) \leq m(G) \leq m(I) + \epsilon$. As $\epsilon > 0$ was arbitrary we have $m^*(I) \leq m(I)$. By definition of m^* there exists a sequence of open bounded intervals $\{G_j\}$ covering I such that

$$\sum_{j=1}^{\infty} m(G_j) \leq m^*(I) + \epsilon.$$

There also exists a bounded closed interval F, $F \subset I$, such that $m(F) \geq m(I) - \epsilon$. As F is compact, there is some N such that

$$F \subset G_1 \cup G_2 \cup \cdots \cup G_N.$$

and so

$$m(I) \leq \epsilon + m(F) \leq \epsilon + \sum_{j=1}^{N} m(G_j) \leq \epsilon + \sum_{j=1}^{\infty} m(G_j) \leq m^*(I) + 2\epsilon.$$

So $m(I) \leq m^*(I)$. Thus $m(I) = m^*(I)$.

Let us show countable subadditivity. Suppose that $A = \cup_{j=1}^{\infty} A_j$. If $m^*(A_j) = \infty$ for any j, then we are done, so suppose that $m^*(A_j)$ is finite for every j.

Each A_j has a covering G_{jk} of bounded open intervals such that

$$\sum_{k=1}^{\infty} m(G_{jk}) \leq m^*(A_j) + \epsilon 2^{-j}.$$

So as all the G_{jk} together cover A

$$m^*(A) \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} m(G_{jk}) \leq \sum_{j=1}^{\infty} \Big(m^*(A_j) + \epsilon 2^{-j} \Big) \leq \left(\sum_{j=1}^{\infty} m^*(A_j) \right) + \epsilon.$$

□

Note that by the same argument as for the example we started the section with, we have:

Corollary: If $S \subset \mathbb{R}$ is countable, then $m^*(S) = 0$.

It will be useful to have the following result about open subsets of $\mathbb{R}$:

Proposition: An open subset $W \subset \mathbb{R}$ is a countable union of pairwise disjoint open intervals.

Proof. For each point $x \in W$, let I_x be the largest open interval such that $I_x \subset W$ and $x \in I_x$ (that is, I_x is the union of all open intervals contained in W that contain x). Every I_x contains rational points. Furthermore if $y \in I_x$, then $I_y = I_x$. So

$$W = \bigcup_{x \in \mathbb{Q} \cap W} I_x$$

We take some enumeration of the rationals and pick one rational point in every I_x, then we have W written as a countable union of pairwise disjoint open intervals. □

Here we depart a little from Rudin again to have a simpler definition:

Definition: A set $E \subset \mathbb{R}$ is said to be *Lebesgue measurable* if for each subset $A \subset \mathbb{R}$ we get

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c).$$

We will denote the measurable sets by $\mathcal{M}$. And unless otherwise stated (that is, when talking about Lebesgue measure m or the associated outer measure m^*) $\mathcal{M}$ will mean Lebesgue measurable sets.

Note that

$$m^*(A) \leq m^*(A \cap E) + m^*(A \cap E^c)$$

is always true by subadditivity of m^*. So to show that E is measurable, what we need to show is that

$$m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c).$$

Furthermore, this inequality is always true when $m^*(A) = \infty$, so we only really need to worry about A such that $m^*(A) < \infty$.

If E is measurable then E^c is measurable by symmetry of the condition. It is not hard to see that $\emptyset$ and $\mathbb{R}$ are measurable.

Proposition: If $m^*(E) = 0$, then E is Lebesgue measurable.

Proof. For any set E we have

$$m^*(A \cap E) \leq m^*(E)$$

so $m^*(A \cap E) = 0$. Also

$$m^*(A \cap E^c) \leq m^*(A).$$

So

$$m^*(A \cap E) + m^*(A \cap E^c) \leq m^*(A).$$

□

So for example countable sets and their complements are Lebesgue measurable.

Sets of measure 0 are called *null sets*. We have seen above that all countable subsets of $\mathbb{R}$ are null sets, but there exist uncountable null sets as well.

Proposition: The set of Lebesgue measurable sets $\mathcal{M}$ is an algebra of sets.

Proof. As we said above, $\mathcal{M}$ is closed under complements. So we need to show that it is closed under finite unions.

Let E and F be measurable. Given any A we have

$$m^*(A \cap E^c) = m^*(A \cap E^c \cap F) + m^*(A \cap E^c \cap F^c) = m^*(A \cap E^c \cap F) + m^*\big(A \cap (E \cup F)^c\big)$$

and

$$m^*(A \cap E^c) = m^*(A) - m^*(A \cap E).$$

Also $A \cap (E \cup F) = (A \cap E) \cup (A \cap E^c \cap F)$ so

$$m^*\big(A \cap (E \cup F)\big) \leq m^*(A \cap E) + m^*(A \cap E^c \cap F).$$

Hence,

$$\begin{aligned} m^*\big(A \cap (E \cup F)\big) + m^*\big(A \cap (E \cup F)^c\big) &= m^*\big(A \cap (E \cup F)\big) + m^*(A \cap E^c) - m^*(A \cap E^c \cap F) \\ &= m^*(A) + m^*\big(A \cap (E \cup F)\big) - m^*(A \cap E) - m^*(A \cap E^c \cap F) \\ &\leq m^*(A). \end{aligned}$$

□

Proposition: Let $E_1, \dots, E_n$ be pairwise disjoint and measurable, then for any set A we have

$$m^*\left(A \cap \left(\bigcup_{j=1}^{n} E_j\right)\right) = \sum_{j=1}^{n} m^*(A \cap E_j).$$

Proof. The set E_n is measurable and hence

$$\begin{aligned} m^*\left(A \cap \left(\bigcup_{j=1}^{n} E_j\right)\right) &= m^*\left(A \cap \left(\bigcup_{j=1}^{n} E_j\right) \cap E_n\right) + m^*\left(A \cap \left(\bigcup_{j=1}^{n} E_j\right) \cap E_n^c\right) \\ &= m^*(A \cap E_n) + m^*\left(A \cap \left(\bigcup_{j=1}^{n-1} E_j\right)\right) \end{aligned}$$

and the proof follows by induction. □

Theorem: The set of Lebesgue measurable sets is a σ-algebra.

Proof. Suppose that $E = \cup_{j=1}^{\infty} E_j$ where all the E_j are measurable. Define $F_1 = E_1$ and $F_j = E_j \setminus \cup_{k=1}^{j-1} E_k$. We have that F_j is measurable for every j as $\mathcal{M}$ is an algebra. We have that $F_j \cap F_k = \emptyset$ if $j \neq k$, and also that $E = \cup_{j=1}^{\infty} F_j$.

Let A be any set. Then,

$$\begin{aligned} m^*(A) &= m^*\left(A \cap \bigcup_{j=1}^{n} F_j\right) + m^*\left(A \cap \left(\bigcup_{j=1}^{n} F_j\right)^c\right) \\ &\geq m^*\left(A \cap \bigcup_{j=1}^{n} F_j\right) + m^*(A \cap E^c) \\ &= \sum_{j=1}^{n} m^*(A \cap F_j) + m^*(A \cap E^c). \end{aligned}$$

Taking limits we have

$$m^*(A) \geq \sum_{j=1}^{\infty} m^*(A \cap F_j) + m^*(A \cap E^c) \geq m^*\left(A \cap \bigcup_{j=1}^{\infty} F_j\right) + m^*(A \cap E^c) = m^*(A \cap E) + m^*(A \cap E^c).$$

So E is measurable. □

Theorem: All intervals are Lebesgue measurable, and hence all open sets are measurable.

Proof. Let I be an interval of the form $(-\infty, x)$, $(-\infty, x]$, (x, ∞), or $[x, \infty)$. Let $\epsilon > 0$ be given and A be an arbitrary set such that $m^*(A) < \infty$. Let $\{I_n\}$ be a countable collection of open bounded intervals such that

$$A \subset \bigcup_{j=1}^{\infty} I_j,$$

and such that

$$\sum_{j=1}^{\infty} m(I_j) \leq m^*(A) + \epsilon.$$

Note that $I_j \cap I$ and $I_j \cap I^c$ are bounded intervals (could be empty). We have

$$m^*(A \cap I) \leq \sum_{j=1}^{\infty} m(I_j \cap I), \qquad \text{and}$$

$$m^*(A \cap I^c) \leq \sum_{j=1}^{\infty} m(I_j \cap I^c).$$

We have that $m(I_j) = m(I_j \cap I) + m(I_j \cap I^c)$. So

$$m^*(A \cap I) + m^*(A \cap I^c) \leq \sum_{j=1}^{\infty} m(I_j) \leq m^*(A) + \epsilon.$$

As $\epsilon > 0$ was arbitrary we obtain the required inequality. If $m^*(A) = \infty$ the inequality was trivial.

Any bounded interval is an intersection of two half infinite intervals as above, and so is measurable. Any open set is a countable union of open intervals, and so it is also measurable. □

We of course also get that all closed sets are measurable. But we get a lot more. We get that countable unions of closed sets are measurable, and so are countable intersections of open sets, and so on and so forth.

It is not hard to prove that an intersection of σ-algebras is still a σ-algebra. Therefore, there exists a smallest σ-algebra that contains the open sets (it's the intersection of all σ-algebras containing the open sets). This σ-algebra is denoted by $\mathcal{B}$ and the sets in it are called the *Borel sets*. As $\mathcal{B} \subset \mathcal{M}$, we have that all Borel sets are measurable. Sometimes it is just convenient to talk about $\mathcal{B}$ rather than $\mathcal{M}$.

Let us now define

$$m \colon \mathcal{M} \to [0, \infty]$$

by defining $m(E) = m^*(E)$. As m^* agreed with the earlier definition of m on intervals, this new m agrees with our earlier definition of m (on intervals). We have still not shown that m is a measure on $\mathcal{M}$. We call m the *Lebesgue measure* (we will show momentarily that it really is a measure, so the name is justified).

Theorem (like 11.10 in Rudin): m is countably additive, and hence a measure.

Proof. Let $\{E_j\}$ be a family of pairwise disjoint Lebesgue measurable sets and let $E = \cup_{j=1}^{\infty} E_j$. If $m(E_j) = \infty$ for any j, then $m(E) = \infty$ and additivity is trivial. So assume that $m(E_j) < \infty$ for all j.

Using $A = \mathbb{R}$ with an above proposition we have for any n

$$m(E) = m\left(\bigcup_{j=1}^{\infty} E_j\right) \geq m\left(\bigcup_{j=1}^{n} E_j\right) = \sum_{j=1}^{n} m(E_j).$$

Taking limits we have

$$m(E) \geq \sum_{j=1}^{\infty} m(E_j).$$

The opposite inequality follows by subadditivity. □

Proposition: If $E \subset \mathbb{R}$ is Lebesgue measurable, then for every $\epsilon > 0$ there exist an open set G and a closed set F such that $F \subset E \subset G$,

$$m(E \setminus F) < \epsilon, \qquad \text{and} \qquad m(G \setminus E) < \epsilon.$$

Proof. If $m(E) < \infty$ then G is found directly by definition of m^*. If $m(E) = \infty$, then we have to work a little harder. So look at the sets $E_j = E \cap [j, j+1)$. We have that $m(E_j) \leq m\big([j, j+1)\big) < 1 < \infty$, and $E = \cup_{j=-\infty}^{\infty} E_j$. For every j we can find an open set G_j such that, $E_j \subset G_j$ and $m(G_j \setminus E_j) < \epsilon 2^{-|j|}$. Let $G = \cup_{j=-\infty}^{\infty} G_j$. So

$$m(G \setminus E) = m\left(\bigcup_{j=-\infty}^{\infty} (G_j \setminus E)\right) \leq m\left(\bigcup_{j=-\infty}^{\infty} (G_j \setminus E_j)\right) \leq \sum_{j=-\infty}^{\infty} m(G_j \setminus E_j) < \sum_{j=-\infty}^{\infty} \epsilon 2^{-|j|} = 3\epsilon.$$

Then to find F, take the complement E^c and find an open set that covers it and take a complement of that. Details are left to student. □

We remark that by letting ϵ go to 0, we can show (left to students) that there exists a Borel set G that is a countable intersection of open sets, and a Borel set F that is a countable union of closed sets, such that $F \subset E \subset G$

$$m(E \setminus F) = m(G \setminus E) = 0.$$

Note that, of course, $m(E) = m(F) = m(G)$. So every Lebesgue measurable set is almost like a Borel set; the difference is a null set.

Measurable functions

If we want to integrate functions, we want to know which functions play nicely with the measure, or actually with the measurable sets. For example if S is a nonmeasurable set, then we don't expect to be able to integrate the characteristic function χ_S, as its integral should be the measure of S.

Let us work in a general *measurable space* $(X, \mathcal{M})$, that is, a set X and a σ-algebra of sets $\mathcal{M}$. If you want to, you can think of $(\mathbb{R}, \mathcal{M})$, where $\mathcal{M}$ are the Lebesgue measurable sets. Note that we will not worry about the actual measure.

Definition 11.13: Let $(X, \mathcal{M})$ is a measurable space. $f\colon X \to \overline{\mathbb{R}}$ is said to be *measurable* if

$$f^{-1}\big((a, \infty]\big) = \{x \in X : f(x) > a\} \in \mathcal{M}$$

for all $a \in \mathbb{R}$.

If $X = \mathbb{R}$ and $\mathcal{M}$ is the set of Lebesgue measurable sets, then we say that f is said to be *Lebesgue measurable.* If $X = \mathbb{R}$ and $\mathcal{M} = \mathcal{B}$ is the σ-algebra of Borel sets, then f is said to be *Borel measurable.* Note that if a function is Borel measurable then it is, of course, Lebesgue measurable.

When people speak of just "measurable" functions on the real line, they will generally mean Lebesgue measurable.

Proposition: If $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, then it is Borel measurable (and hence Lebesgue measurable).

Proof. The interval (a, ∞) is open and so $f^{-1}\big((a, \infty)\big)$ is open and so Borel (and so Lebesgue measurable as well). □

Theorem 11.15: Let $(X, \mathcal{M})$ is a measurable space and $f\colon X \to \overline{\mathbb{R}}$ a function. The following are equivalent:

(i) f is measurable, that is $\{x \in X : f(x) > a\}$ is measurable for all $a \in \mathbb{R}$.
(ii) $\{x \in X : f(x) \geq a\}$ is measurable for all $a \in \mathbb{R}$.
(iii) $\{x \in X : f(x) < a\}$ is measurable for all $a \in \mathbb{R}$.
(iv) $\{x \in X : f(x) \leq a\}$ is measurable for all $a \in \mathbb{R}$.

Proof. The implications (i) implies (ii) implies (iii) implies (iv) implies (i) are shown by the following equalities:

$$\begin{aligned}
\{x \in X : f(x) \geq a\} &= \bigcap_{n=1}^{\infty} \{x \in X : f(x) > a - 1/n\}, \\
\{x \in X : f(x) < a\} &= X \setminus \{x \in X : f(x) \geq a\}, \\
\{x \in X : f(x) \leq a\} &= \bigcap_{n=1}^{\infty} \{x \in X : f(x) < a + 1/n\}, \\
\{x \in X : f(x) > a\} &= X \setminus \{x \in X : f(x) \leq a\}.
\end{aligned}$$

□

Similarly we also can prove that $f^{-1}(\{\infty\})$ and $f^{-1}(\{-\infty\})$ are measurable. So we could let a vary over all of $\overline{\mathbb{R}}$.

Theorem 11.16 (and corollary): Let $(X, \mathcal{M})$ is a measurable space and $f \colon X \to \overline{\mathbb{R}}$ and $g \colon X \to \overline{\mathbb{R}}$ are measurable then

(i) $|f|$ is measurable.
(ii) $\max(f, g)$ and $\min(f, g)$ are measurable.
(iii) $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are measurable. (Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$)

Proof. First item follows by $\{x : |f(x)| < a\} = \{x : f(x) < a\} \cap \{x : f(x) > -a\}$

Second item follows by writing $\{x : \max(f, g)(x) < a\} = \{x : f(x) < a\} \cap \{x : g(x) < a\}$ and $\{x : \min(f, g)(x) < a\} = \{x : f(x) < a\} \cup \{x : g(x) < a\}$.

Last item follows by the second item. □

In fact essentially any reasonable (see below about composition) operation we do to measurable functions lands us back in the set of measurable functions.

Theorem 11.17: Let $(X, \mathcal{M})$ is a measurable space and let $\{f_n\}$ be a sequence of measurable functions defined on X. Define

$$\begin{aligned}
g_1(x) &= \sup_{n \in \mathbb{N}} f_n(x), \\
g_2(x) &= \inf_{n \in \mathbb{N}} f_n(x), \\
g_3(x) &= \limsup_{n \to \infty} f_n(x), \\
g_4(x) &= \liminf_{n \to \infty} f_n(x).
\end{aligned}$$

Then g_1, g_2, g_3, and g_4 are all measurable. In particular, if $\{f_n\}$ converges pointwise to f, then f is measurable.

Proof. If $g_1(x) > a$, then there is some n such that $f_n(x) > a$. Similarly, if $f_n(x) > a$ for some n, then obviously $g_1(x) > a$. So

$$\{x : g_1(x) > a\} = \{x : \sup_{n \in \mathbb{N}} f_n(x) > a\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) > a\}.$$

In other words g_1 is measurable.

Similarly,

$$\{x : g_2(x) < a\} = \{x : \inf_{n \in \mathbb{N}} f_n(x) < a\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) < a\}.$$

So g_2 is measurable.

Next notice that

$$g_3(x) = \limsup_{n\to\infty} f_n(x) = \inf_{m\in\mathbb{N}} \left(\sup_{n\geq m} f_n(x) \right),$$

$$g_4(x) = \liminf_{n\to\infty} f_n(x) = \sup_{m\in\mathbb{N}} \left(\inf_{n\geq m} f_n(x) \right).$$

So g_3 and g_4 are also measurable.

If the sequence is convergent, then limit is equal to limsup (or liminf) and hence f is measurable. □

Composition is somewhat tricky. Even if $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ are Lebesgue measurable, doesn't mean that $f \circ g$ is measurable. First we notice that what really happens is that a Lebesgue measurable function is a function that takes Borel sets on $\mathbb{R}$ into Lebesgue measurable sets, that is, if A is a Borel set then $g^{-1}(A)$ is Lebesgue measurable. The inverse image of a Lebesgue measurable set need not be Lebesgue measurable for a Lebesgue measurable function. We would need something stronger:

Proposition: If $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ are both Borel measurable, then $f \circ g$ is Borel measurable.

The proof is left to student. On the other hand there exist examples of even a continuous g and Lebesgue measurable f so that $f \circ g$ is not Lebesgue measurable.

Theorem 11.18: Let $(X, \mathcal{M})$ is a measurable space, $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ be measurable functions, and $F\colon \mathbb{R}^2 \to \mathbb{R}$ be a continuous function, then $h(x) = F\big(f(x), g(x)\big)$ is a measurable function.

In particular $f + g$ and fg are measurable.

Proof. Fix $a \in \mathbb{R}$, then look at the open set

$$G = \{(y_1, y_2) : F(y_1, y_2) > a\}.$$

An open set contains a whole ball around every point. So for every point $y = (y_1, y_2)$ in G there is a $\delta > 0$ such that

$$\{(z_1, z_2) : y_1 - \delta < z_1 < y_1 + \delta,\ y_2 - \delta < z_2 < y_2 + \delta\} \subset G.$$

Since $\mathbb{R}^2$ contains a dense countable subset (the set of points with rational coordinates), there are countably many such sets whose union is G. That is, there exist sequences $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, and $\{d_n\}$ and

$$I_n = \{(z_1, z_2) : a_n < z_1 < b_n, c_n < z_2 < d_n\},$$

such that

$$G = \bigcup_{n=1}^{\infty} I_n.$$

Then

$$\begin{aligned}
\{x : h(x) > a\} &= \{x : (f(x), g(x)) \in G\} \\
&= \bigcup_{n=1}^{\infty} \{x : a_n < f(x) < b_n,\ c_n < g(x) < b_n\} \\
&= \bigcup_{n=1}^{\infty} \big(\{x : a_n < f(x)\} \cap \{x : f(x) < b_n\} \cap \{x : c_n < g(x)\} \cap \{x : g(x) < b_n\}\big).
\end{aligned}$$

And so $\{x : h(x) > a\}$ is measurable. □

Let us motivate what we will do next. For Riemann integral (using the Darboux approach) we really took step functions that were less than the function, integrated those and took their supremum (that was the lower Darboux integral). A step function is a function that is constant on intervals, that is a function such that if $I_1, I_2, \ldots, I_n$ are disjoint intervals and $c_1, \ldots, c_n$ are numbers then a step function is a function of the form

$$s(x) = \sum_{j=1}^{n} c_j \chi_{I_j}(x),$$

where χ_{I_j} is the characteristic function of I_j (the function that is 1 on I_j and 0 elsewhere). The integral of s was easy to define then

$$\int s(x)\,dx = \sum_{j=1}^{n} c_j m(I_j),$$

and $m(I_j)$ is just the length of the jth interval. Then if we take the supremum of those sums, that is the integrals of those step functions less than f, we get the integral of f.

For the Lebesgue approach we will do something very similar, except that we now know how to measure a lot more sets, so we we can replace the I_j with arbitrary measurable sets. Let us first see what we replace the step function with.

Definition: Let $(X, \mathcal{M})$ is a measurable space. A function $s\colon X \to \mathbb{R}$ is said to be a *simple function* if the range is finite. In other words, s is simple if it attains only finitely many values.

Suppose that s is a simple function and $s(X) = \{c_1, c_2, \ldots, c_n\}$. Then let

$$E_j = \{x : s(x) = c_j\},$$

and we can write

$$s(x) = \sum_{j=1}^{n} c_j \chi_{E_j}(x),$$

where χ_{E_j} is the characteristic function of E_j (the function that is 1 on E_j and 0 elsewhere). We note that s is measurable if and only if E_1, E_2, ..., E_n are measurable.

Be careful though, just because s has this form and is measurable doesn't mean that the E_j are measurable. For example if E is a nonmeasurable set then $1 = \chi_E + \chi_{E^c}$, which is a measurable simple function. The reason we made the "if and only if" statement is because the c_j are all distinct numbers and the E_j are disjoint.

It turns out that every function can be approximated by simple functions.

Theorem 11.20: Let $(X, \mathcal{M})$ is a measurable space. Let $f\colon X \to \mathbb{R}$ be a function. Then there is a sequence $\{s_n\}$ of simple functions converging pointwise to f. If $f \geq 0$, we can choose $\{s_n\}$ to be monotonically increasing, that is $\{s_n(x)\}$ is a monotonically increasing sequence for every x. Finally, if f is measurable, then we can choose all the s_n to be measurable.

Proof. First suppose that $f \geq 0$ and define for each $n \in \mathbb{N}$, and all $j = 1, 2, \ldots, n2^n$, define

$$E_{n,j} = \left\{ x : \frac{j-1}{2^n} \leq f(x) < \frac{j}{2^n} \right\},$$

and

$$F_n = \{x : f(x) \geq n\}.$$

Let

$$s_n = \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \chi_{E_{n,j}} + n\chi_{F_n}.$$

A moment's reflection will show that $\{s_n(x)\}_{n=1}^{\infty}$ really does converge to $f(x)$. Furthermore, by construction all the sets are measurable if f is measurable.

Finally if f is not nonnegative, write $f = f^+ - f^-$ and apply the above construction to f^+ and f^- separately. □

Note that in the proof, if the function f is bounded, then beyond a certain n, the F_n are all empty. Then we must be at most 2^{-n} from the value. That means that the sequence s_n converges uniformly to f in this case (only if f is bounded).

The integral

Let X be a set and $\mathcal{M}$ a σ-algebra, and μ a measure. The triple $(X, \mathcal{M}, \mu)$ is then called a *measure space*. We will from now work in such an abstract measure space. Again, if you wish, you can just think of

$X = \mathbb{R}$, $\mathcal{M}$ the Lebesgue measurable sets and $\mu = m$, the Lebesgue measure, but most of what we prove will work for an arbitrary measure space.

Definition: Suppose that

$$s(x) = \sum_{j=1}^{n} c_j \chi_{E_j}(x)$$

is measurable (and all the E_j's are measurable) and suppose that $c_j > 0$. Then define

$$\int s \, d\mu = \sum_{j=1}^{n} c_j \mu(E_j).$$

Given a measurable nonnegative function f, let $\mathcal{S}$ be the set of measurable nonnegative simple functions s such that $0 \leq s \leq f$

$$\int f \, d\mu = \sup_{s \in \mathcal{S}} \int s \, d\mu.$$

We leave it to the student to check that this is well defined if f is a simple function. We call $\int f \, d\mu$ the *Lebesgue integral* with respect to μ. We sometimes write

$$\int f(x) \, d\mu(x),$$

in case the variable is important. If the set X needs to be emphasized we write

$$\int_X f \, d\mu.$$

And for a measurable subset E we can define

$$\int_E f \, d\mu = \int f \chi_E \, d\mu.$$

In the special case of Lebesgue measure we may write

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_{\mathbb{R}} f \, dm, \qquad \int_a^b f(x) \, dx = \int_{[a,b]} f \, dm.$$

We will later prove that the notation is justified as we will obtain the same values as the Riemann integral for Riemann integrable functions.

Also note that we could take $X \subset \mathbb{R}$ to be a measurable subset, and then we could let μ be the restriction of m to the measurable subsets of X. Then

$$\int_X f|_X \, d\mu = \int_{\mathbb{R}} f \chi_X \, dm,$$

where one integral exists if and only if the other one does.

Definition: For an arbitrary measurable function f write $f = f^+ - f^-$ and if at least one of the integrals

$$\int f^+ \, d\mu \qquad \text{and} \qquad \int f^- \, d\mu$$

is finite, we define

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

If both $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ are finite then we say f is *integrable* (or *summable*) or perhaps more precisely f is Lebesgue integrable with respect to μ and we write $f \in L^1(\mu)$ or $f \in L^1(X, \mu)$. If $E \subset X$ is measurable, then $L^1(E, \mu)$ has the obvious meaning. We may write L^1 or $L^1(X)$ if the measure is clear from context.

Note that we require both of the integrals to be finite to say integrable.

Proposition:

(i) If $a \leq f(x) \leq b$ for all $x \in E$ and $\mu(E) < \infty$, then
$$a\mu(E) \leq \int_E f \, d\mu \leq b\mu(E).$$
In particular, if $\mu(E) < \infty$ and a real-valued f is bounded on E, then $f \in L^1(E, \mu)$.

(ii) Suppose that f, g are either integrable or f, g are nonnegative and measurable. If $f(x) \leq g(x)$ for all x, then
$$\int f \, d\mu \leq \int g \, d\mu.$$

(iii) If $f \geq 0$ is measurable, and A and B are disjoint and measurable then
$$\int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu.$$

Proof. For part (i) note that $a\chi_E(x) \leq f(x)\chi_E(x) \leq b\chi_E(x)$ and $a\chi_E$ and $b\chi_E$ are simple functions. Without loss of generality assume that $E = X$. If $a \geq 0$, then $f = f^+$ and $f^- = 0$, and also $a \leq f$. So the first inequality follows. Any simple function less than f is also less than b showing the second inequality. The cases $b \leq 0$ and $a < 0 < b$ follow similarly.

Part (ii) can be proved by noting that $f^+ \leq g^+$ and $f^- \geq g^-$. So we only need to prove the result for nonnegative measurable functions. If s is simple and $s \leq f$, then $s \leq g$ and the result follows.

Let us prove part (iii). Let $s \leq f\chi_{A \cup B}$ be a nonnegative measurable simple function $s = \sum_{j=1}^n c_j \chi_{E_j}$ then
$$\int_{A \cup B} s \, d\mu = \sum_{j=1}^n c_j \mu(E_j) = \sum_{j=1}^n c_j \mu(E_j \cap A) + \sum_{j=1}^n c_j \mu(E_j \cap B) = \int_A s \, d\mu + \int_B s \, d\mu.$$
Note that if $0 \leq s \leq f\chi_{A \cup B}$ then $s\chi_A \leq f\chi_A$ and $s\chi_A \leq f\chi_A$. Therefore taking suprema over all such s we get
$$\int_{A \cup B} f \, d\mu \leq \int_A f \, d\mu + \int_B f \, d\mu.$$
If $\int_A f \, d\mu = \infty$ or $\int_B f \, d\mu = \infty$, then $\int_{A \cup B} f \, d\mu = \infty$ and equality follows. So let's assume that all 3 are finite. Given $\epsilon > 0$ find a measurable simple $s \leq f\chi_{A \cup B}$ such that
$$\int_A s \, d\mu \geq \int_A f \, d\mu - \epsilon \qquad \text{and} \qquad \int_B s \, d\mu \geq \int_B f \, d\mu - \epsilon.$$
This is not hard to do as A and B are disjoint, so just find s_1 that works on A (and is zero outside of A) and s_2 that works for B (and is zero outside of B) and let $s = s_1 + s_2$. Then
$$\int_{A \cup B} f \, d\mu \geq \int_{A \cup B} s \, d\mu = \int_A s \, d\mu + \int_B s \, d\mu \geq \int_A f \, d\mu + \int_B f \, d\mu - 2\epsilon.$$
$\square$

Let us integrate complex valued functions.

Definition: Suppose that $f : X \to \mathbb{C}$ is a function. If $f = u + iv$ where u and v are real-valued, then we say that f is *measurable* if u and v are.

If u and v are integrable, then we say that f is *integrable* and we write
$$\int f \, d\mu = \int u \, d\mu + i \int v \, d\mu.$$

Note that if f is measurable then $|f| = \sqrt{u^2 + v^2}$ is also measurable.

In general when we write $L^1(X, \mu)$ from now on we will mean complex valued functions. It turns out there is no loss in generality by not allowing the values $\pm\infty$ for integrable functions. The set where an L^1 function could be ∞ must be a null set.

Proposition:

(i) If $\mu(E) < \infty$ and $f\colon X \to \mathbb{C}$ is measurable and bounded on E, then $f \in L^1(E, \mu)$.

(ii) If $f \in L^1(\mu)$ and A and B are disjoint and measurable, then

$$\int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu.$$

(iii) If $f \in L^1(\mu)$ and $c \in \mathbb{C}$, then $cf \in L^1(\mu)$ and

$$\int cf \, d\mu = c \int f \, d\mu.$$

(iv) If $\mu(E) = 0$ and $f\colon X \to \mathbb{C}$ is measurable then $f \in L^1(E, \mu)$ and

$$\int_E f \, d\mu = 0.$$

(v) If $f \in L^1(\mu)$ and A and B are measurable with $B \subset A$ and $\mu(A \setminus B) = 0$ then

$$\int_A f \, d\mu = \int_B f \, d\mu.$$

(vi) If $f \in L^1(X, \mu)$ and $E \subset X$ is measurable, then $f \in L^1(E, \mu)$.

Proof. We leave the proof to the reader. Note that for example parts (i) and (ii) follow almost trivially from parts (i) and (iii) of the proposition for real functions. $\square$

We note that the above proposition, among other things shows that measure zero sets are not relevant to integration, that is the integral doesn't see something that happens on a measure zero set. This leads us to the following definition.

Definition: Let $(X, \mathcal{M}, \mu)$ be a measure space as above and let f and g be functions defined on X. We write

$$f = g \quad \textit{almost everywhere}$$

if the set

$$E = \{x : f(x) \neq g(x)\}$$

is a null set, that is $\mu(E) = 0$. We will say that $f = g$ *almost everywhere on* A, where $A \subset X$, if $f|_A = g|_A$ almost everywhere, or in other words if

$$\mu(\{x : f(x) \neq g(x)\} \cap A) = 0.$$

If something happens outside of a measure zero set we say it happens almost everywhere. For example, we write

$$f \leq g \qquad \text{almost everywhere,}$$

if the set where $f(x) \not\leq g(x)$ is of measure zero. Sometimes we just write

$$f = g \text{ a.e.} \qquad \text{or} \qquad f(x) = g(x) \text{ a.e.}$$

Proposition:

(i) The relation $f = g$ almost everywhere is an equivalence relation.

(ii) If $f = g$ almost everywhere, then

$$\int f \, d\mu = \int g \, d\mu.$$

The proof is easy. For equivalence relation you must prove that First, we have that $f = f$ a.e. Further, if $f = g$ a.e., then $g = f$ a.e. Finally, if $f = g$ a.e. and $g = h$ a.e., then $f = h$ a.e. The second item follows by integrating only on the set where f and g are equal.

When talking about $L^1(X, \mu)$, we usually talk about the equivalence class of functions under equality almost everywhere. That is, if $f = g$ a.e., then we just consider f and g the same element of $L^1(X, \mu)$. It is a common abuse of notation to consider $L^1(X, \mu)$ to be either the set of integrable functions or the set of equivalence classes. So we write $f \in L^1$ even though we really mean that f is a member of an equivalence class that itself is a member of L^1. Notice also that when talking about $L^1(X, \mu)$, we only need to consider complex-valued (or real-valued) functions, and ignore where the set where the function is infinite; if a function is integrable and has values in the extended reals, then it is equal almost everywhere to a function that is just real-valued.

Many results involving the integral only require a hypothesis that holds almost everywhere. It is generally very easy to see when this is possible, for example suppose that $f \leq g$ almost everywhere and f and g are either nonnegative or in L^1 (so that the integral is defined). Then using the proposition above we obtain

$$\int f \, d\mu \leq \int g \, d\mu.$$

Theorem 11.24: Suppose that $(X, \mathcal{M}, \mu)$ is a measure space, f is measurable and $f \geq 0$. The function $\varphi \colon \mathcal{M} \to \overline{\mathbb{R}}$ defined by

$$\varphi(A) = \int_A f \, d\mu$$

is countably additive. Furthermore, if $f \in L^1(X, \mu)$ then $\varphi \colon \mathcal{M} \to \mathbb{C}$ defined in the same way is also countably additive.

Proof. If the theorem is true for $f \geq 0$, then it follows for $f \in L^1$ by writing $f = u + iv$, $u = u^+ - u^-$, and $v = v^+ - v^-$. So let us just assume that $f \geq 0$. Notice that this makes φ nonnegative as well.

Let $\{E_n\}$ be a countable collection of pairwise disjoint measurable sets and let $E = \cup_{n=1}^{\infty} E_n$. If $\varphi(E_n) = \infty$ for any n, then as

$$\varphi(E_n) = \int \chi_{E_n} f \, d\mu \leq \int \chi_E f \, d\mu = \varphi(E)$$

we also get that $\varphi(E) = \infty$. So countable additivity follows trivially. So from now on assume that $\varphi(E_n) < \infty$ for all n.

If $f = \sum_{j=1}^m c_j \chi_{A_j}$ is a measurable nonnegative simple function (all the $c_j \geq 0$ and all the A_j are measurable) then

$$\begin{aligned}
\varphi(E) &= \int_E \sum_{j=1}^m c_j \chi_{A_j} \, d\mu = \int \sum_{j=1}^m c_j \chi_{A_j} \chi_E \, d\mu = \int \sum_{j=1}^m c_j \chi_{A_j \cap E} \, d\mu \\
&= \sum_{j=1}^m c_j \mu\big(A_j \cap (\cup_{n=1}^{\infty} E_n)\big) = \sum_{j=1}^m c_j \mu\big(\cup_{n=1}^{\infty} (A_j \cap E_n)\big) = \sum_{j=1}^m c_j \sum_{n=1}^{\infty} \mu(A \cap E_n) \\
&= \sum_{n=1}^{\infty} \sum_{j=1}^m c_j \mu(A \cap E_n) = \sum_{n=1}^{\infty} \int \sum_{j=1}^m c_j \chi_{A_j \cap E_n} \, d\mu = \sum_{n=1}^{\infty} \int_{E_n} \sum_{j=1}^m c_j \chi_{A_j} \, d\mu = \sum_{n=1}^{\infty} \varphi(E_n).
\end{aligned}$$

So suppose that $f \geq 0$ is any measurable function. If $0 \leq s \leq f$ and s is simple then

$$\int_E s \, d\mu = \sum_{n=1}^{\infty} \int_{E_n} s \, d\mu \leq \sum_{n=1}^{\infty} \int_{E_n} f \, d\mu = \sum_{n=1}^{\infty} \varphi(E_n).$$

By definition of the integral when we take the supremum of the simple functions less than f we get

$$\varphi(E) = \int_E f \, d\mu \leq \sum_{n=1}^{\infty} \varphi(E_n).$$

Remember that $\varphi(E_n) < \infty$ for all n. Let $\epsilon > 0$ be given. Find a measurable simple $s \geq 0$ such that for all $j = 1, \ldots, n$ we have

$$\int_{E_j} s \, d\mu \geq \int_{E_j} f \, d\mu - \epsilon = \varphi(E_j) - \epsilon.$$

Again this is easy directly from the definition as all the E_j are pairwise disjoint.

$$\varphi(\cup_{j=1}^n E_j) \geq \int_{\cup_{j=1}^n E_j} s\, d\mu = \sum_{j=1}^n \int_{E_j} s\, d\mu \geq \sum_{j=1}^n \bigl(\varphi(E_j) - \epsilon\bigr) = \left(\sum_{j=1}^n \varphi(E_j)\right) - n\epsilon.$$

As $\epsilon > 0$ we obtain

$$\varphi\left(\bigcup_{j=1}^n E_j\right) \geq \sum_{j=1}^n \varphi(E_j).$$

Next,

$$\varphi(E) \geq \varphi\left(\bigcup_{j=1}^n E_j\right) \geq \sum_{j=1}^n \varphi(E_j).$$

Taking limits we get

$$\varphi(E) \geq \sum_{j=1}^\infty \varphi(E_j).$$

And we obtain countable additivity. □

Theorem (Triangle inequality for the integral): (extended 11.26 from Rudin) For a measurable function f on a measure space $(X, \mathcal{M}, \mu)$ we have $f \in L^1(X, \mu)$ if and only if $|f| \in L^1(X, \mu)$, and in this case,

$$\left| \int f\, d\mu \right| \leq \int |f|\, d\mu.$$

Often we write

$$\|f\|_{L^1} = \|f\|_{L^1(X,\mu)} = \int |f|\, d\mu.$$

This *norm* provides the "distance from the origin" for the space L^1, and will actually make L^1 into a complete metric space (this will be an exercise) if we consider elements of L^1 to be the equivalence classes of functions under equality almost everywhere as we mentioned above. The proposition gives a way of testing that f is in L^1 by testing that $\|f\|_{L^1} < \infty$. The left hand side of the inequality in the theorem does not always make sense, but the right hand side makes sense for any measurable function if we allow it to be infinite.

Proof. First suppose that f is real-valued and write $f = f^+ - f^-$. Let $A = \{x : f(x) \geq 0\}$ and $B = \{x : f(x) < 0\}$. Then A and B are measurable and disjoint and $X = A \cup B$. So

$$\int |f|\, d\mu = \int_A |f|\, d\mu + \int_B |f|\, d\mu = \int_A f^+\, d\mu + \int_B f^-\, d\mu = \int f^+\, d\mu + \int f^-\, d\mu.$$

If $f \in L^1$, then the right hand side is finite and so $|f|$ (which is a nonnegative function) must be in L^1. Similarly if the left hand side is finite then the right hand side must be finite, because a sum of two nonnegative extended real numbers is finite if and only if they are both finite.

Now assume that f complex valued. First suppose that $|f| \in L^1$. Then $\bigl(\mathrm{Re}(f)\bigr)^+ \leq |f|$ and $\bigl(\mathrm{Re}(f)\bigr)^- \leq |f|$. As

$$\int \bigl(\mathrm{Re}(f)\bigr)^+ d\mu \leq \int |f|\, d\mu < \infty \qquad \text{and} \qquad \int \bigl(\mathrm{Re}(f)\bigr)^- d\mu \leq \int |f|\, d\mu < \infty,$$

we have that $\mathrm{Re}(f)$ is integrable. Similarly, $\mathrm{Im}(f)$ is integrable and therefore f itself is integrable.

Next suppose that $f \in L^1$. That means that if $f = u + iv$, then u and v are in L^1 and so $|u|$ and $|v|$ are in L^1 as we saw above. By triangle inequality we have $|f| \leq |u| + |v|$. Let $A = \{x : |u(x)| \geq |v(x)|\}$ and $B = \{x : |u(x)| < |v(x)|\}$. Then A and B are measurable and disjoint and $X = A \cup B$. On A we have $|f| \leq 2\,|u|$ and on B we have $|f| \leq 2\,|v|$ and

$$\int |f|\, d\mu = \int_A |f|\, d\mu + \int_B |f|\, d\mu \leq 2 \int_A |u|\, d\mu + 2 \int_B |v|\, d\mu.$$

And that's finite. Note that the argument could be somewhat simpler if we already knew linearity of the integral; we will prove linearity little later.

To show the inequality in case $f \in L^1$, we find a $c \in \mathbb{C}$ such that $|c| = 1$ and

$$\left| \int f \, d\mu \right| = c \int f \, d\mu = \int cf \, d\mu.$$

And cf is also L^1. Next, the integral of cf is real so

$$\int cf \, d\mu = \int \operatorname{Re}(cf) \, d\mu + i \int \operatorname{Im}(cf) \, d\mu = \int \operatorname{Re}(cf) \, d\mu.$$

And finally we have that for every x

$$\operatorname{Re}\bigl(cf(x)\bigr) \leq |cf(x)| = |f(x)| \, .$$

So

$$\left| \int f \, d\mu \right| = \int \operatorname{Re}(cf) \, d\mu \leq \int |f| \, d\mu.$$

$\square$

One way the theorem sometimes arises is that if we find a $g \in L^1(X, \mu)$ such that $|f| \leq g$ almost everywhere (or perhaps even everywhere), then $f \in L^1(X, \mu)$ (see Theorem 11.27 in Rudin). This just follows trivially.

We now get to one of the main theorems in the theory of the Lebesgue integral, one of those that make the Lebesgue theory so useful. The three theorems I am talking about is Lebesgue's monotone convergence theorem, Fatou's lemma (Rudin calls it a theorem), and Lebesgue's dominated convergence theorem. (This is a hint: these theorems will almost surely (look up "almost surely" on wikipedia) be on the exam).

Theorem 11.28 (Lebesgue's monotone convergence theorem): Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\{f_n\}$ be a sequence of nonnegative measurable functions such that

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots$$

for all x. Let

$$f(x) = \lim_{n\to\infty} f_n(x) \qquad \left(= \sup_{n \in \mathbb{N}} f_n(x)\right).$$

Then

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

That is, for a monotone sequence of functions we can always swap the limit and the integral.

Proof. The sequence $\int f_n \, d\mu$ is monotone, so there is some L (possibly infinity) with

$$L = \lim_{n\to\infty} \int f_n \, d\mu.$$

We also have by monotonicity that $\int f_n \, d\mu \leq \int f \, d\mu$, so

$$L \leq \int f \, d\mu.$$

Let $c \in (0, 1)$ be a number and let s be a measurable simple function such that $0 \leq s \leq f$. Further, let

$$E_n = \{x : f_n(x) \geq cs(x)\}.$$

It is clear that $E_1 \subset E_2 \subset \cdots$ by monotonicity of the sequence $\{f_n\}$. As $s(x) \leq f(x)$ we have $cs(x) < f(x)$ and so eventually for any x, there is an n such that $f_n(x) \geq cs(x)$. Hence, $X = \cup_{n=1}^{\infty} E_n$.

$$L \geq \int f_n \, d\mu \geq \int_{E_n} f_n \, d\mu \geq c \int_{E_n} s \, d\mu.$$

The integral of s over a set is a countably additive function by Theorem 11.24, and so by Theorem 11.3. So the right hand side converges to $c \int s \, d\mu$, and hence

$$L \geq c \int s \, d\mu.$$

As this is true for arbitrary $c \in (0,1)$ we get $L \geq \int s \, d\mu$. This was an arbitrary simple measurable function s less than f, so

$$L \geq \int f \, d\mu.$$

And we are done. □

Let us use the monotone convergence theorem to prove linearity of the integral.

Theorem 11.29: Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose f, g are nonnegative and measurable then

$$\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu.$$

Furthermore, if $f, g \in L^1(X, \mu)$, then $h = f + g$ is also in L^1 and we also get

$$\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu.$$

Proof. First suppose that f, g are nonnegative. It is not hard to see linearity for simple functions, so the result holds for simple functions. Now choose a monotone sequences of simple functions $\{s_n\}$ and $\{r_n\}$ converging to f and g from below (Theorem 11.20). We have

$$\int (s_n + r_n) \, d\mu = \int s_n \, d\mu + \int r_n \, d\mu.$$

Note that $\{s_n + r_n\}$ is a monotone sequence approaching $f + g$ from below. So by monotone convergence theorem we can take the limit to get

$$\int h \, d\mu = \int f \, d\mu + \int g \, d\mu.$$

Now suppose that $f \geq 0$ and $g \leq 0$. Let $A = \{x : h(x) \geq 0\}$ and $B = \{x : h(x) < 0$. Then on A, h, $-g$, and f are nonnegative and so

$$\int_A f \, d\mu = \int_A \big(h + (-g)\big) \, d\mu = \int_A h \, d\mu + \int_A (-g) \, d\mu = \int_A h \, d\mu - \int_A g \, d\mu.$$

On B, $-h$, $-g$, and f are nonnegative.

$$-\int_B g \, d\mu = \int_B (-g) \, d\mu = \int_B \big(f + (-h)\big) \, d\mu = \int_B f \, d\mu - \int_B h \, d\mu.$$

We now can write

$$\int h \, d\mu = \int_A h \, d\mu + \int_B h \, d\mu = \int_A f \, d\mu + \int_A g \, d\mu + \int_B f \, d\mu + \int_B g \, d\mu = \int f \, d\mu + \int g \, d\mu.$$

We divide the space into 4 pairwise disjoints sets where f and g have constant sign. We apply the two above cases to get the result in each of the four sets and we put them together just like above. We leave the details to the reader.

Similarly, if f and g are complex valued, then we just apply the result to the real and imaginary parts. □

In other words for any finite sum of nonnegative or integrable functions we have

$$\int \sum_{j=1}^{n} f_j(x)\, d\mu = \sum_{j=1}^{n} \int f_j(x)\, d\mu.$$

Therefore we have a corollary of the monotone convergence theorem.

Corollary 11.30: Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose $\{f_n\}$ are nonnegative and measurable functions. Then

$$\int \sum_{n=1}^{\infty} f_n(x)\, d\mu = \sum_{n=1}^{\infty} \int f_n(x)\, d\mu.$$

What can we say if we don't have monotonicity? The following is classically called the Fatou Lemma, though Rudin calls it the Fatou Theorem.

Theorem 11.31 (Fatou's lemma): Let $(X, \mathcal{M}, \mu)$ be a measure space. If $\{f_n\}$ is a sequence of nonnegative measurable functions then

$$\int \liminf_{n\to\infty} f_n(x)\, d\mu(x) \leq \liminf_{n\to\infty} \int f_n(x)\, d\mu(x).$$

Example: The way to remember which way the inequality goes (and to see why we really need an inequality) is to think of the following example: Let $f_n = \chi_{[n,n+1]}$. Then $\liminf_{n\to\infty} f_n(x) = 0$ for all x, but $\int f_n dm = 1$ for all n.

Proof. For any n let

$$g_n(x) = \inf_{k\geq n} f_k(x)$$

The g_n are measurable and now they are also monotone increasing

$$0 \leq g_1(x) \leq g_2(x) \leq \cdots .$$

Furthermore $\lim_{n\to\infty} g_n(x) = \liminf_{n\to\infty} f_n(x)$ by definition of $\liminf$. So using the monotone convergence theorem,

$$\int \liminf_{n\to\infty} f_n\, d\mu = \int \lim_{n\to\infty} g_n\, d\mu = \lim_{n\to\infty} \int g_n\, d\mu = \liminf_{n\to\infty} \int g_n\, d\mu \leq \liminf_{n\to\infty} \int f_n\, d\mu.$$

The last inequality because $g_n \leq f_n$ for all n. □

Theorem 11.32 (Lebesgue's dominated convergence theorem): Let $(X, \mathcal{M}, \mu)$ be a measure space. Let $\{f_n\}$ be a sequence of measurable functions converging pointwise almost everywhere to a function $f\colon X \to \mathbb{C}$, and suppose that there exists a function $g \in L^1(X, \mu)$ such that

$$|f_n(x)| \leq g(x)$$

for almost every x and all n. Then

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu.$$

It is instructive to think about why the dominated convergence theorem does not apply to the sequence in the example after Fatou's lemma, that is $f_n = \chi_{[n,n+1]}$. We see that a g would have to be at least identically 1 from some point onwards, and such a g would never be integrable.

Another sequence that is useful to think about is $f_n = n\chi_{(0,1/n]}$. $\{f_n\}$ goes pointwise to 0, but $\int_0^1 f_n(x)\, dx = 1$ for all n. Note that there is no g again. This time because the sequence "blows up" too quickly near the origin.

These two behaviours are the two things that can in general "go wrong." Either the set where all the action happens is "escaping to infinity," or the sequence "blows up" somewhere. Having a dominating $g \in L^1$ avoids both of these types of behaviours.

Proof. First we note that by changing f_n's and g on a set of measure zero doesn't change their integrals. Therefore, if we redefine $f_n(x) = f(x) = g(x) = 0$ for all the x where convergence did not happen, we can just assume without loss of generality that f_n goes to f pointwise everywhere, and furthermore we can for the same reason assume that $|f_n(x)| \leq g(x)$ for all x.

We have that $f_n \in L^1$ and by taking a limit we have that $|f(x)| \leq g(x)$ and so $f \in L^1$.

Also note that $\bigl|\operatorname{Re}\bigl(f_n(x)\bigr)\bigr| \leq |f_n(x)| \leq g(x)$ for all x, and same for the imaginary part. Therefore the hypotheses apply to the real and imaginary part of f_n and f. If we prove the theorem for real functions, it is easy to see that the theorem applies for complex valued functions. So assume from now on that $\{f_n\}$ and f are all real-valued.

Now $f_n + g \geq 0$, so apply Fatou's lemma to get

$$\int (f+g)\,d\mu \leq \liminf_{n\to\infty} \int (f_n + g)\,d\mu.$$

By linearity we get

$$\int f\,d\mu \leq \liminf_{n\to\infty} \int f_n\,d\mu.$$

Similarly $g - f_n \geq 0$ and so by Fatou,

$$\int (g-f)\,d\mu \leq \liminf_{n\to\infty} \int (g - f_n)\,d\mu.$$

Again by linearity we get

$$-\int f\,d\mu \leq \liminf_{n\to\infty} \left(-\int f_n\,d\mu \right),$$

or

$$\int f\,d\mu \geq \limsup_{n\to\infty} \int f_n\,d\mu.$$

In other words

$$\int f\,d\mu \geq \limsup_{n\to\infty} \int f_n\,d\mu \geq \liminf_{n\to\infty} \int f_n\,d\mu \geq \int f\,d\mu.$$

This implies the theorem. □

Exercise: Prove reverse Fatou: Let $(X, \mathcal{M}, \mu)$ be a measure space. If $\{f_n\}$ is a sequence of measurable functions and $g \in L^1(\mu)$ such that $f_n \leq g$ for all n, then

$$\limsup_{n\to\infty} \int f_n(x)\,d\mu(x) \leq \int \limsup_{n\to\infty} f_n(x)\,d\mu(x).$$

Define

$$f_n = {}^1\!/_n \chi_{[n,2n]}.$$

Then f_n's go to 0 uniformly on $\mathbb{R}$, yet $\int f_n = 1$ for all n. But we do have the following. If the space is of finite measure though, we can in fact swap limits.

Exercise: Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < \infty$. Let $\{f_n\}$ be a sequence of measurable functions that converges uniformly to $f\colon X \to \mathbb{C}$. Then show that

$$\lim_{n\to\infty} \int f_n\,d\mu = \int f\,d\mu.$$

In fact a far stronger result is true.

Exercise: Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < \infty$. Let $\{f_n\}$ be a uniformly bounded (there exists an M such that $|f_n(x)| \leq M$ for all x and all n) sequence of measurable functions that converges pointwise to $f\colon X \to \mathbb{C}$. Then show that

$$\lim_{n\to\infty} \int f_n\,d\mu = \int f\,d\mu.$$

Exercise: Let $L^1(X,\mu)$ denote the equivalence classes of functions equal almost everywhere. Prove that $L^1(\mu)$ is a complete metric space with the metric

$$d(f,g) = \|f-g\|_{L^1} = \int |f-g| \, d\mu,$$

where we take any representative f and g of the equivalence class.

Let us prove a strong version of the "differentiate under the integral sign" theorem.

Corollary: Let $I \subset \mathbb{R}$ be an open interval and let $(Y,\mathcal{M},\mu)$ be a measure space. Suppose $f\colon I \times Y \to \mathbb{C}$ satisfies all of the following:

(i) For every fixed $x \in I$, the function $y \mapsto f(x,y)$ is in $L^1(Y,\mu)$.
(ii) For almost every $y \in Y$, the derivative $\frac{\partial f}{\partial x}(x,y)$ exists for all $x \in I$.
(iii) There is a $g \in L^1(Y,\mu)$ such that $\left|\frac{\partial f}{\partial x}(x,y)\right| \leq g(y)$ for all $x \in I$ and almost every $y \in Y$ (in particular only when the derivative is defined).

Then

$$\frac{\partial}{\partial x}\left[\int_Y f(x,y)\, d\mu(y)\right] = \int_Y \frac{\partial f}{\partial x}(x,y)\, d\mu(y)$$

for all $x \in I$.

Here we may be committing a slight abuse of notation $\frac{\partial f}{\partial x}(x,y)$ is defined almost everywhere only. But since we are integrating it, this doesn't matter, we can just set it to whatever we wish on the set where it is not defined.

Proof. Fix $x \in I$. Pick $\{x_n\}$ in I such that $\lim x_n = x$. Now for any $y \in Y$ take

$$\varphi_n(y) = \frac{f(x_n,y)-f(x,y)}{x_n - x}.$$

We have that φ_n goes to $\frac{\partial f}{\partial x}(x,y)$ pointwise almost everywhere. So suppose that y is such that the derivative exists. Then by mean value theorem there is a t between x_n and x such that

$$\varphi_n(y) = \frac{\partial f}{\partial x}(t,y).$$

So

$$|\varphi_n(y)| = \left|\frac{\partial f}{\partial x}(t,y)\right| \leq g(y)$$

almost everywhere. We can now apply dominated convergence theorem to

$$\frac{\int f(x_n,y)\, d\mu(y) - \int f(x,y)\, d\mu(y)}{x_n - x} = \int \frac{f(x_n,y)-f(x,y)}{x_n-x}\, d\mu(y) = \int \varphi_n \, d\mu.$$

□

To avoid the "almost everywhere"s in the argument, we could have also only taken the subset of Y for which the derivative exists to begin with, and just work there. The result would be the same.

Exercise: Prove the following generalization: Let $I \subset \mathbb{R}$ be an open interval and let $(Y,\mathcal{M},\mu)$ be a measure space. Suppose $f\colon I \times Y \to \mathbb{C}$ satisfies all of the following:

(i) For every fixed $x \in I$, the function $y \mapsto f(x,y)$ is in $L^1(Y,\mu)$.
(ii) There is an $x_0 \in I$ such that for almost every $y \in Y$, there exists an $\epsilon_y > 0$, such that the derivative $\frac{\partial f}{\partial x}(x,y)$ exists for all $x \in (x_0-\epsilon_y, x_0+\epsilon_y) \subset I$.
(iii) There is a $g \in L^1(Y,\mu)$ such that for almost every $y \in Y$, the inequality $\left|\frac{\partial f}{\partial x}(x,y)\right| \leq g(y)$ holds for all $x \in (x_0-\epsilon_y, x_0+\epsilon_y)$.

Then

$$\frac{\partial}{\partial x}\Big|_{x=x_0}\left[\int_Y f(x,y)\,d\mu(y)\right] = \int_Y \frac{\partial f}{\partial x}(x_0,y)\,d\mu(y).$$

Note: By $\frac{\partial}{\partial x}\big|_{x=x_0}$ we mean the derivative at x_0.

Exercise: Prove the following classical version: If $f\colon [a,b]\times[c,d]\to\mathbb{C}$ is continuous, and $\frac{\partial f}{\partial x}(x,y)$ exists and is continuous on $[a,b]\times[c,d]$, then

$$\frac{\partial}{\partial x}\left[\int_c^d f(x,y)\,dy\right] = \int_c^d \frac{\partial f}{\partial x}(x,y)\,dy.$$

The Riemann integral via the Lebesgue integral

We still have not shown that the Lebesgue integral is an integral in the sense that we are used to. That is, that the Lebesgue integral and the Riemann integral agree on Riemann integrable functions.

To distinguish the Riemann and the Lebesgue integral, let us write

$$\mathcal{R}\!\int_a^b f(x)\,dx$$

for the Riemann integral. In the following we use the Lebesgue measure m on $\mathbb{R}$ and we write

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,dm.$$

Theorem 11.33:

(i) If $f\colon [a,b]\to\mathbb{C}$ is Riemann integrable, then it is Lebesgue integrable on $[a,b]$ and

$$\int_a^b f(x)\,dx = \mathcal{R}\!\int_a^b f(x)\,dx.$$

(ii) The function $f\colon [a,b]\to\mathbb{C}$ is Riemann integrable if and only if f is bounded and continuous almost everywhere on $[a,b]$.

Proof. If we prove the result for real-valued functions it is easy to extend it to complex valued functions. Let $f\colon [a,b]\to\mathbb{R}$ be a bounded function. Let $P=\{x_0,\ldots,x_n\}$ be a partition of $[a,b]$, that is a finite set of points such that $a=x_0<x_1<\cdots<x_n=b$. Define

$$m_j = \inf\{f(x) : x\in[x_{j-1},x_j]\} \qquad \text{and} \qquad M_j = \sup\{f(x) : x\in[x_{j-1},x_j]\}.$$

Define the step functions

$$s = m_1\chi_{[x_0,x_1]} + \sum_{j=2}^n m_j\chi_{(x_{j-1},x_j]} \qquad \text{and} \qquad r = M_1\chi_{[x_0,x_1]} + \sum_{j=2}^n M_j\chi_{(x_{j-1},x_j]}.$$

Note that for all $x\in[a,b]$ we have $s(x)\le f(x)\le r(x)$.

It is not hard to see that we can pick a sequence $\{P_k\}$ of partitions with $P_k\subset P_{k+1}$ (a sequence of refinements) and such that

$$\underline{\int_a^b} f(x)\,dx = \lim_{k\to\infty} L(P_k,f) \qquad \text{and} \qquad \overline{\int_a^b} f(x)\,dx = \lim_{k\to\infty} U(P_k,f),$$

where $L(P_k,f)$ and $U(P_k,f)$ are the lower and upper Darboux sums, and $\underline{\int_a^b}$ and $\overline{\int_a^b}$ are the lower and the upper Darboux integrals.

Let s_k and r_k be the step functions corresponding to P_k. It is easy to see that

$$\int_a^b s_k(x)\,dx = L(P_k,f) \qquad \text{and} \qquad \int_a^b r_k(x)\,dx = U(P_k,f).$$

Because the P_k are successive refinements, we have that $s_k(x) \leq s_{k+1}(x) \leq f(x) \leq r_{k+1}(x) \leq r_k(x)$ for all x. We have that $\{s_k\}$ and $\{r_k\}$ are monotone and pointwise bounded and so they have a pointwise limit. Let

$$g(x) = \lim_{k\to\infty} s_k(x) \qquad \text{and} \qquad h(x) = \lim_{k\to\infty} r_k(x).$$

By monotone convergence theorem we have that

$$\underline{\int_a^b} f(x)\,dx = \int_a^b g(x)\,dx \qquad \text{and} \qquad \overline{\int_a^b} f(x)\,dx = \int_a^b h(x)\,dx.$$

Recall f is Riemann integrable if and only if $\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx$. Or in other words if and only if

$$\int_a^b h(x) - g(x)\,dx = 0.$$

As $h(x) \geq g(x)$ for all x, we have (by an exercise) that $h(x) = g(x)$ a.e. Now suppose that $h(x) = g(x)$ a.e. Then as $g(x) \leq f(x) \leq g(x)$ a.e., we have $g(x) = f(x)$ a.e. So $f(x) \in L^1$ (in particular it is measurable), and

$$\int_a^b f(x)\,dx = \int_a^b g(x)\,dx = \underline{\int_a^b} f(x)\,dx = \mathcal{R}\int_a^b f(x)\,dx.$$

This proves the first part of the theorem.

Now suppose that $h(x) = g(x)$ a.e. Fix x such that $x \notin P_k$ for all k, and such that $h(x) = g(x)$. It is not hard to see that f must be continuous at x: Given an $\epsilon > 0$, simply choose k large enough such that for the s_k and r_k the interval that contains x satisfies $M_j - m_j < \epsilon$. Then we must have that f is stuck between m_j and M_j for a whole interval around x (because x is not an endpoint of one of the subintervals of the partition P_k).

For the opposite direction let us make a further assumption that P_k has width at most $1/k$, that is, the size of the largest interval in P_k is at most $1/k$. Suppose that f is bounded and continuous almost everywhere. Let x be a point where f is continuous and $x \notin P_k$ for all k. Then given $\epsilon > 0$ find a $K > 0$ such that $|f(x) - f(y)| < \epsilon$ for all y such that $|x - y| < 1/k$ for all $k \geq K$. If $k \geq K$, and $x \in [x_{j-1}, x_j]$ in the partition P_k, then from continuity we conclude that $f(x) - s_k(x) = f(x) - m_j \leq \epsilon$ and $r_k(x) - f(x) = M_j - f(x) \leq \epsilon$. Hence $g(x) = h(x)$.

Now note that f is Riemann integrable if and only if f is bounded and $h(x) = g(x)$ a.e. The union of all the P_k is still only a countable (and hence measure zero) set. So f is Riemann integrable if and only if it is bounded and continuous almost everywhere. □

Notice a funky thing: we have proved a result about Riemann integral (classification of Riemann integrable functions) using the Lebesgue integral machinery. For example, we have seen last semester that the popcorn function defined on $(0, 1)$

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1/n & \text{if } x = m/n \text{ in lowest terms} \end{cases}$$

is continuous at all the irrational points, and hence is continuous almost everywhere. So as an immediate consequence we obtain that f is Riemann integrable, and furthermore since it equals 0 almost everywhere, then

$$\int_0^1 f(x)\,dx = 0.$$

Anything we know about the Riemann integral carries over to Lebesgue integral. Although some theorems do require a bit more work if we want to state them in full generality. For example, we leave it to the reader to prove that if $f \in L^1(\mathbb{R})$ then the function

$$F(x) = \int_{-\infty}^x f(x)\,dx$$

is continuous. The proofs are often similar to those for the Riemann integral.

Be careful about using this theorem and improper Riemann integrals. For example,

$$\int_0^\infty \frac{\sin(x)}{x}\,dx = \lim_{b\to\infty}\int_0^b \frac{\sin(x)}{x}\,dx = \frac{\pi}{2}$$

when thought of as an improper Riemann integral. Let's not worry now about how to prove that, a proof requires complex analysis. It is not too difficult to show that the limit exists by explicit estimation. But $\frac{\sin(x)}{x}$ is not in L^1 as

$$\int_0^\infty \left|\frac{\sin(x)}{x}\right| dx = \infty.$$

Which is also not too hard to show. We leave it as an exercise to show the two facts we mentioned. The hint is to use the harmonic series.

Examples of Lebesgue integration over other measures.

Example: Suppose that $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ is a measure space where μ is the counting measure (that is $\mu(A) = |A|$). Then for $f\colon \mathbb{N} \to \mathbb{C}$ is integrable if and only if $\sum f(n)$ is absolutely summable, and in this case we have,

$$\int_{\mathbb{N}} f(n)\,d\mu = \sum_{n=1}^{\infty} f(n).$$

Example: The δ-function that we have mentioned before is also a measure. Take the set $\mathbb{R}$ with the σ-algebra $\mathcal{P}(\mathbb{R})$ of all subsets of $\mathbb{R}$. The δ-function is really the measure defined by

$$\delta(A) = \begin{cases} 1 & \text{if } 0 \in A, \\ 0 & \text{if } 0 \notin A. \end{cases}$$

We leave it to the reader that this really is a measure. Note that all functions are measurable, and all functions where $|f(0)| < \infty$ are integrable, and we get that

$$\int f\,d\delta = f(0).$$

This is usually written as

$$\int_{-\infty}^{\infty} f(x)\delta(x)\,dx = f(0),$$

although that is somewhat of an abuse of notation as $\delta(x)$ is not a function. There is no need to only use 0. We could define δ_y to be the measure that tests if $y \in A$, and then $\int f\,d\delta_y = f(y)$.

Example: You could also combine measures. The measure $\mu = m + \delta$ is a measure such that

$$\int f\,d\mu = \int f\,dm + f(0).$$

Example: Another example is the measure defined by $d\mu(x) = f(x)\,dm(x)$ (that is, $\mu(A) = \int_A f\,dm$) for some measurable $f \geq 0$. Then $\int g\,d\mu = \int g(x)f(x)\,dm(x)$.

Exercise: Let $\{f_n\}$ be a sequence of measurable functions converging uniformly to 0, show that

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} \frac{f_n(x)}{1+x^2}\,dx = 0.$$